Learning about the Long Run

Leland E. Farmer

University of Virginia

Emi Nakamura

University of California Berkeley

Jón Steinsson

University of California Berkeley

Forecasts of professional forecasters are anomalous: they are biased, and forecast errors are autocorrelated and predictable by forecast revisions. We propose that these anomalies arise because professional forecasters do not know the model that generates the data. We show that Bayesian agents learning about hard-to-learn features of the world can generate all the prominent aggregate anomalies emphasized in the literature. We show this for professional forecasts of nominal interest rates and Congressional Budget Office forecasts of gross domestic product growth. Our learning model for interest rates can explain observed deviations from the expectations hypothesis of the term structure without relying on time variation in risk premia.

I. Introduction

For almost half a century, the assumption that people form rational expectations has dominated economic modeling in macroeconomics and

We thank Abhi Gupta, Ethan McClure, Venance Riblier, and Sharath Sonti for excellent research assistance. We thank Assaf Ben-Shoham, Daniel Benjamin, John Campbell, Ryan Chahrour, Gary Chamberlain, Anna Cieslak, Nicolas Chopin, Ian Dew-Becker, Benjamin

Electronically published August 7, 2024

Journal of Political Economy, volume 132, number 10, October 2024.

https://doi.org/10.1086/730207

finance. During this time, a substantial empirical literature has formulated and evaluated tests of rational expectations. One finding from this literature has been that even professional forecasters consistently fail such tests. Professional forecasts seem to suffer from a long list of anomalies. For example, they are biased, forecast errors are autocorrelated, and forecast revisions predict future forecast errors.

A related literature has tested the expectations hypothesis of the term structure. If the expectations hypothesis holds, yields on long-term bonds are the bond market's forecast of future short rates (modulo a constant risk premium). Empirical tests of the expectations hypothesis fail spectacularly (e.g., Campbell and Shiller 1991). One reaction to this finding is that risk premia in the bond market are time varying (Wachter 2006; Bansal and Shaliastovich 2013; Vayanos and Vila 2021). An alternative view is that the this finding reflects forecasting anomalies among bond traders (Froot 1989).[1]

The traditional reaction to forecasting anomalies in macroeconomics is that they imply that professional forecasters are irrational, that is, that forecasters are not making efficient use of the information available to them (Mincer and Zarnowitz 1969; Friedman 1980; Nordhaus 1987; Maddala 1991; Croushore 1998; Schuh 2001). Recent behavioral work develops this perspective (e.g., Bordalo et al. 2020). An alternative reaction is that these anomalies result from information frictions (Mankiw, Reis, and Wolfers 2003; Coibion and Gorodnichenko 2012, 2015). The most prominent models of information frictions in macroeconomics are sticky information models (Mankiw and Reis 2002) and noisy information models (Sims 2003; Woodford 2003). These models seem eminently plausible for households and firms. Arguably, they are less well suited to explain the behavior of professional forecasters (and bond traders). Professional forecasters read the news every day and have no trouble observing the relevant data precisely (i.e., without noise).

In this paper, we consider another explanation. Standard tests of rational expectations impose the very strong assumption that agents know the model that generates the variables that are being forecast (parameter values and all). In reality, nobody knows the correct model of the world. Since professional forecasters do not know the correct model of the world, they use incoming data to learn about how the world works. But such learning can

Friedman, Gopi Gaswami, Robert Hodrick, Michael Johannes, David Laibson, Lars Lochstoer, Sydney Ludvigson, Omiros Papaspiliopoulos, Monika Piazzesi, Martin Schneider, Andrei Shleifer, Richard Thaler, and seminar participants at various institutions for valuable comments and discussions. We thank the Alfred P. Sloan Foundation, the Smith Richardson Foundation, and the Bankard Fund for Political Economy at the University of Virginia for financial support. This paper was edited by Andrew Atkeson.

[1] See also Bekaert, Hodrick, and Marshall (2001), Piazzesi, Salomao, and Schneider (2015), Cieslak (2018), Xu (2019), and Nagel and Xu (2021).

fundamentally change the dynamics of even perfectly rational Bayesian forecasts. This idea has been recognized by researchers at least since Friedman (1979).[2] Models in which learning has been shown to be important include long-run risk models and models with disasters (Cogley and Sargent 2008; Croce, Lettau, and Ludvigson 2015; Bidder and Dew-Becker 2016; Collin-Dufresne, Johannes, and Lochstoer 2016; Kozlowski, Veldkamp, and Venkateswaran 2020).[3]

Realistic learning models are difficult to solve. As a consequence, early work on learning used relatively simple models. But in such models, Bayesian learning occurs quickly, suggesting that rational learning cannot explain forecasting anomalies that persist over multiple decades. Structural breaks have sometimes been invoked as a reason why learning might persist over long periods of time, but such arguments have been informal.

Bayesian learning can, however, be extremely slow in richer, more realistic models (Johannes, Lochstoer, and Mou 2016). Consider, for example, models with multiple unobserved components, some stationary and others containing a unit root. A key property of such models is that the long-run trajectory of a variable may move quite independently from the short-run dynamics of that variable (if the short-run dynamics are dominated by the stationary components). This means that the quarter-to-quarter dynamics of the variable may be quite uninformative about its longer-run properties. Since information about low-frequency properties accumulates slowly, learning about the long run can be extremely slow. In such models, several different parameter combinations may yield a similar fit for the high-frequency behavior of the series but may have very different implications about the low-frequency behavior of the series. We show that in such cases it can take many decades to learn the true parameters.

We develop two applications of these ideas, one for forecasting nominal interest rates and another for forecasting real gross domestic product (GDP) growth. In each case, we endow Bayesian forecasters with an unobserved components model and initial beliefs about the parameters of this model. Each period, these agents use real-time US data to update their beliefs about the parameters and states of the model. They then forecast the variable in question, and we assess whether the resulting forecasts are anomalous.

Our main result is that we are able to match all the main aggregate forecasting anomalies emphasized in the prior literature for both interest rates

[2] Other important papers that emphasize this idea include Caskey (1985), Lewis (1989a, 1989b), Barsky and De Long (1993), Timmermann (1993), Brav and Heaton (2002), Lewellen and Shanken (2002), Cogley and Sargent (2005), Guo and Wachter (2019), Singleton (2021).

[3] Bianchi, Ludvigson, and Ma (2022) analyze the performance of a machine-learning algorithm for forecasting in a data-rich environment in which the true model is unknown. Andolfatto, Hendry, and Moran (2008) point out the potential for small sample rejections of rational expectations tests in an exercise that is conceptually similar to ours but without attempting to fit the quantitative magnitude of the deviations from rational expectations.

and real GDP when forecasters are endowed with reasonable initial beliefs. In addition, we construct long-term yield data from our model-generated forecasts of nominal interest rates, assuming that the expectations hypothesis holds. We then run a battery of standard tests of the expectations hypothesis on these data. The model-generated yield data fail the tests of the expectations hypothesis in exactly the same way as do real-world bond yields. Notably, our sample period is roughly 40 years for the forecast data (60 years for the term structure data), and we endow our Bayesian agents with data back to the early post–World War II period. Even though they learn for quite a few decades, agents' forecasts continue to display anomalies.

Since learning is slow in our unobserved components model, agents' initial beliefs matter for a long time. An important question is whether these findings rely on very tight (dogmatic) initial beliefs. This is not the case. The initial beliefs we endow agents with are quite dispersed. In this sense, we show that we can match the anomalous features of the forecast data with reasonable initial beliefs. Furthermore, the initial beliefs we endow agents with accord well with historical experience prior to our sample period. For example, our agents place small weight in 1951 on the possibility that the nominal interest rate has a large random walk component. This is consistent with the fact that (outside of war) the United States had been on a gold (or silver) standard almost continuously from its founding until that point in time and interest rates had therefore been quite stable. The large and persistent rise and fall in nominal interest rates that occurred subsequently was far outside of what had been experienced up to that point in history.[4]

Our findings demonstrate that many apparent anomalies can be rationalized by the same underlying phenomenon: initial beliefs that turn out (ex post) not to be centered on the right location in the parameters space. While the initial beliefs required for our explanation to work are quite dispersed, they are not flat. One might reasonably ask whether it is irrational for agents to deviate from flat initial beliefs. Interestingly, however, we show that flat initial beliefs would not have led to appreciably smaller root mean squared errors (RMSEs). This finding echoes the more general finding in the forecasting literature that allowing for unrestricted priors in complicated learning models often does not improve forecasting performance. Bianchi, Ludvigson, and Ma (2022) show that including lagged forecast revisions in a forecasting model actually worsens out-of-sample forecasting performance, though the predictive content of this variable leads to failures of standard rational expectations tests.

A potential concern with our results is that perhaps we are able to match the forecast anomalies we emphasize because we endow agents with a misspecified model. To address this concern and understand better what drives our results, we conduct a Monte Carlo simulation of our

[4] See Fama (2006, 360–61) for a narrative description of these ideas.

model for nominal interest rates. In this case, we know the true model and thus know that the agents in our model are not learning using a misspecified model. We show that when initial beliefs are centered on parameters that imply too little persistence in interest rates relative to the truth, our model generates the kinds of anomalies we find in the data.[5] In contrast, if initial beliefs are centered on parameters that imply too much persistence, our model generates anomalies in the opposite direction (e.g., negatively autocorrelated forecast errors and overreaction rather than underreaction in Coibion and Gorodnichenko [2015] regressions). If initial beliefs happen to be exactly centered on the true values in our Monte Carlo, no anomalies arise.

In the Monte Carlo simulations, we know what the truth is. When it comes to the real world, there is no way of knowing what the truth is without learning, and learning about the long run can be extremely slow. In our Monte Carlo simulation, a decade is a blink of an eye in terms of learning about key parameters of in our model. Even after agents have been learning for 70 years, they are still very far from the truth and are inching toward the truth extremely slowly. These results illustrate how rational expectations tests can be very misleading even when run over long periods of time. They are also related to the fact that unit root tests have low power in short samples (short often being many decades).

Whether anomalies arise from Bayesian learning about parameters, however, depends crucially on the nature of the data. If the fluctuations in a variable of interest are homoscedastic and not very persistent, information about model parameters will accumulate quickly. The same is true when a variable displays a regular pattern over and over again (such as daily and annual cycles in the weather). In these cases, agents will learn the value of model parameters relatively quickly and none of the issues we emphasize will persist for very long.

Bordalo et al. (2020) document that while underreaction to news is a pervasive phenomenon for consensus (i.e., mean) forecasts, the forecasts of individual forecasters tend to overreact to news for a number of macroeconomic variables (although not for interest rates). They propose a model with two features to match these facts: (1) noisy information to generate underreaction of consensus forecasts and (2) diagnostic expectations to generate overreaction of individual forecasts. We view our model as an alternative to the first feature in Bordalo et al. (2020): uncertainty about the data-generating process is (arguably) a more plausible information friction than noisy information for professional forecasters. One could layer diagnostic expectations on top of our model to match overreaction at

[5] This result is similar in spirit to results in Gourinchas and Tornell (2004) about exchange rates.

the individual level, just as Bordalo et al. (2020) combine diagnostic expectations with noisy information.

We show that for the anomalies we study, the behavior of individual forecasts is very similar to the consensus forecast.[6] An important literature has sought to understand the behavior of individual forecasts relative to the mean forecast as well as forecast dispersion (e.g., Patton and Timmermann 2011; Andrade et al. 2016; Angeletos, Huo, and Sastry 2020; Cao et al. 2021; Crump et al. 2021; Singleton 2021; Broer and Kohlhas 2024). Patton and Timmermann (2010) document that disagreement among forecasters is largest about long-run outcomes and persists over time. They argue that this points to the disagreement arising because of heterogeneity in priors rather than differences in information sets. Analyzing this prediction is beyond the scope of our current analysis (what we do is already very computationally demanding.) However, we view this as an important topic for future work.

Our work also relates to a rich literature on boundedly rational learning in macroeconomics (e.g., Evans and Honkapohja 2001; Sargent 2001; Eusepi and Preston 2011, 2018; Giacoletti, Laursen, and Singleton 2018; Molavi, Tahbaz-Salehi, and Vedolin 2021). Kohlhas and Robertson (2022) show that rational forecasters seeking to minimize mean squared error optimally react cautiously to incoming signals and that this gives rise to forecasting anomalies. Ben-David, Graham, and Harvey (2013) provide evidence for Bayesian learning among firm chief financial officers.

The paper proceeds as follows. Section II describes our data. Section III reviews forecasting anomalies for interest rates and real GDP data. Section IV presents our model and results for nominal interest rates. Section V presents our model and results for real GDP growth. Section VI presents Monte Carlo simulation exercises aimed to shed light on why our results turn out the way they do. Section VII concludes.

II. Data

The paper discusses two applications, one to interest rate forecasting and the other to real GDP forecasting. This section describes the data we use for these two applications in turn.

A. Interest Rate Data and Forecasts

The forecast data we use for the 3-month Treasury bill (T-bill) rate come from the Survey of Professional Forecasters (SPF) conducted by the Federal

[6] There is some difference for the underreaction anomaly. But quantitatively it is relatively minor.

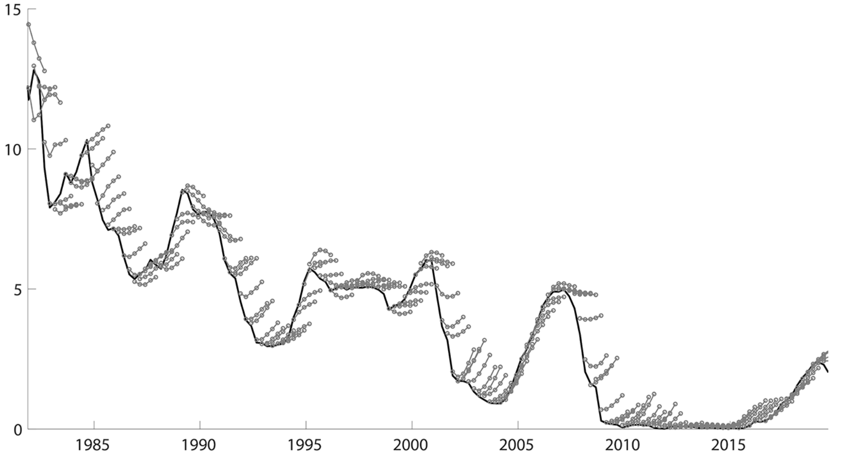

FIG. 1.—SPF forecasts of 3-month T-bill rate. The black line is the 3-month T-bill rate. Each short gray line with five circles represents the SPF forecasts made in a particular quarter about the then present quarter (first circle) and following four quarters (subsequent four circles).

Reserve Bank of Philadelphia. Our sample period for these forecasts is 1981:3 to 2019:4. The SPF is a quarterly survey sent out to a rotating panel of forecasters. Our main analysis uses the mean forecast across forecasters, but we also present results for individual forecasters. Figure 1 plots the mean forecasts.

The survey is sent out near the end of the first month of each quarter. The forecast therefore roughly coincides with the advance report of the national income and product accounts from the US Bureau of Economic Analysis. Survey response deadlines are in the second to third week of the second month of the same quarter. Survey respondents are asked to provide nowcasts and one- to four-quarter-ahead forecasts of the quarterly average 3-month T-bill secondary market rate. The timing of these forecasts is as follows: the nowcast pertains to the quarterly average rate at the end of the quarter when the survey is received, and the subsequent forecasts pertain to quarterly averages for each of the following four quarters.

The data we use on the 3-month T-bill secondary market rate are from the Board of Governors of the Federal Reserve System.[7] Our sample period for this series is 1951:2 to 2019:4. Figure 2 plots the series. To be consistent with the forecast data, we use quarterly averages of the daily interest rate. We also use daily estimates of the zero coupon yield curve from Liu and Wu (2021). Liu and Wu estimate the zero coupon yield curve for bonds of maturity 1 month to 30 years (360 months), dating back to June 1961. We convert

[7] Specifically, we use the following series: https://fred.stlouisfed.org/series/TB3MS.

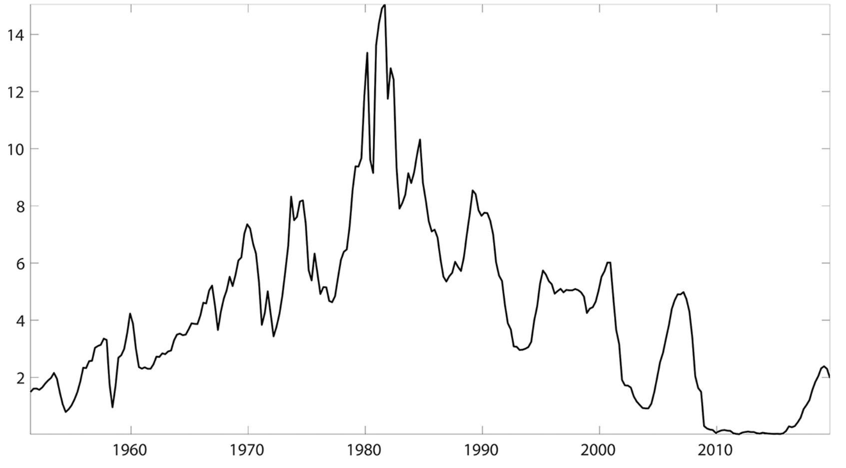

Fig. 2.—3-month T-bill rate.

these data to quarterly data by computing the average yield in a quarter. Our sample period for these zero coupon bond yields is 1961:3 to 2019:4.

B. *Real GDP Growth Data and Forecasts*

The real GDP growth forecasts we analyze are from the Congressional Budget Office (CBO). Our sample period for these forecasts is 1976 to 2019. The CBO releases its annual economic outlook at the beginning of each year, where it provides projections for current and future real economic growth. Since 1996, the CBO has made projections out to a horizon of 11 years. Before that, they made projections out to a horizon of 6 years. The CBO forecasts the annual average level of real output over each calendar year. Growth rates are then computed as percentage changes in these average levels across years. Up to and including their 1992 report, the CBO forecasts real gross national product. Since then, they have forecast real GDP. For expositional simplicity, we refer to these as real GDP forecasts throughout the paper. Figure 3 plots the CBO forecasts. Salient features of this figure include a series of large positive forecast errors in the 1990s (truth above the forecast) and a series of large negative forecast errors in the aftermath of the Great Recession.

The data we use on actual real GDP growth are from the Real-Time Data Set of the Federal Reserve Bank of Philadelphia. This source publishes monthly vintages of real-time real output back to November 1965. Most vintages contain data back to 1947:1. However, a few vintages are missing

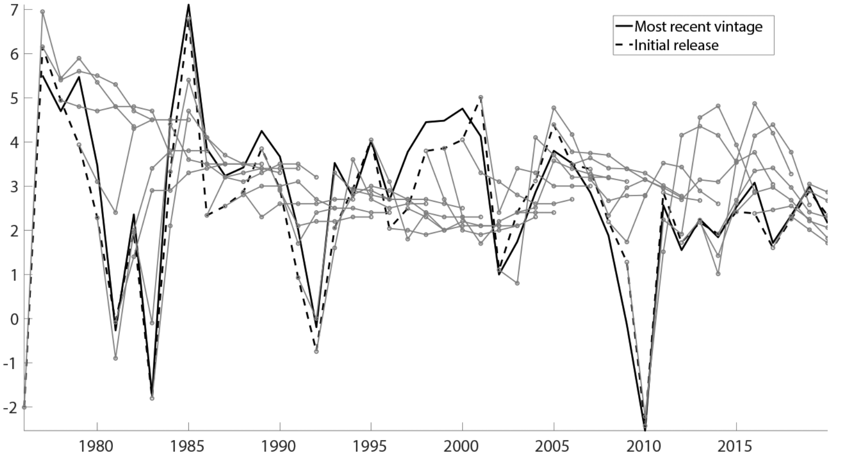

FIG. 3.—CBO forecasts of real GDP growth. The solid black line is the 2021:1 vintage of real GDP growth from 1976 to 2019. The dashed black line is the initial release of GDP growth at each point in time. Each short gray line with seven circles represents the initial release of real GDP for the previous year (first circle) and the CBO forecasts made in a particular year about GDP growth in the following 6 years (subsequent six circles).

data before 1959:3, which limits our sample period, as we discuss in greater detail in section V.

III. Forecasting Anomalies

As we discuss in the introduction, the forecasts of professional forecasters exhibit a number of anomalies—that is, patterns that previous researchers have argued suggest deviations from forecast rationality. Here we document a number of such anomalies for professional forecasts of the 3-month nominal T-bill rate and real GDP growth. We also document deviations from the expectations hypothesis of the term structure, which may arise from forecast anomalies on the part of the bond market (but may alternatively be due to time-varying risk premia). The facts we document in this section will be key empirical targets we seek to match with our models later in the paper.

The null hypotheses we consider below constitute tests of forecast rationality given two assumptions: (1) that forecasters aim to minimize the mean squared error of their forecasts, implying that optimal forecasts are equivalent to conditional expectations ($F_t y_{t+h} = E_t y_{t+h}$), and (2) that forecasters know the true model of the world. For the 3-month T-bill, we focus on forecast horizons of one to four quarters. For real GDP growth, however, we focus on forecast horizons of 1–5 years. These different forecasting horizons reflect differences in the horizons at which the forecast anomalies are most striking for the 3-month T-bill yield versus real GDP growth.

A. *Bias*

A straightforward prediction of full-information rational expectations models is that forecasts should be unbiased at all horizons. Let y_t be the variable to be forecast, and let $F_t y_{t+h}$ denote the h-period-ahead forecast of y_t given time t information. Define the forecast error as $e_{t+h|t} \equiv y_{t+h} - F_t y_{t+h}$. The bias in forecasts can then be estimated using the following regression:

$$e_{t+h|t} = \alpha + u_{t+h}, \tag{1}$$

with $\alpha = 0$ indicating that forecasts are unbiased at a given horizon h.

Panel A of table 1 displays our estimates of α for the 3-month T-bill rate and real output growth. Our estimates indicate that professional forecasts of the T-bill rate display negative bias—the truth being lower than the forecast on average—at all horizons and the magnitude of this bias increases with the horizon. At the four-quarter forecast horizon, SPF forecasters overestimate the true T-bill rate by an average of 0.7 percentage points. These biases are statistically significant at the 1% level at all horizons. In contrast, there is little evidence of statistically significant bias in CBO forecasts of GDP growth at the horizons we study.

B. *Autocorrelated Forecast Errors*

Another prediction of full-information rational expectations models is that forecast errors should be serially uncorrelated. To assess this prediction, we consider the following regression of h-period-ahead forecast errors on their own past value h periods earlier (i.e., we consider the correlation of contiguous, nonoverlapping h-period forecasts):

$$e_{t+h|t} = \alpha + \beta e_{t|t-h} + u_{t+h}. \tag{2}$$

In a full-information setting, forecast rationality implies that $\alpha = 0$ and $\beta = 0$, that is, there should be no bias and forecast errors should not be predictable by known information (the time t forecast error).

Panel B of table 1 reports our estimates of β from equation (2). SPF forecasts of the T-bill display substantial positive autocorrelation. The one-quarter forecast has an autocorrelation of 0.30. This falls to 0.24 at three quarters. These estimates are statistically significantly different from zero, especially at horizon 2. CBO forecasts of GDP growth also display positive autocorrelation. But in this case the autocorrelation is smaller and not statistically significantly different from zero.

C. *Mincer-Zarnowitz Regressions*

A classic test of forecast rationality proposed by Mincer and Zarnowitz (1969) investigates the intuitive prediction that the truth should on

TABLE 1
FORECAST ANOMALIES

	Forecast Horizon				
	1	2	3	4	5
			A. Bias		
T-bill	−.18***	−.34***	−.52***	−.70***	. . .
	(.05)	(.09)	(.14)	(.19)	
GDP growth	.27	−.27	−.54	−.62	−.52
	(.25)	(.35)	(.50)	(.53)	(.49)
			B. Autocorrelation		
T-bill	.30*	.27**	.24*	.13	. . .
	(.14)	(.12)	(.12)	(.13)	
GDP growth	.22	.16	.11	.08	.08
	(.12)	(.14)	(.13)	(.18)	(.10)
			C. Mincer-Zarnowitz		
T-bill	.97*	.94**	.90**	.86**	. . .
	(.02)	(.02)	(.04)	(.05)	
GDP growth	.94	.60	.03**	−.42***	−.43***
	(.10)	(.38)	(.27)	(.18)	(.29)
			D. Coibion-Gorodnichenko		
T-bill	.23*	.34*	.62***	. . .	. . .
	(.12)	(.16)	(.16)		
GDP growth	.08	.00	.50	−1.63**	−1.46**
	(.08)	(.28)	(.58)	(.36)	(.40)

NOTE.—The forecast horizons for the T-bill are quarters, while the forecast horizons for the GDP growth are years. Standard errors are in parentheses. Asterisks represent significance relative to the following hypotheses: $\alpha = 0$ for bias, $\beta = 0$ for autocorrelation, $\beta = 1$ for Mincer-Zarnowitz, and $\beta = 0$ for Coibion-Gorodnichenko. p-values are computed using Newey-West standard errors with lag length selected as $L = \lceil 1.3 \times T^{1/2} \rceil$ and fixed b critical values, as proposed in Lazarus et al. (2018). This corresponds to a bandwidth of 17 for the T-bill regressions and nine for the GDP growth regressions.
* $p < .10$.
** $p < .05$.
*** $p < .01$.

average move one-for-one with a rational forecast: when the forecast rises by 1%, on average, the realized value should also rise by 1%. This prediction can be analyzed using the regression

$$y_{t+h} = \alpha + \beta F_t y_{t+h} + u_{t+h}. \tag{3}$$

In a full-information setting, forecast rationality implies that $\alpha = 0$ and $\beta = 1$, that is, there should be no bias and realized values should move one-for-one with forecasts.

Panel C of table 1 reports our estimates of β from (3). In this case, it is the GDP growth forecasts that display substantial deviations from the null of forecast rationality. While the estimate of β for the 1-year-ahead forecast is close to 1, it falls sharply at longer horizons. For the 3-year-ahead forecast, we estimate a β close to zero. In other words, actual GDP growth is no more likely to be high when it was forecast to be high 3 years earlier than when it was forecast to be low 3 years earlier. For the 4- and 5-year-ahead forecast, we estimate negative values (high forecasted growth predicts low growth on average). These three estimates are strongly statistically significantly different from 1. In contrast, our estimates of β for the T-bill forecasts are close to 1. They are somewhat below 1, and the difference is statistically significant. But the deviation from the null of 1 is much less stark than in the case of GDP forecasts.

D. *Coibion-Gorodnichenko Test*

Another property of rational forecasts under full information is that they should not underreact or overreact to new information. Coibion and Gorodnichenko (2015) propose the following regression to assess this:

$$e_{t+h|t} = \alpha + \beta(F_t y_{t+h} - F_{t-1} y_{t+h}) + u_{t+h}.$$

Forecast rationality in a full-information setting implies that $\alpha = 0$ and $\beta = 0$. $F_t y_{t+h} - F_{t-1} y_{t+h}$ is known at time t, and forecast errors should not be predictable by known information. If $\beta > 0$, the forecasts are said to suffer from underreaction. In this case, an increase in the forecast predicts a situation where the new forecast is still too low on average, that is, did not increase enough. If $\beta < 0$, the forecasts are said to suffer from overreaction.

Panel D of table 1 reports our estimates of β from (3). In this case, we see opposite anomalies for the two applications we consider. For the T-bill forecasts, we see evidence of underreaction: we estimate positive values for β rising from 0.22 at the one-quarter horizon to 0.64 at the three-quarter horizon. For GDP growth forecasts, however, we estimate neither over- nor underreaction at short horizons. At the 4- and 5-year horizons, however, we estimate negative values of β, indicating overreaction.

E. *Individual Forecasts*

The T-bill results presented in table 1 are for the mean forecast among SPF forecasters. Table A.1 (tables A.1–G.4 are available online) presents analogous results for individual forecasters. Following Bordalo et al. (2020), we present results where the forecasts of individual SPF forecasters are pooled as well as the median estimate from forecaster-by-forecaster regressions.[8]

[8] We exclude forecasts that are more than five interquartile ranges away from the median and forecasters with fewer than 10 forecasts. These are the same sample restrictions that Bordalo et al. (2020) employ.

For the bias, autocorrelation, and Mincer-Zarnowitz regressions, the results are very similar for individual forecasters as they are for the mean SPF forecast (which we refer to as the consensus forecast in table A.1). For the Coibion-Gorodnichenko regressions, the anomalies we document are smaller at the individual level than at the aggregate level (but of the same sign). This last fact has been documented by Bordalo et al. (2020). Given the high degree of similarity between the anomalies at the consensus and individual level, we focus on consensus forecasts in the rest of the paper.

F. *Failures of the Expectations Hypothesis*

The expectations hypothesis of the term structure implies that the yield on an n-period bond should equal the average expected value of the yield on a 1-period bond over the lifetime of the n-period bond, up to a constant risk premium. This should hold regardless of the process followed by the short rate. Following Campbell and Shiller (1991) and others, we can test this implication with the following regression:

$$\frac{1}{n}\sum_{i=0}^{n-1} y_{t+i}^{(1)} - y_t^{(1)} = \alpha + \beta(y_t^{(n)} - y_t^{(1)}) + u_t, \quad (4)$$

where $y_t^{(n)}$ denotes the yield of an n-period bond at time t. The expectations hypothesis implies that when the yield spread between short-term and long-term bonds $(y_t^{(n)} - y_t^{(1)})$ is high, short-term bond yields will rise in the future (the dependent variable will be large). Specifically, the expectations hypothesis implies that $\beta = 1$. Early papers estimating equation (4) include Fama (1984) and Fama and Bliss (1987).

The first row in table 2 presents our estimates of β in equation (4) for bonds of maturity of two to 40 quarters. Consistent with a large earlier literature, we find that the null hypothesis of $\beta = 1$ is resoundingly rejected at short horizons. At short horizons, our estimates of β are close to zero. As the horizon grows, our estimate of β rises closer to 1 but remains below 1 for all horizons we consider.

Another implication of the expectations hypothesis of the term structure is that at times when the yield spread is unusually high, the yield on long bonds will rise. One intuition for this is that returns must be equalized (modulo a constant) for short-term and long-term bonds. If the yield spread is high, then the long-bond yield needs to rise to reduce the return on the long bond so that it can be equal to that of the short bond. Another intuition is that the high yield spread implies that the short yield will rise over the life of the long bond. As time passes, the relatively low current short rate will then drop out of the sum of future short rates that determines the long yield (according to the expectations hypothesis). As this happens, the sum increases and so the long yield should increase.

TABLE 2
FAILURES OF EXPECTATIONS HYPOTHESIS

	Long-Horizon n						
	2	3	4	8	12	20	40
Future short rates	−.01***	.11***	.18***	.39**	.57	.74	.71
	(.23)	(.23)	(.23)	(.23)	(.26)	(.23)	(.20)
Change in long rate	−1.02***	−.91***	−1.03***	−1.29***	−1.61***	−2.04***	−2.75***
	(.45)	(.59)	(.62)	(.59)	(.57)	(.55)	(.87)

NOTE.—The sample period is from 1961:3 to 2019:4. "Future short rates" reports estimates of β from regression (4). "Change in long rate" reports estimates of β from regression (5). Standard errors are in parentheses. Asterisks represent significance relative to the hypothesis that $\beta = 1$. p-values are computed using Newey-West standard errors with lag length selected as $L = \lceil 1.3 \times T^{1/2} \rceil$ and fixed b critical values, as proposed in Lazarus et al. (2018). This corresponds to a bandwidth of 19.
** $p < .05$.
*** $p < .01$.

We can test this implication of the expectations hypothesis with the following regression:

$$y_{t+1}^{(n-1)} - y_t^{(n)} = \alpha + \beta\left(\frac{1}{n-1}\right)(y_t^{(n)} - y_t^{(1)}) + u_t. \qquad (5)$$

It is straightforward to show that the expectations hypothesis implies $\beta = 1$. Early papers estimating equation (5) include Shiller (1979), Shiller, Campbell, and Schoenholtz (1983), and Campbell and Shiller (1991).

The second row of table 2 presents our estimates of β in equation (5). Consistent with earlier research, we find large deviations from the null of $\beta = 1$ implied by the expectations hypothesis. We estimate values for β around -1 at short horizons and even larger negative values at longer horizons. This means that when the yield spread is large, the long rate has tended to fall rather than rise, as the expectation hypothesis implies that it should. The conventional interpretation of this result is that it implies large predictable excess returns on the long bond when the yield spread is high.

The previous literature has identified a number of potential econometric issues associated with these tests of the expectations hypothesis. One issue is that in regression (5), the long-term yield appears in the dependent variable with a negative sign and in the regressor with a positive sign. As a consequence, measurement error in the long yield will bias the estimated coefficient downward and may even result in a negative estimate. Campbell and Shiller (1991) use instrumental variable techniques to assess whether measurement error is the cause of the negative estimates but find that the negative coefficients are quite robust. A second issue is small sample bias. This issue was emphasized for regressions (4) and (5) by

Bekaert, Hodrick, and Marshall (1997), who show that for these regressions, taking account of small sample bias strengthens the evidence against the null of $\beta = 1$. We conduct Monte Carlo analysis in section VI based on our model from section IV. This analysis does find evidence of some small sample biases. But the quantitative magnitude of these biases is small.

IV. Learning about Nominal Interest Rates

Traditional tests of forecast rationality evaluate the joint hypothesis that agents form conditional expectations rationally and that they know the true model that generates the data. Our goal is to assess whether we can explain the forecast anomalies documented in section III by relaxing the assumption that forecasters know the true model while maintaining the assumption of Bayesian updating. To this end, we consider agents who update their beliefs about how the world works using Bayesian learning and then form real-time Bayesian forecasts.

Our first application is to learning about the 3-month T-bill rate (short rate). We begin by presenting the model we assume the agents use to learn about and forecast the short rate. We then describe the details of how they learn and forecast. Finally, we compare the resulting forecasts with the SPF forecasts and longer-term yields.

A. *An Unobserved Components Model for the Nominal Short Rate*

Following Kozicki and Tinsley (2001), we propose a "shifting end point" model for the short rate.[9] Specifically, the model we assume agents use to learn about and forecast the short rate is

$$y_t = \mu_t + x_t, \tag{6}$$

$$\mu_t = \mu_{t-1} + \sqrt{\gamma}\sigma\eta_t, \eta_t \sim N(0, 1), \tag{7}$$

$$x_t = \rho x_{t-1} + \sqrt{1 - \gamma}\sigma\omega_t, \omega_t \sim N(0, 1). \tag{8}$$

Here, the short rate y_t is modeled as the sum of two unobserved components: a permanent random walk component μ_t and a transitory AR(1) component x_t. The transitory component x_t is assumed to have mean zero and persistence ρ. Shocks to μ_t and x_t are independent, normally distributed. The total variance of these two innovations to y_t conditional on time $t - 1$ information is σ^2. The share of the variance of

[9] See also van Dijk et al. (2014), Cieslak and Povala (2015), Bauer and Rudebusch (2020), Crump et al. (2021), and Bianchi, Lettau, and Ludvigson (2022).

these innovations that is attributable to shocks to the permanent component μ_t is assumed to be γ, with the complementary share $1 - \gamma$ attributable to the transitory component x_t. We refer to this as the unobserved components model.

To gain intuition about the model, consider the h-period forecast of the short rate, assuming that the unobserved components at time t and parameters of the model are known:

$$E_t y_{t+h} = \mu_t + \rho^h x_t. \tag{9}$$

This shows that μ_t corresponds to the long-run forecast of the short rate (as $h \to \infty$), while x_t captures short-run deviations of the short rate from this long-run forecast. The expectations hypothesis implies that the yield on an n-period zero coupon bond is

$$y_t^{(n)} = c^{(n)} + \frac{1}{n}\sum_{h=0}^{n-1} E_t y_{t+h} = c^{(n)} + \mu_t + \frac{1}{n}\sum_{h=0}^{n-1}\rho^h x_t, \tag{10}$$

where $c^{(n)}$ denotes the constant risk premium on n-period bonds. Using language from the term structure literature, we can say that μ_t represents a level factor for bond yields, while the slope and curvature of the term structure are governed by x_t.[10]

Our model for the short rate abstracts from stochastic volatility. We have extensively analyzed a version of the model with stochastic volatility ($\log \sigma^2$ following a random walk). This version of the model yields similar results to the baseline model, but the stochastic volatility adds substantial computational complexity.

B. Bayesian Learning and Forecasting about the Nominal Short Rate

We assume that agents do not know the value of the unobserved components (states) μ_t and x_t. We furthermore assume that they do not know the value of the parameters ρ, γ, and σ. We endow them with initial beliefs about these unknown states and parameters and data on the short rate. We assume that they use Bayes's law to update their beliefs about the states and parameters over time and then in each period construct forecasts of future short rates on the basis of their then current beliefs. More specifically, we start the agents off with initial beliefs in 1951:2. The agents then use data on the short rate from 1951:2 onward to update

[10] Nominal interest rates are nonstationary because inflation is nonstationary: over our sample period, nominal interest rates and inflation are cointegrated. That is, we fail to reject cointegration using the tests of Engle and Granger (1987) and Johansen (1991).

their beliefs. Starting in 1961:3, they perform online forecasting of the short rate. In other words, each quarter they forecast the short rate on the basis of their beliefs at that point in time.

The world did not begin in 1951:2. So, why do we not use data going further back in time? The reason for this is that the monetary policy regime in the United States was fundamentally different before 1951:2. In March 1951, the US Treasury and the Federal Reserve reached an agreement—commonly referred to as the Treasury-Fed Accord—to separate government debt management and monetary policy (Romero 2013). Between 1942 and the accord, the Federal Reserve abdicated its monetary independence by committing to fix the short rate at a low value to aid the financing of World War II and manage the massive government debt left after World War II. Before 1942, the United States had for the most part been on a gold (or silver) standard. Rather than model these fundamentally different monetary regimes explicitly, we start our analysis at the time of the Treasury-Fed Accord and simply endow agents with initial beliefs at that date (which presumably reflect information gleamed from the prior history).

We use a Gibbs sampling algorithm (augmented with random walk Metropolis-Hastings steps when needed) to sample from the posterior distribution of the model parameters and the latent states at each time period t. We describe this algorithm in more detail in appendix B (apps. A–J are available online). Armed with an estimate of agent's belief distribution for the unknown parameters and states in each time period t, we use our unobserved components model to construct Bayesian forecasts of the future evolution of the short rate—that is, we calculate the posterior predictive distribution of future short rates given beliefs at time t. We describe the algorithm we use to do this in appendix C. We do this for each quarter starting in 1961:3, which is the first quarter for which we have zero coupon yield curve data.

An advantage of the fact that agents in our model are Bayesian is that it does not matter how we write our model. For example, our unobserved components model has an autoregressive integrated moving average (ARIMA) representation. The Bayesian agents in our model see through the superficial difference between the unobserved components and ARIMA representation of our model. Whether we write the model one way or the other therefore does not matter for our results (something that is not true in the case of boundedly rational learning).

We assume that agents make their forecasts on the final day of each quarter. This implies that they have access to the average level of the short rate in that quarter and their nowcast is the true realized interest rate for the quarter. This is an approximation: in reality, the SPF forecasters have information only up to the second to third week of the second month of the quarter, as we discuss above.

The short rate was constrained by the zero lower bound toward the end of our sample period. We define the period when the target federal funds rate was at or below 25 basis points as the zero lower bound period. This corresponds to 2009:1 to 2015:4 in our sample period. We view this as a period when the desired short rate is censored (but for simplicity follows the same process as before). Our approximation to Bayesian learning for this period is to assume that agents do not update their beliefs about the parameters (ρ, γ, and σ) but that they continue to filter the hidden states (μ_t and x_t) using the parameter estimates from 2008:4. Full learning then resumes in 2016:1. This shortcut allows us to avoid substantial additional complications, which we believe are unlikely to materially affect our results.[11] Our results are very similar if we end the sample in 2008:4.

C. *Initial Beliefs about the Nominal Short Rate*

If learning is fast, beliefs converge quickly to the truth and initial beliefs quickly cease to matter. If learning is slow, beliefs will not converge quickly to the truth and initial beliefs will continue to influence later beliefs nontrivially for a long time—as long as it takes for beliefs to converge to the truth. In our setting, learning about the parameters ρ, γ, and σ is slow, while learning about the states μ_t and x_t is reasonably fast. Our choice of initial beliefs about μ_t and x_t therefore does not matter for our results as long as they are reasonable. (Recall that there is a 10-year burn-in period from 1951:2 to 1961:3.) We assume that initial beliefs about μ_t in 1951:2 are $N(y_{1951:12}, 1)$ and initial beliefs about x_t in 1951:2 are $N(0, 1)$. These initial conditions are assumed to have a correlation of -1 because of the form of the observation equation (6).

For ρ, γ, and σ, we specify initial beliefs in 1951:2 of the following form:

$$\rho \sim N(\mu_\rho, \sigma_\rho^2), \gamma \sim \mathcal{B}(\alpha_\gamma, \beta_\gamma), \sigma^2 \sim \mathcal{IG}(\alpha_{\sigma^2}, \beta_{\sigma^2}),$$

where $\mathcal{B}$ denotes a beta distribution and $\mathcal{IG}$ denotes an inverse gamma distribution. As we discuss above, these initial beliefs encode professional forecaster's understanding of how the world works as of 1951:2, on the basis of prior history. We search over the space of initial beliefs specified above for the initial beliefs that can best rationalize the forecast anomalies we document in section III. If we can find a belief (or perhaps a set of beliefs) that can rationalize the forecast anomalies, then we ask whether any of these beliefs can be viewed as a reasonable initial belief for professional

[11] A fuller treatment would explicitly allow for censoring of the desired short rate. This would require us to shift to nonlinear sampling methods and would thus increase run times by an order of magnitude. Intuitively, however, the information learned about ρ, γ, and σ during this period would likely be limited since the desired short rate is censored.

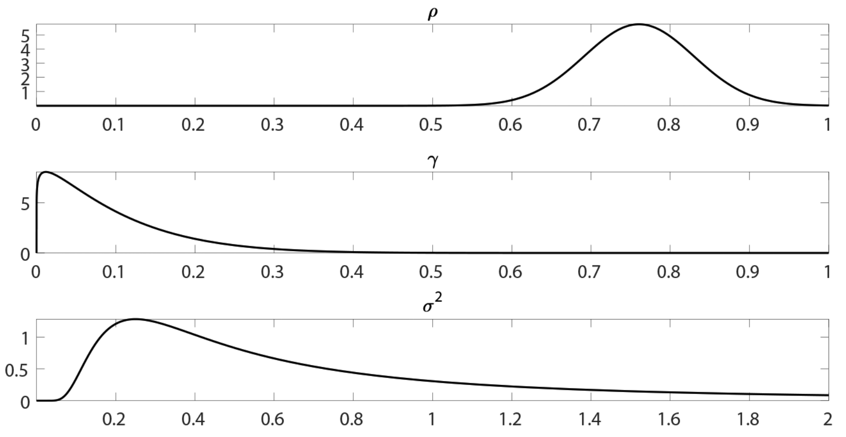

FIG. 4.—Marginal initial beliefs distributions: T-bill rate model. Each panel plots the initial beliefs held in 1951:2 by agents in our T-bill rate model for each of the three model parameters: ρ, γ, and σ^2.

forecasters to have in 1951:2. If so, we conclude that the forecast anomalies we have documented can be explained by Bayesian learning and are therefore not necessarily evidence of forecaster irrationality.[12]

To keep our analysis manageable, we fix the initial beliefs for σ by setting $\alpha_{\sigma^2} = 1.25$ and $\beta_{\sigma^2} = 0.5625$. This belief distribution is plotted in the bottom panel of figure 4.[13] This leaves four parameters: μ_ρ, σ^2_ρ, α_γ, β_γ. We search over the space of these parameters to find beliefs that match the forecast anomalies as well as possible. Specifically, for each point in this space, we construct forecasts as described above and estimate the forecasting regressions discussed in section III. We then minimize an unweighted average of the square of the difference between the regression coefficients from the regressions based on model-generated forecasts and the regression coefficients we estimated in section III based on real-world data. To focus on the subspace of reasonable initial beliefs, we constrain the mean of the prior for ρ, μ_ρ, to be larger than 0.5. Appendix D provides more detail.

[12] In app. E, we present an alternative set of results where—rather than targeting the anomalies we document in sec. III—we directly target the time series of consensus T-bill forecasts from the SPF at horizons 1–4 and also the 5- and 10-year zero coupon nominal yields from the Liu and Wu (2021) data. This yields very similar results to our baseline results report below. The main difference is that the initial belief we estimate for ρ is concentrated on smaller values and the initial belief for γ has a somewhat smaller standard deviation. As a result, the interest rate forecasts have slightly more slope. The fit to the anomalies and deviations from the expectations hypothesis is quite similar.

[13] This belief distribution has a mode of 0.25. The standard deviation of the distribution is undefined for values of $\alpha_{\sigma^2} \leq 2$. Our choice of $\alpha_{\sigma^2} = 1.25$ is thus a very dispersed initial belief.

The top and middle panels of figure 4 plot the initial belief distributions for ρ and γ that minimize the objective function discussed above. The belief distribution for ρ is concentrated on moderately large values. It is centered at 0.76 and has a standard deviation of 0.07. With a $\rho = 0.76$, the half-life of innovations to x_t is roughly 8 months. The belief distribution for γ is concentrated on relatively small values. It has a mean of 0.09 and a standard deviation of 0.08. This implies that forecasters believed in 1951:2 that most of the variation in the short rate was due to transitory fluctuations of moderate persistence (i.e., an x_t with a ρ around 0.76) rather than permanent fluctuations (μ_t).

Are the initial belief distributions plotted in figure 4 reasonable? We argue they are for two reasons. First, they are quite dispersed, that is, they put substantial mass on a wide range of parameter combinations, a sufficiently wide range that we think they constitute plausible beliefs forecasters might have had in 1951:2. Second, the belief that γ was relatively small is arguably consistent with the history of interest rates prior to 1951:2. Outside of war, the United States had been on a gold (or silver) standard almost continuously from its founding, and the United Kingdom had been on a gold (or bimetallic) standard for hundreds of years before that. Over this long time span, risk-free interest rates had been quite stable at low frequencies, with most variation being rather transient (because of seasonal cycles and financial crises).[14]

This can be seen clearly in figure 5, which plots the yield on UK consoles from 1727 to 2016.[15] The UK console rate is arguably the best proxy of a long-term risk-free rate prior to the twentieth century. The figure shows that this rate fluctuated very little prior to the start of our sample period (marked by the vertical line in the figure), never rising above 6%. Even extreme events, such as the Napoleanic Wars and World War I—both cases when the United Kingdom suspended convertibility to gold—did not cause large swings in long-term rates. In other words, throughout this period, investors believed that any sizable short-term fluctuations in interest rates would be short-lived.[16]

The beliefs we estimate for forecasters in 1951:2 line up well with this history in that they put a substantial amount of weight on the notion that most fluctuations in interest rates were relatively transient. Figure 5 shows that the long upward march of interest rates in the 1960s, 1970s, and early 1980s and subsequent downward march since then was completely without

[14] Mankiw and Miron (1986) show that violations of the expectations hypothesis of the term structure were smaller before the founding of the Federal Reserve, when there was sizable seasonality in interest rates (which was presumably relatively easy to learn).

[15] Data on secondary market yields on UK consoles are first available in 1727.

[16] Consistent with this, Payne et al. (2022) show that long-run inflation expectations in the United States were anchored during much of the nineteenth century (even during the Civil War greenback devaluation period).

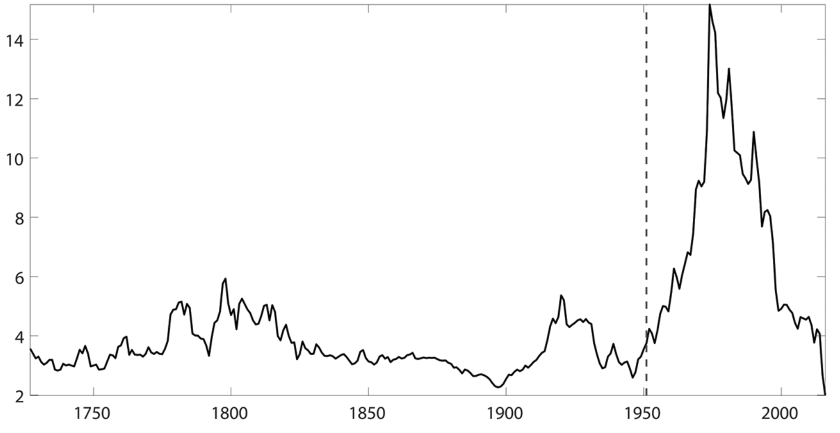

FIG. 5.—UK console rate. The figure plots the yield on UK consoles from 1727 to 2016. The vertical dashed line is 1951, the year our sample starts.

parallel in history. As Homer and Sylla (2005, 1) put it in their *A History of Interest Rates*, "A long view, provided by this history, shows that recent peak yields were far above the highest prime long-term rates reported in the United States since 1800, in England since 1700, or in Holland since 1600. In other words, since modern capital markets came into existence, there have never been such high long-term rates." It seems unlikely that forecasters in 1951:2 would put much weight on this unprecedented multidecade run-up and fall in interest rates occurring.

D. *Model's Fit to the Data*

Figure 6 offers a visual depiction of the fit of the model's forecasts to the data. The top panel plots SPF forecasts of the short rate (the same data that are plotted in fig. 1). The bottom panel plots the forecasts generated by our model with the initial beliefs discussed above. Our model captures the fact that SPF forecasters tend to predict that the short rate will mean revert slowly toward a normal value that is shifting over time, that is, something close to the average value of the short rate over the past business cycle. For example, in the easing cycle of 1985–87, SPF forecasters consistently expect the short rate to rise. This leads them to be wrong in their forecast in the same direction over and over again. The same is true for agents in our model. This pattern repeats in later easing cycles, such as 1991–93 and 2001–3. When rates are rising, SPF forecasters expect them to rise more slowly than they actually do. This occurs in 1988–89, 1994, and 1999–2000 and leads to highly autocorrelated forecast errors. Our model matches this pattern.

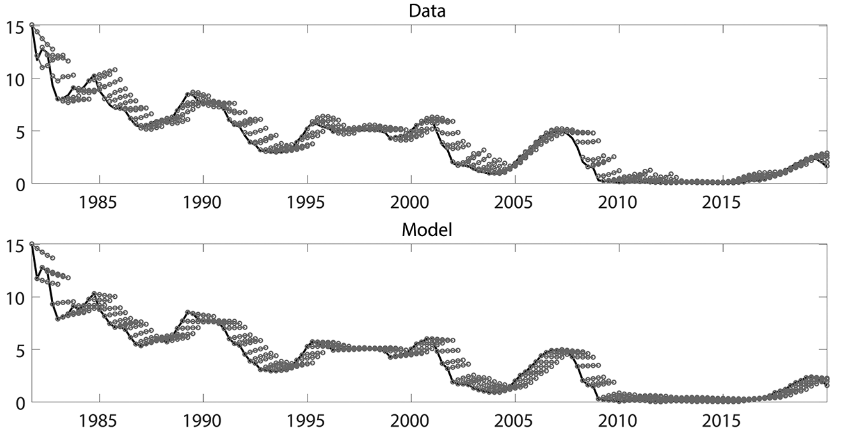

FIG. 6.—Forecasted T-bill rate: data versus model. The solid black line is the 3-month T-bill rate. Each short gray line with five circles represents forecasts made in a particular quarter about the then present quarter (first circle) and following four quarters (subsequent four circles). In the top panel, these forecasts are SPF forecasts. In the bottom panel, these forecasts are mean forecasts generated from the unobserved components model estimated in real time.

More recently, the increasing use of forward guidance has led SPF forecasts to diverge from what our model predict on occasion. A prominent example of this is the period 2012–15, when the Fed explicitly stated that they would keep the short rate at 0.25% for several years. Our model does not incorporate this forward guidance and therefore fails to capture its effect on SPF forecasts. Something similar occurs in 2004–7 and 2018, when the Fed used forward guidance to inform the market about the speed of tightening.

Table 3 presents results for the forecast anomaly regressions we analyze in section III for our model-generated data (“UC model”) and compares these with analogous results for the real-world data (“SPF”). Despite our model having very few parameters, we are able to match almost all the anomalies we have emphasized. For all four types of regressions and at all horizons, our model matches the magnitude and statistical significance of the real-world estimates quite closely. Specifically, our model generates a negative bias that increases in size with the horizon, as in the data; autocorrelation in forecast errors of about 0.35 at horizons 1–3 and much less at horizon 4, as in the data; Mincer-Zarnowitz coefficients slightly below 1 and decreasing with the horizon, as in the data; and underreaction that grows with the horizon, as in the data.

Table 4 presents results for the expectations hypothesis regressions we discuss in section III based on model-generated data and compares these results with those based on real-world data. Again, our model matches the

TABLE 3
T-Bill Rate Forecast Anomalies: Model versus Data

	Forecast Horizon			
	1	2	3	4
	A. Bias			
SPF	−.18***	−.34***	−.52***	−.70***
	(.05)	(.09)	(.14)	(.19)
UC model	−.15**	−.27**	−.40**	−.51**
	(.06)	(.11)	(.16)	(.21)
	B. Autocorrelation			
SPF	.30*	.27**	.24*	.13
	(.14)	(.12)	(.12)	(.13)
UC model	.36*	.39**	.35**	.23*
	(.17)	(.14)	(.11)	(.12)
	C. Mincer-Zarnowitz			
SPF	.97*	.94**	.90**	.86**
	(.02)	(.02)	(.04)	(.05)
UC model	.96*	.93**	.88**	.84***
	(.02)	(.03)	(.04)	(.05)
	D. Coibion-Gorodnichenko			
SPF	.23*	.34*	.62***	. . .
	(.12)	(.16)	(.16)	
UC model	.39*	.56	.89*	. . .
	(.18)	(.37)	(.42)	

Note.—The forecast horizons are quarters. Standard errors are in parentheses. Asterisks represent significance relative to the following hypotheses: $\alpha = 0$ for bias, $\beta = 0$ for autocorrelation, $\beta = 1$ for Mincer-Zarnowitz, and $\beta = 0$ for Coibion-Gorodnichenko. p-values are computed using Newey-West standard errors with lag length selected as $L = \lceil 1.3 \times T^{1/2} \rceil$ and fixed b critical values, as proposed in Lazarus et al. (2018). This corresponds to a bandwidth of 17.

* $p < .10$.
** $p < .05$.
*** $p < .01$.

real-world anomalies both qualitatively and quantitatively. For the future short rate regressions in panel A, we estimate β coefficients close to zero at short horizons, as in the data. The estimates then rise for longer-term bonds as they do for the data. For the change in long rate regressions in panel B, we estimate β coefficients that are negative at all horizons and increasingly so as the horizon increases. Quantitatively, our estimates are close to -1 at short horizons and decrease to -2.5 at long horizons. These patterns are quite consistent with those in the real-world data.

Table 4 shows that our model provides an explanation for why the long rate has tended to fall when the yield spread is large rather then rise as full-information rational expectations models predict. In our model, this arises from learning. When the yield spread is large, agents in our model

TABLE 4
FAILURES OF EXPECTATIONS HYPOTHESIS: MODEL VERSUS DATA

	Long-Horizon n						
	2	3	4	8	12	20	40
	A. Future Short Rates						
Data	−.01***	.11***	.18***	.39**	.57	.74	.71
	(.23)	(.23)	(.23)	(.23)	(.26)	(.23)	(.20)
UC model	−.11***	.08**	.17**	.56	.81	.93	.99
	(.32)	(.32)	(.33)	(.38)	(.37)	(.31)	(.36)
	B. Change in Long Rate						
Data	−1.02***	−.91***	−1.03***	−1.29***	−1.61***	−2.04***	−2.75***
	(.45)	(.59)	(.62)	(.59)	(.57)	(.55)	(.87)
UC model	−1.21***	−1.25***	−1.28***	−1.40***	−1.54***	−1.84***	−2.55**
	(.63)	(.64)	(.65)	(.70)	(.76)	(.88)	(1.52)

NOTE.—The sample period is from 1961:3 to 2019:4. Panel A reports estimates of β from regression (4). Panel B reports estimates of β from regression (5). Standard errors are in parentheses. Asterisks represent significance relative to the hypothesis that $\beta = 1$. p-values are computed using Newey-West standard errors with lag length selected as $L = \lceil 1.3 \times T^{1/2} \rceil$ and fixed b critical values, as proposed in Lazarus et al. (2018). This corresponds to a bandwidth of 19.
** $p < .05$.
*** $p < .01$.

tend to revise their estimate of the long-run level of the short rate (μ_t) downward by enough to offset the forces embedded in full-information rational expectations models.

E. Parameter and State Estimates

Figure 7 plots the evolution of the mean of the posterior distributions of ρ, γ, and σ along with 90% credible intervals between 1961:3 and 2019:4. Relative to the initial beliefs presented in figure 4, agents' estimates of ρ rise noticeably. The mean estimate of ρ is around 0.8 early in the sample as compared with about 0.76 for the initial beliefs (in 1951:2). It then gradually rises further over the sample and is around 0.9 toward the end of the sample. Agents also revise their beliefs about γ upward relative to the initial beliefs. The mean estimate of γ hovers between 0.1 and 0.2 for most of the sample. In both cases, agents are revising their beliefs in the direction of believing that interest rate fluctuations are more persistent. The mean estimate of σ is around 0.4 early in the sample. It rises sharply during the Volcker disinflation and gradually decreases after the early 1980s.

Figure 8 plots the mean estimate of μ_t over the course of the sample. The solid black line is the mean of the real-time filtering distribution, that is, the belief distribution about μ_t conditional on data up to time t, while the solid gray line is the mean of the ex post smoothing distributions, that is, the belief distribution about μ_t conditional on data up to

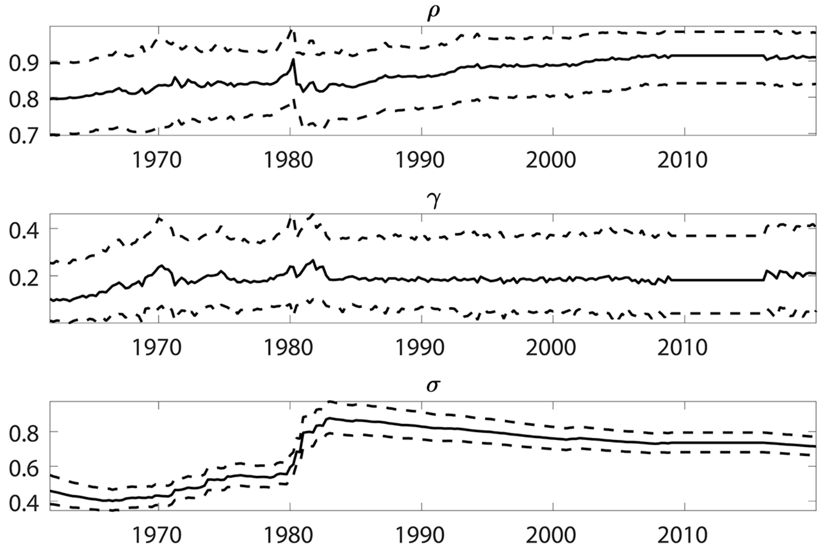

FIG. 7.—Parameter estimates: T-bill rate model. Each panel plots the evolution of beliefs about one of the three unobserved components model parameters: ρ, γ, and σ. The solid black line is the mean, and the dashed black lines are the 5th and 95th percentiles of the posterior distribution for the parameter in question. Recall that we update beliefs about these parameters only every fourth quarter.

2019:4. The dashed black lines plot 90% credible intervals for the real-time filtering distribution.

It is interesting to compare the real-time filtering distribution and the ex post smoothing distribution in figure 8. The real-time filtering distribution is consistently below the ex post smoothing distribution from the beginning of our sample until the early 1980s and then consistently above from the early 1980s until very late in our sample. This reflects the fact that in real time the agents in our model underestimate the persistence of the run-up of interest rates in the 1960s and 1970s and again underestimate the persistence of the fall in interest rates after the early 1980s. Ex post, agents revise their view of history and conclude that both the run-up and fall in interest rates were more persistent than they believed at the time. This helps explain the persistent downward drift of long rates in the 1980s at a time when the yield spread was high.

F. Allowing for a Break in 1990

Much recent work on the term structure of interest rates restricts attention to data after 1990 because of a break in the behavior of the term structure around 1990. A possible reason for such a break is that bond

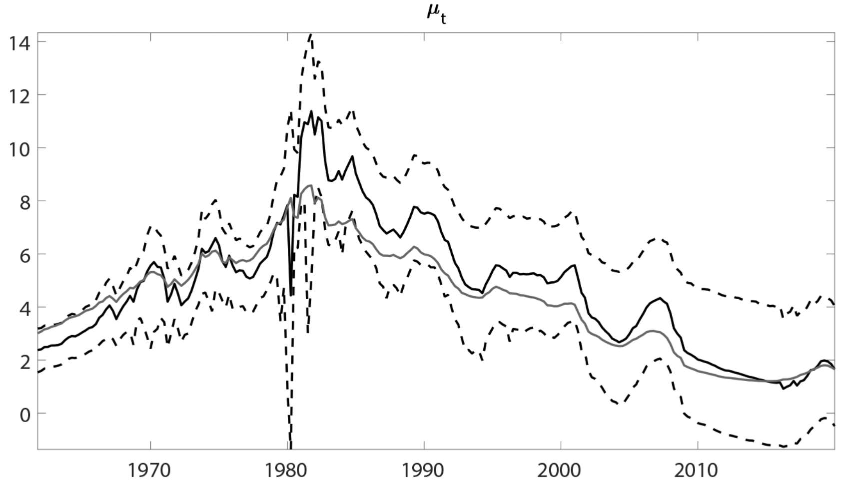

FIG. 8.—State estimates: T-bill rate model. The figure plots the evolution of beliefs about the permanent component μ_t. The solid black line is the posterior mean of the real-time filtering distributions, the dashed black lines are the 5th and 95th percentiles of the posterior real-time filtering distributions, and the gray line is the posterior mean of the ex post smoothing distributions.

market traders at some point became convinced that the change in monetary policy implemented by Paul Volcker and carried on by Alan Greenspan—which focused monetary policy on maintaining low and stable inflation—was likely to be permanent.

In our baseline model, we do not allow forecasters to learn about the process that the short rate follows from any other source than past data on the short rate itself. In reality, it is likely that forecasters' views are to some extent shaped by other sources of information. In particular, it seems likely that the relentless rhetorical focus of Federal Reserve officials in the 1980s on their commitment to keep inflation low going forward may have affected the views of bond market traders and forecasters about the future evolution of short-term interest rates.

In our model, this amounts to forecasters and the bond market becoming convinced that γ was likely to be smaller going forward than in the past. To capture this, we now consider a case where we allow for a break in γ in 1990. Specifically, we redo our baseline short rate analysis exactly as before except that we allow the agents in the model to reset their beliefs about γ in 1990. We assume that the new belief distribution of agents about γ in 1990 is $\gamma \sim \mathcal{B}(\alpha_{\gamma,2}, \beta_{\gamma,2})$, and we search over the values of $\alpha_{\gamma,2}$ and $\beta_{\gamma,2}$ as well as the hyperparameters in the baseline case to best match the forecast anomalies.

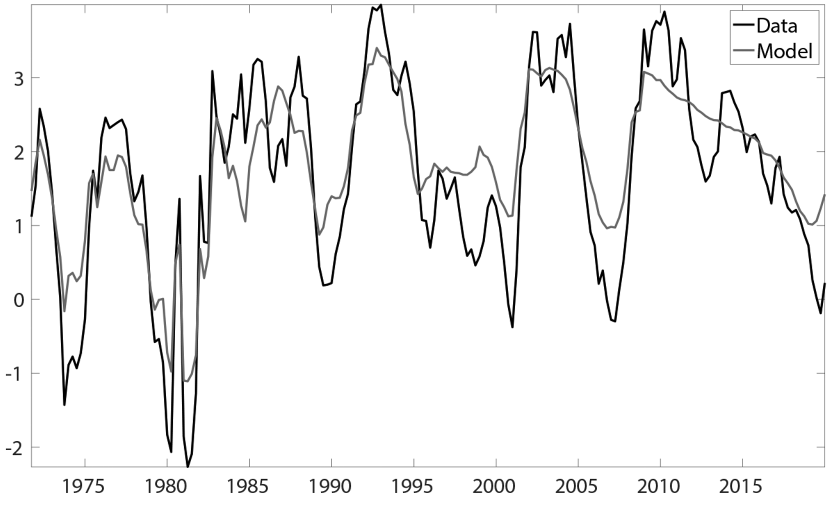

FIG. 9.—Yield spread in data and model. The figure plots the spread between the yield on a 10-year zero coupon bond and the 3-month T-bill rate for the data and the model.

We find that beliefs about γ do indeed shift down in 1990: the mean of the distribution of γ shifts from 0.19 to 0.11. In addition, the belief distribution becomes much more concentrated on low values. The standard deviation of the belief distribution falls from 0.09 to 0.03. The fit of the model to the forecast anomalies and expectations hypothesis regressions we focus on above improves somewhat but is fairly similar to the baseline case. However, the model with this break allows us to match several additional features of the term structure quite well. (The full results for this case are presented in app. G.)

Figure 9 plots the yield spread between a 10-year zero coupon bond and the 3-month T-bill rate in the data and in the model. We see that the model is able to match quite well the many ups and downs of the yield spread over this 50-year period. The main way in which the model-implied spread differs from the spread in the data is that it is slightly less volatile.

Cochrane and Piazzesi (2005) present even more spectacular evidence of return predictability than earlier work by Fama and Bliss (1987), Campbell and Shiller (1991), and others. They show that a single factor predicts 1-year excess returns on 2–5-year maturity bonds with an R^2 in excess of 0.4. We estimate a return predictability factor using the procedure of Cochrane and Piazzesi on data from our model with the break in 1990. Our model can match the high R^2 for 1-year excess returns on 2–5-year zero coupon bonds observed in the data: the R^2 for these predictive regressions on data from our model are between 0.46 and 0.50.

We find it quite plausible that ρ, γ, and σ have in fact undergone a number of structural breaks over our sample period and will do so again in the future. A likely benefit of incorporating further parameter breaks into our model would be to further perpetuate learning. In our model, agents eventually learn the parameters and the anomalies disappear, although this takes many, many decades. In a model where the parameters undergo occasional breaks, learning would continue for much longer, potentially many centuries.

G. Alternative Initial Beliefs

In appendix F, we present results for two alternative assumptions about the initial beliefs of agents in our model. The first of these is a case where agents have much more dispersed initial beliefs about both ρ and γ. The second case is one where agents have "look-ahead" initial beliefs, that is, their initial beliefs are set to approximate the beliefs agents have at the end of our sample in our baseline analysis. In both of these cases, agents produce forecasts that are closer to the forecasts from a random walk model (i.e., no change) than in our baseline case. This is particularly the case when agents start with very dispersed initial beliefs. In both cases, this leads to a moderate deterioration of the fit to the forecasting anomalies and a much more dramatic deterioration of the fit to the expectations hypothesis statistics and the yield spread.

We also consider whether more dispersed initial beliefs yield better forecasts in the sense of lower RMSEs. More dispersed initial beliefs yield ever so slightly smaller RMSEs at short horizons but slightly larger RMSEs at longer horizons. Averaging over horizons, the two cases are virtually identical in terms of RMSEs. Both cases yield RMSEs that are about 3% smaller than the SPF forecasts.

V. Learning about the Real GDP Growth

Our second application is to learning about real GDP growth. As in section IV, we begin by presenting the model we assume agents use to learn about and forecast GDP growth. We then describe the details of how they learn and forecast. Finally, we compare the resulting forecasts with the CBO forecasts we discussed in section III.

A. An Unobserved Components Model for GDP

A key issue for GDP forecasting has to do with the extent to which fluctuations in GDP are trend stationary versus difference stationary. The model we assume agents use to learn about and forecast real GDP allows for both trend stationary and difference stationary shocks:

$$y_t = z_t + x_t, \tag{11}$$

$$\Delta z_t = \mu + \sqrt{\gamma}\sigma u_t,\ u_t \sim N(0,1), \tag{12}$$

$$x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + \sqrt{1-\gamma}\sigma v_t,\ v_t \sim N(0,1), \tag{13}$$

where y_t denotes quarterly log real GDP, z_t is a difference stationary component, and x_t is a trend stationary component x_t. The difference stationary component z_t is assumed to follow a random walk with drift μ. The trend stationary component x_t is assumed to follow a mean zero AR(2) process with autoregressive coefficients ρ_1 and ρ_2. The conditional standard deviation of y_t is denoted σ. The share of innovations to y_t that hit the difference stationary component z_t is γ, with the complementary share $1 - \gamma$ hitting the trend stationary component x_t. The parameter γ therefore governs how big the random walk component of GDP is (Cochrane 1988). We refer to this model as an unobserved components model. This model is slightly more complicated than our model for interest rates. It has two extra parameters: μ to allow for a trend and ρ_2 to allow for hump-shaped dynamics.

B. *Bayesian Learning and Forecasting about GDP*

As in the interest rate application discussed in section IV, we assume that agents in the model do not know the value of the unobserved components (states) z_t and x_t or parameters μ, ρ_1, ρ_2, σ, and γ. We start the agents off with an initial belief distribution about these unknown states and parameters in 1959:3. This is the first date for which we have a full set of real-time GDP vintages with which to do our analysis. The agents then observe (real-time) data on GDP and update their beliefs about the states and parameters using Bayes's law. Below we plot results starting in 1976:1. This corresponds to the first period for which CBO forecasts are available.

We assume that agents have access to the first release of fourth-quarter GDP for the prior year (the advance release for that quarter from the Bureau of Labor Statistics) when they forecast. This is meant to approximate the information set the CBO has access to when it forecasts GDP each year. The CBO's forecasts (contained in its economic outlook report) are typically released in January or February of each year. While this is usually before the advance release of fourth-quarter GDP for the previous year, much of the underlying data that are used to construct the advance release have been made public at this point. This implies that the fourth-quarter advance release can be predicted fairly accurately. We therefore think that endowing our model agents with the fourth-quarter advance release is the best way to approximate the information set of the CBO at the time it constructs its annual GDP forecast.

The parameters of the model and latent state estimates are updated every four quarters to line up with the timing of when the CBO constructs forecasts. We describe the algorithm we use to update agent's beliefs in appendix H. Armed with estimates of agent's beliefs, we use our unobserved components model to construct forecasts of GDP growth. We describe the algorithm we use to do this in appendix I.

C. *Initial Beliefs about GDP*

As in the interest rate application in section IV, learning about the parameters in our model for GDP is slow, and agents' initial beliefs about the parameters matters. In contrast, learning about the states z_t and x_t is reasonably fast, implying that initial beliefs about these states do not affect our results. (In this case, there is a roughly 15-year burn-in period from 1959:3 to 1976.) We assume that agents' initial beliefs about z_t and x_t in 1959:3 are $z_t \sim N(y_{1959:3}, 0.01^2)$ and $x_t \sim N(0, 0.01^2)$.

We specify initial beliefs for the parameters in 1959:3 of the following form:

$$\rho_1 + \rho_2 \sim N(\mu_\rho, \sigma_\rho^2)\mathcal{I}(\rho_1, \rho_2), \rho_2 \sim N(\mu_{\rho_2}, \sigma_{\rho_2}^2)\mathcal{I}(\rho_1, \rho_2),$$
$$\gamma \sim \mathcal{B}(\alpha_\gamma, \beta_\gamma), \mu \sim N(\mu_\mu, \sigma_\mu^2), \sigma \sim \mathcal{IG}(\alpha_\sigma, \beta_\sigma),$$

where $\mathcal{I}(\rho_1, \rho_2)$ is an indicator function that is 1 if the x_t process is stationary and 0 otherwise. For more detail, see appendix J.

We fix $\mu_\mu = 0.01$ and $\sigma_\mu = 0.01$, corresponding to an initial belief for average annual long-run growth of 4%. We fix $\alpha_\sigma = 7.0625$ and $\beta_\sigma = 0.0014$, corresponding to a mean initial belief for σ^2 of 0.015^2 and standard deviation of 0.01. That leaves six parameters to estimate to fit the forecast anomalies presented in section II, which we denote $\boldsymbol{\theta} = (\mu_\rho, \sigma_\rho, \mu_{\rho_2}, \sigma_{\rho_2}, \alpha_\gamma, \beta_\gamma)'$. We do this in a similar fashion to what we do in the interest rate application in section IV. Appendix J provides details.[17]

The resulting initial beliefs are plotted in figure 10. We view these as reasonable initial beliefs in that they are quite dispersed. For example, the initial belief distribution on $\rho_1 + \rho_2$ puts substantial weight on values between 0.7 and 1. This range spans cases were the transitory component x_t has a

[17] We place some bounds on the values of parameters that can be chosen in this estimation. Namely, we restrict the standard deviation of the initial beliefs on $\rho_1 + \rho_2$, ρ_2, and γ to be greater than or equal to 0.05. For the initial belief distribution for γ, we additionally put an upper bound on the standard deviation of 0.15 and restrict the mode of the distribution to be less than 0.6. The latter restriction imposes that agents believe that at least 40% of the variation in output comes from trend stationary fluctuations. These restrictions are useful to guarantee dispersed initial beliefs and to generate initial beliefs where a significant fraction of output fluctuations are trend stationary.

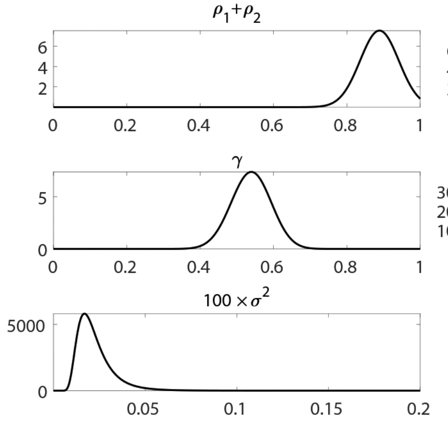

Fig. 10.—Marginal initial beliefs distributions: real GDP growth model. Each panel plots the initial beliefs held in 1959:3 by agents in our model for the following five parameter combinations: $\rho_1 + \rho_2$, ρ_2, γ, μ, and σ^2.

modest half-life of less than a year and cases where it is very persistent. Likewise, the initial belief for γ is centered close to 0.5 and has high variance. The initial belief for ρ_2 embeds a belief that the transitory component of GDP is hump shaped. But again, this distribution has substantial variance.

D. *Model's Fit to the Data*

Figure 11 offers a visual depiction of the fit of the forecasts that our model generates to the data. The top panel plots CBO forecasts of real GDP growth (the same data that are plotted in fig. 3). The bottom panel plots the forecasts generated by our model with the initial beliefs discussed above. The model is able to match the broad characteristics of CBO forecast errors. For example, the model matches the large forecast errors the CBO made in the early 2010s when it forecast that the economy would grow unusually fast after the Great Recession but growth turned out to be more modest. Also, the model generates persistent forecast errors in the late 1990s when growth was high for several years but the CBO persistently forecast lower growth.

Table 5 presents results for the forecast anomaly regressions we analyze in section III for our model-generated data ("UC model") and compares these with analogous results for the real-world data ("CBO"). Our model is able to match the anomalies in the CBO forecasts quite well. The most spectacular anomaly in the case of the CBO forecasts is for the Mincer-Zarnowitz regressions in panel C. These start off close to 1 at the 1-year horizon but fall to zero at the 3-year horizon and to roughly -0.4

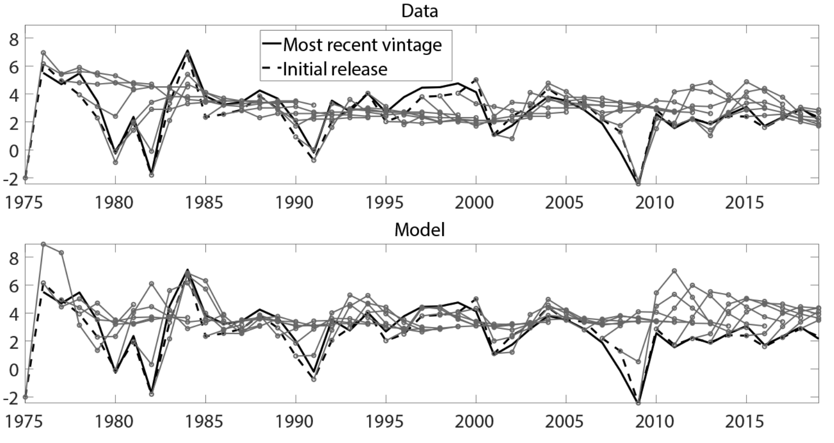

FIG. 11.—Forecast whisker plots: real economic output growth. The solid black line is the most recent vintage of GDP growth estimates. The dashed black line is the initial release of GDP growth for each period. Each short gray line with seven circles represents forecasts made in a particular year about that year (first circle) and the following 6 years (subsequent six circles). In the top panel, these forecasts are CBO forecasts. In the bottom panel, these forecasts are mean forecasts generated from the unobserved components model estimated in real time.

at the 4- and 5-year horizons. Our model is able to match this pattern quite well. The model also yields positively autocorrelated forecast errors, overreaction at long horizons in the Coibion-Gorodnichenko regression, and negative bias. For almost all of the anomaly statistics, the model estimate is not statistically significantly different from the data estimate, though the exact numerical fit is not as impressive as in our interest rate application.

E. Parameter Estimates

Figure 12 plots the evolution of the mean of the posterior distributions of the five parameters of our model for GDP along with 90% credible intervals over the period 1976 and 2019. Perhaps the most striking feature of figure 12 is how little beliefs about the parameters change over time. We do see that σ trends downward by a modest amount, likely reflecting the Great Moderation. Also, ρ_2 trends modestly upward. But ρ_1, γ, and μ change very little. This lack of change reflects a combination of at least two things. First, it may be that some of the parameters are close to their true values. Second, for those parameters that are further away from their true values, little information can be gleaned from the data about their true values resulting in posterior beliefs being little changed even over a 40-year period. This is perhaps not surprising given how difficult it

TABLE 5
Real GDP Forecast Anomalies: Model versus Data

	Forecast Horizon				
	1	2	3	4	5
	A. Bias				
CBO	.27	−.27	−.54	−.62	−.52
	(.25)	(.35)	(.50)	(.53)	(.49)
UC model	−.65	−1.65**	−1.36**	−.85	−.66
	(.32)	(.45)	(.45)	(.42)	(.40)
	B. Autocorrelation				
CBO	.22	.16	.11	.08	.08
	(.12)	(.14)	(.13)	(.18)	(.10)
UC model	.39*	.31	.23*	.06	−.05
	(.17)	(.16)	(.10)	(.10)	(.05)
	C. Mincer-Zarnowitz				
CBO	.94	.60	.03**	−.42***	−.43***
	(.10)	(.38)	(.27)	(.18)	(.29)
UC model	.84	.35**	.34*	−.38***	−.98**
	(.11)	(.17)	(.31)	(.19)	(.53)
	D. Coibion-Gorodnichenko				
CBO	.08	.00	.50	−1.63**	−1.46**
	(.08)	(.28)	(.58)	(.36)	(.40)
UC model	.06	−.76	−.11	−.78	−1.22**
	(.09)	(.44)	(.26)	(.39)	(.38)

Note.—The forecast horizons are years. Standard errors are in parentheses. Asterisks represent significance relative to the following hypotheses: $\alpha = 0$ for bias, $\beta = 0$ for autocorrelation, $\beta = 1$ for Mincer-Zarnowitz, and $\beta = 0$ for Coibion-Gorodnichenko. p-values are computed using Newey-West standard errors with lag length selected as $L = \lceil 1.3 \times T^{1/2} \rceil$ and fixed b critical values, as proposed in Lazarus et al. (2018). This corresponds to a bandwidth of 9.

* $p < .10$.

** $p < .05$.

*** $p < .01$.

is to distinguish between difference stationary time series and persistent but trend stationary time series.

VI. Why Does It Work?

To understand better why it is that our Bayesian learning model can match the forecast anomalies that we document in section III, we now simulate data from the model we use in section IV and assess how learning occurs in this model. Relative to the analysis earlier in the paper, in this section, we know the true data-generating process. We can therefore assess how long it takes agents to learn the truth and how initial beliefs that differ in various ways from the truth affect results from the forecasting regressions we consider in section III.

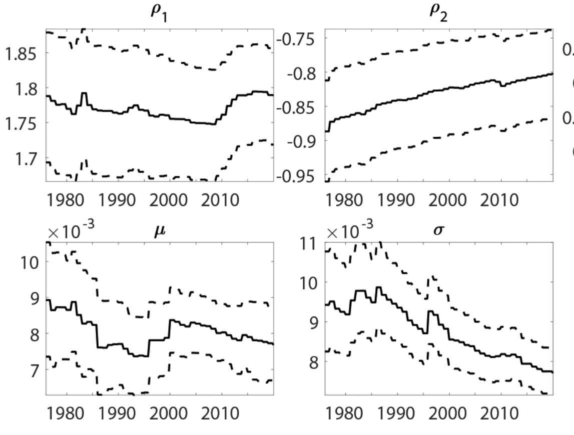

FIG. 12.—Parameter estimates: real economic output growth. Each panel plots the evolution of beliefs about one of the five unobserved components model parameters: ρ_1, ρ_2, γ, μ, and σ. The solid line is the mean, and the dashed lines are the 5th and 95th percentiles of the posterior distribution for the parameter in question. Recall that we update beliefs about these parameters only every fourth quarter.

Recall that the model we use for the short rate in section IV is

$$y_t = \mu_t + x_t, \tag{14}$$

$$\mu_t = \mu_{t-1} + \sqrt{\gamma}\sigma\eta_t, \eta_t \sim N(0,1), \tag{15}$$

$$x_t = \rho x_{t-1} + \sqrt{1-\gamma}\sigma\omega_t, \omega_t \sim N(0,1). \tag{16}$$

We present results for three cases, which we refer to as a case of unbiased initial beliefs, downward-biased initial beliefs, and upward-biased initial beliefs. Figure 13 plots the true parameter values (gray lines) and initial belief distributions (black lines) for these three cases. A more detailed description follows:

- *Unbiased initial beliefs.*—In this case, we set the true parameters to values $\rho = 0.95$, $\gamma = 0.3$, and $\sigma = 0.5$. These values are close to the mean of the belief distribution we estimate from the real-world data in the second half of our sample. We assume that agents in the model have an initial belief distribution with the property that the mode of the belief distribution for each parameter is equal to the truth:

$$\rho \sim N(0.95, 0.01), \gamma \sim \mathcal{B}(9.052, 19.788), \sigma^2 \sim \mathcal{IG}(1.25, 0.5625).$$

- *Downward-biased initial beliefs.*—In this case, we again set the true parameters to values $\rho = 0.95$, $\gamma = 0.3$, and $\sigma = 0.5$. However, we assume that agents in the model have an initial belief distribution

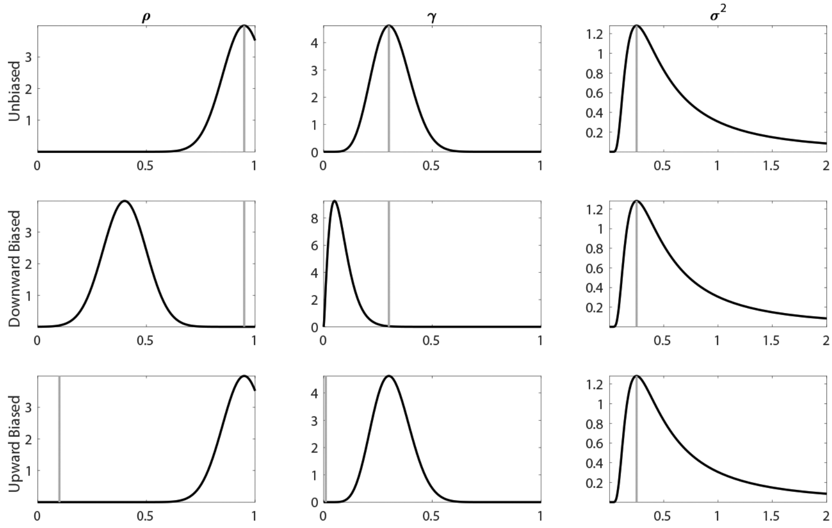

FIG. 13.—Truth and initial beliefs for three simulations. The figure plots the truth (gray line) and initial belief distribution (black line) for ρ (*left*), γ (*middle*), and σ (*right*) for the three cases we consider. The top row is the unbiased initial beliefs case, the middle row is the downward-biased initial beliefs case, and the bottom row is the upward-biased initial beliefs case.

with the property that the modes of the belief distributions for ρ and γ are smaller than the truth:

$$\rho \sim N(0.4, 0.01), \gamma \sim \mathcal{B}(2.34, 26.5), \sigma^2 \sim \mathcal{IG}(1.25, 0.5625).$$

- *Upward-biased initial beliefs.*—In this case, we set the true parameters to values $\rho = 0.1, \gamma = 0.01$, and $\sigma = 0.5$. We then assume that agents in the model have an initial belief distribution with the property that the modes of the belief distributions for ρ and γ are larger than the truth:

$$\rho \sim N(0.95, 0.01), \gamma \sim \mathcal{B}(9.052, 19.788), \sigma^2 \sim \mathcal{IG}(1.25, 0.5625).$$

The reason why we choose different true values for this case is that the true value of ρ used in the other two cases is sufficiently large that it is difficult to illustrate the effects of beliefs that are upward biased relative to this truth.

For each of these three sets of assumptions, we simulate 500 samples of the same length as the short rate data we use in section IV, that is, 275 periods corresponding to the sample period from 1951:2 to 2019:4. For each of these simulated data series, we then perform the same exercise as we did in section IV. Given their initial beliefs, the agents in the model learn about the parameters of the model using the short rate series and

Bayes's law. They then construct Bayesian forecasts. The length of the sample period for the Bayesian forecasts is the same as for the real-world data. We then run the same forecast rationality and expectations hypothesis tests on the resulting data as we did on the real-world data in section IV.

Tables 6 and 7 present the results from this analysis. Table 6 presents results on autocorrelation of forecast errors, the Mincer-Zarnowitz test, and Coibion-Gorodnichenko tests of over- and underreaction, while table 7 presents results on the two tests of the expectations hypothesis we consider in section III. In each case, we report three statistics. The first is the mean estimated coefficient across the 500 simulations, the second is the standard deviation of the estimated effects across simulations (in parentheses), and the third is the fraction of simulations that give a smaller estimate than the estimate based on real-world data.

The main finding from this analysis is that the downward-biased initial beliefs simulation roughly matches all of the anomalies we document in real-world data. This simulation yields positively autocorrelated forecast errors, underreaction in the Coibion-Gorodnichenko regression, values below 1 in the future short rate regression, and negative values for the change in long rate regressions. In virtually all cases, the downward-biased initial beliefs simulation is quantitatively consistent with our real-world estimates of the anomalies in the sense that the real-world estimate is well within the 95% central probability mass of the distribution of estimates from the simulation.

In sharp contrast, the upward-biased initial belief simulation yields anomalies with the opposite sign from what we see in the real-world data. This simulation yields negatively autocorrelated forecast errors, overreaction in the Coibion-Gorodnichenko regression, and values above 1 in both the future short rate regression and the change in long rate regression. In addition to this, the upward-biased initial beliefs simulation also yields a very different pattern from the real-world data for the Mincer-Zarnowitz regression, while the downward-biased initial beliefs simulation matches the real-world data for this regression as well.

Finally, the unbiased initial beliefs simulation yields results that are in most cases consistent with full-information rational expectations on average. It yields virtually no autocorrelation of forecast errors and a coefficient very close to zero in the Coibion-Gorodnichenko regressions (i.e., neither underreaction nor overreaction). For the expectations hypothesis regressions, it yields coefficients that are on average slightly larger than 1 at longer horizons. But the value 1 is not far from the middle of the distribution of coefficients across simulations.

From these results, we conclude that beliefs in society about interest rates in 1951 that underestimated the extent to which fluctuations in interest rates would be persistent relative to what turned out to be the case provide an explanation for the forecast anomalies and failures of the expectations hypothesis that we discuss in section III. As we discuss earlier in the

TABLE 6
Forecast Anomalies in Simulated Data

Initial Beliefs	Forecast Horizon 1	2	3	4
	A. Autocorrelation			
Unbiased:				
Mean estimate	.01	.00	−.00	−.01
SD across simulations	(.08)	(.09)	(.11)	(.13)
Fraction of simulations	1.00	1.00	.99	.84
Downward biased:				
Mean estimate	.16	.19	.19	.18
SD across simulations	(.09)	(.10)	(.12)	(.14)
Fraction of simulations	.93	.78	.61	.33
Upward biased:				
Mean estimate	−.34	−.32	−.28	−.26
SD across simulations	(.06)	(.07)	(.08)	(.08)
Fraction of simulations	1.00	1.00	1.00	1.00
	B. Mincer-Zarnowitz			
Unbiased:				
Mean estimate	.96	.92	.88	.83
SD across simulations	(.03)	(.05)	(.08)	(.11)
Fraction of simulations	.58	.62	.57	.53
Downward biased:				
Mean estimate	.98	.95	.90	.85
SD across simulations	(.03)	(.05)	(.08)	(.11)
Fraction of simulations	.27	.40	.45	.47
Upward biased:				
Mean estimate	.37	.33	.34	.35
SD across simulations	(.17)	(.22)	(.25)	(.27)
Fraction of simulations	1.00	1.00	1.00	.99
	C. Coibion-Gorodnichenko			
Unbiased:				
Mean estimate	.01	.01	.01	. . .
SD across simulations	(.09)	(.12)	(.15)	
Fraction of simulations	.99	.99	1.00	
Downward biased:				
Mean estimate	.18	.32	.41	. . .
SD across simulations	(.11)	(.19)	(.25)	
Fraction of simulations	.66	.55	.79	
Upward biased:				
Mean estimate	−.52	−.55	−.53	. . .
SD across simulations	(.10)	(.13)	(.17)	
Fraction of simulations	1.00	1.00	1.00	

Note.—For each case, the table shows the mean estimate across simulations, the standard deviation across simulations, and the fraction of simulations that give a smaller estimate than the real-world data.

paper, such beliefs seem reasonable given the prior history of interest rate movements. Outside of war, the United States had been on a gold or silver standard and a run-up and run-down of interest rates such as was experiences from the 1960s to the 2000s had never before happened.

TABLE 7
Failures of Expectations Hypothesis in Simulated Data

Initial Beliefs	Long-Horizon n						
	2	3	4	8	12	20	40
	A. Future Short Rates						
Unbiased:							
Mean estimate	.95	1.01	1.05	1.19	1.31	1.51	2.06
SD across simulations	(.64)	(.63)	(.66)	(.72)	(.76)	(.82)	(1.03)
Fraction of simulations	.07	.07	.08	.12	.16	.16	.08
Downward biased:							
Mean estimate	.17	.20	.23	.33	.42	.57	.97
SD across simulations	(.19)	(.21)	(.22)	(.29)	(.33)	(.40)	(.56)
Fraction of simulations	.17	.30	.38	.57	.65	.66	.29
Upward biased:							
Mean estimate	2.46	2.14	1.97	1.71	1.64	1.59	1.50
SD across simulations	(.21)	(.16)	(.14)	(.09)	(.07)	(.06)	(.05)
Fraction of simulations	.00	.00	.00	.00	.00	.00	.00
	B. Change in Long Rate						
Unbiased:							
Mean estimate	.90	.93	.95	1.01	1.08	1.20	2.08
SD across simulations	(1.27)	(1.32)	(1.36)	(1.50)	(1.63)	(1.92)	(3.00)
Fraction of simulations	.07	.08	.07	.06	.05	.03	.03
Downward biased:							
Mean estimate	−.66	−.69	−.74	−1.03	−1.39	−2.04	−3.62
SD across simulations	(.38)	(.40)	(.42)	(.52)	(.64)	(.91)	(1.84)
Fraction of simulations	.17	.28	.24	.32	.34	.51	.66
Upward biased:							
Mean estimate	3.91	4.13	4.38	5.59	6.90	9.56	13.77
SD across simulations	(.42)	(.42)	(.42)	(.51)	(.62)	(.86)	(1.62)
Fraction of simulations	.00	.00	.00	.00	.00	.00	.00

Note.—For each case, the table gives the mean estimate across simulations, the standard deviation across simulations, and the fraction of simulations that give a smaller estimate than the real-world data.

It is instructive to consider the speed of learning about the key parameters ρ and γ in the simulations with downward-biased initial beliefs. Figure 14 plots the evolution of beliefs about these parameters over time in the simulations. The gray line denotes the true value of the parameters. The solid black line plots the evolution of the mean point estimate of the parameters across simulations from 1951:2 to 2019:4. In 1951, the point estimates of both ρ and γ are substantially below the truth. Over time, both estimates rise, but this happens very slowly, and both continue to be substantially below the truth at the end of the sample—when agents have been learning about these parameters for almost 70 years.

Figure 14 shows that it takes substantially longer than 70 years for the agents in our model to learn the true values of the parameters ρ and γ. One reason for this is that increases in ρ and γ both increase the persistence of

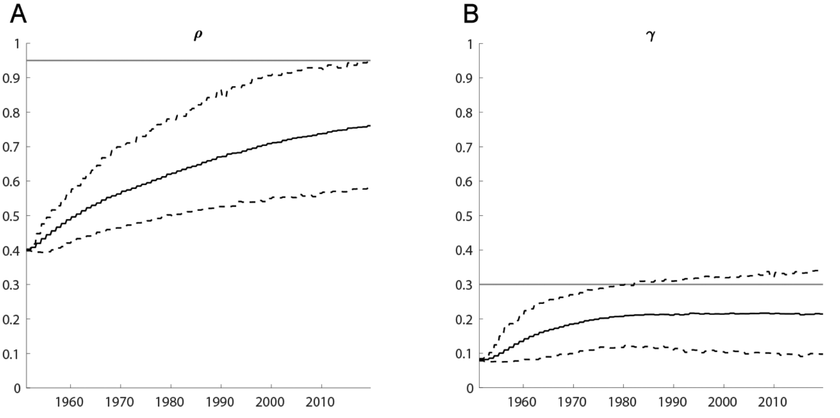

FIG. 14.—Parameter learning with downward-biased initial beliefs. The figure plots the evolution of beliefs about ρ (*A*) and γ (*B*) over time when agents start off with downward-biased initial beliefs. The gray line is the truth. The solid black line is the evolution over time of the mean point estimate across simulation. Recall that the point estimate in a particular simulation is the mean of the belief distribution of the parameter in question in that simulation. The dashed black lines plot the evolution of the 90% and 10% quantiles of the distribution of point estimates across simulations.

fluctuations in the short rate. When agents revise upward their beliefs about the persistence of the short rate, they face the problem of whether the higher persistence is due to a more persistent x_t process (i.e., a higher ρ) or to a more volatile μ_t process (i.e., higher γ). This is an example of what Johannes, Lochstoer, and Mou (2016) refer to as confounded learning, which they argue slows down learning. Figure 15 compares the speed of learning about ρ in the downward-biased case with a case that is the same as the downward-biased case except that γ is set (very close) to zero and agents have a very tight initial belief distribution around the true value of γ—that is, we turn off variation in μ_t and learning about γ. In this case, learning about ρ is much quicker.

Figure 15 shows that confounded learning (i.e., having two unobserved persistent components) slows down learning in our setting. But even when variation in μ_t and learning about γ has been shut down, learning about ρ still takes quite a few decades. This illustrates that learning about the persistence of highly persistent time series processes is quite slow. Unit root tests have low power for similar reasons.

An important point to emphasize is that the agents in the model cannot exploit past forecast anomalies to improve their forecasts. Agents are already optimally incorporating new information to update their beliefs through Bayes's rule. At every point in time, the agents in the model expect that their future forecasts will be free of anomalies. In our simple model, this will eventually be true once their beliefs have converged to

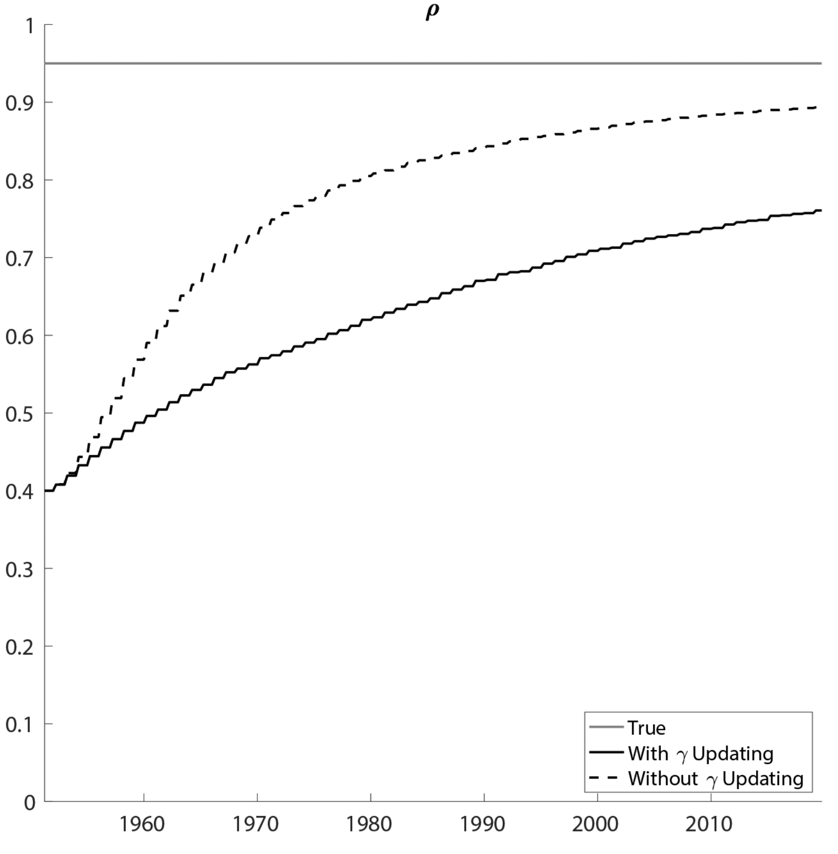

FIG. 15.—Learning about persistence with different values for γ. The figure plots the evolution of beliefs about ρ for the downward-biased case (solid black line) and for a case that is the same as the downward-biased case except that γ is set (very close) to zero and agents have very tight initial beliefs around the true value of γ (dashed black line). The gray line is the truth.

the true parameters. In the short run, each new data point on average moves their beliefs a little bit closer to those true parameters. This process is slowed down by the fact that each data point contains relatively little information about the long-run dynamics of the short rate.

VII. Conclusion

In this paper, we provide a new interpretation of well-known forecast anomalies of professional forecasters. We stress that tests of forecast rationality are joint tests of rationality and the notion that forecasters know the true model of the world. We relax the assumption that forecasters know the true model of the world and show that the anomalies can be explained via Bayesian learning of unobserved components models. Since the anomalies in question persist for decades, it is important that learning is slow in our setting. We show that learning is indeed extremely slow in the type of unobserved components model we consider. This implies that forecasters with reasonable initial beliefs that turn out not to be centered on the truth result in forecast anomalies of the kind we observe in the data that persist for decades. We also perform a simulation exercise in which we know the true value of the parameters. We show in this exercise that reasonably dispersed initial beliefs can yield extremely persistent forecast anomalies. In this simulation exercise, we know that agents are using a correctly specified model to learn and yet learning is extremely slow. Forecast anomalies can thus arise in part for the same reason that it is hard for econometricians to distinguish certain classes of models/parameters even with decades of data, for example, it is hard to reject a unit root in many macroeconomic settings.

Data Availability

Data and code to replicate all results in this paper are available in Farmer, Nakamura, and Steinsson (2024) in the Harvard Dataverse, https://doi.org/10.7910/DVN/JQDGJN.

References

Andolfatto, D., S. Hendry, and K. Moran. 2008. "Are Inflation Expectations Rational?" *J. Monetary Econ.* 55:406–22.

Andrade, P., R. K. Crump, S. Eusepi, and E. Moench. 2016. "Fundamental Disagreement." *J. Monetary Econ.* 83:106–28.

Angeletos, G.-M., Z. Huo, and K. A. Sastry. 2020. "Imperfect Macroeconomic Expectations: Evidence and Theory." *NBER Macroeconomics Ann.* 35:1–86.

Bansal, R., and I. Shaliastovich. 2013. "A Long-Run Risks Explanation of Predictability Puzzles in Bond and Currency Markets." *Rev. Financial Studies* 26:1–33.

Barsky, R. B., and J. B. De Long. 1993. "Why Does the Stock Market Fluctuate?" *Q.J.E.* 108:291–311.

Bauer, M. D., and G. D. Rudebusch. 2020. "Interest Rates under Falling Stars." *A.E.R.* 110:1316–54.

Bekaert, G., R. J. Hodrick, and D. Marshall. 1997. "On Biases in Tests of the Expectations Hypothesis of the Term Structure of Interest Rates." *J. Financial Econ.* 42:309–48.

———. 2001. "Peso Problem Explanations for Term Structure Anomalies." *J. Monetary Econ.* 48:241–70.

Ben-David, I., J. R. Graham, and C. R. Harvey. 2013. "Managerial Miscalibration." *Q.J.E.* 128:1547–84.

Bianchi, F., M. Lettau, and S. C. Ludvigson. 2022. "Monetary Policy and Asset Valuation." *J. Finance* 77:967–1017.

Bianchi, F., S. Ludvigson, and S. Ma. 2022. "Belief Distortions and Macroeconomic Fluctuations." *A.E.R.* 112:2269–315.

Bidder, R., and I. Dew-Becker. 2016. "Long-Run Risk Is the Worst-Case Scenario." *A.E.R.* 106:2494–527.
Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer. 2020. "Overreaction in Macroeconomic Expectations." *A.E.R.* 110:2748–82.
Brav, A., and J. B. Heaton. 2002. "Competing Theories of Financial Anomalies." *Rev. Financial Studies* 15:575–606.
Broer, T., and A. N. Kohlhas. 2024. "Forecaster (Mis-)Behavior." *Rev. Econ. and Statis.* 106:1–18.
Campbell, J. Y., and R. Shiller. 1991. "Yield Spreads and Interest Rate Movements: A Bird's Eye View." *Rev. Econ. Studies* 58:495–514.
Cao, S., R. K. Crump, S. Eusepi, and E. Moench. 2021. "Fundamental Disagreement about Monetary Policy and the Term Structure of Interest Rates." Working paper, Univ. Texas, Austin.
Caskey, J. 1985. "Modelling the Formation of Price Expectations: A Bayesian Approach." *A.E.R.* 75:768–76.
Cieslak, A. 2018. "Short-Rate Expectations and Unexpected Returns in Treasury Bonds." *Rev. Financial Studies* 31:3265–306.
Cieslak, A., and P. Povala. 2015. "Expected Returns in Treasury Bonds." *Rev. Financial Studies* 28:2859–901.
Cochrane, J. 1988. "How Big Is the Random Walk in GNP?" *J.P.E.* 96:893–920.
Cochrane, J. H., and M. Piazzesi. 2005. "Bond Risk Premia." *A.E.R.* 95:138–60.
Cogley, T., and T. J. Sargent. 2005. "The Conquest of US Inflation: Learning and Robustness to Model Uncertainty." *Rev. Econ. Dynamics* 8:528–63.
———. 2008. "The Market Price of Risk and the Equity Premium: A Legacy of the Great Depression?" *J. Monetary Econ.* 55:454–76.
Coibion, O., and Y. Gorodnichenko. 2012. "What Can Survey Forecasts Tell Us about Information Rigidities?" *J.P.E.* 120:116–59.
———. 2015. "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts." *A.E.R.* 105:2644–78.
Collin-Dufresne, P., M. Johannes, and L. A. Lochstoer. 2016. "Parameter Learning in General Equilibrium: The Asset Pricing Implications." *A.E.R.* 106:664–98.
Croce, M. M., M. Lettau, and S. C. Ludvigson. 2015. "Investor Information, Long-Run Risk, and the Term Structure of Equity." *Rev. Financial Studies* 28:706–42.
Croushore, D. 1998. "Evaluating Inflation Forecasts." Working paper, Fed. Reserve Bank Philadelphia.
Crump, R. K., S. Eusepi, E. Moench, and B. Preston. 2021. "The Term Structure of Expectations." In *Handbook of Economic Expectations*, edited by R. Bachmann, G. Topa, and W. van der Klaauw, 507–40. Amsterdam: Elsevier.
Engle, R. F., and C. Granger. 1987. "Co-Integration and Error Correction: Representation, Estimation, and Testing." *Econometrica* 55:251–76.
Eusepi, S., and B. Preston. 2011. "Expectation, Learning, and Business Cycle Fluctuations." *A.E.R.* 101:2844–72.
———. 2018. "The Science of Monetary Policy: An Imperfect Knowledge Perspective." *J. Econ. Literature* 56:3–59.
Evans, G. W., and S. Honkapohja. 2001. *Learning and Expectations in Macroeconomics.* Princeton, NJ: Princeton Univ. Press.
Fama, E. F. 1984. "The Information in the Term Structure." *J. Financial Econ.* 13:509–28.
———. 2006. "The Behavior of Interest Rates." *Rev. Financial Studies* 19:359–79.
Fama, E. F., and R. R. Bliss. 1987. "The Information in Long-Maturity Forward Rates." *A.E.R.* 77:680–92.

Farmer, L. E., E. Nakamura, and J. Steinsson. 2024. "Replication Data for: 'Learning about the Long Run.'" Harvard Dataverse, https://doi.org/10.7910/DVN/JQDGJN.

Friedman, B. M. 1979. "Optimal Expectations and the Extreme Information Assumptions of 'Rational Expectations' Macromodels." *J. Monetary Econ.* 5:23–41.

———. 1980. "Survey Evidence on the 'Rationality' of Interest Rate Expectations." *J. Monetary Econ.* 6:453–65.

Froot, K. A. 1989. "New Hope for the Expectations Hypothesis of the Term Structure of Interest Rates." *J. Finance* 44:283–305.

Giacoletti, M., K. T. Laursen, and K. J. Singleton. 2018. "Learning and Risk Premiums in an Arbitrage-Free Term Structure Model." Working paper, Stanford Univ.

Gourinchas, P.-O., and A. Tornell. 2004. "Exchange Rate Puzzles and Distorted Beliefs." *J. Internat. Econ.* 64:303–33.

Guo, H., and J. Wachter. 2019. "'Superstitious' Investors." Working Paper no. 25603, NBER, Cambridge, MA.

Homer, S., and R. Sylla. 2005. *A History of Interest Rates*, 4th ed. Hoboken, NJ: Wiley.

Johannes, M., L. A. Lochstoer, and Y. Mou. 2016. "Learning about Consumption Dynamics." *J. Finance* 71:551–600.

Johansen, S. 1991. "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica* 59:1551–80.

Kohlhas, A. N., and D. Robertson. 2022. "A Theory of Rational Caution." Working paper, Stockholm Univ.

Kozicki, S., and P. Tinsley. 2001. "Shifting Endpoints in the Term Structure of Interest Rates." *J. Monetary Econ.* 47:613–52.

Kozlowski, J., L. Veldkamp, and V. Venkateswaran. 2020. "The Tail That Wags the Economy: Beliefs and Persistent Stagnation." *J.P.E.* 128:2839–79.

Lazarus, E., D. J. Lewis, J. H. Stock, and M. W. Watson. 2018. "HAR Inference: Recommendations for Practice." *J. Bus. and Econ. Statis.* 36:541–59.

Lewellen, J., and J. Shanken. 2002. "Learning, Asset-Pricing Tests, and Market Efficiency." *J. Finance* 57:1113–45.

Lewis, K. K. 1989a. "Can Learning Affect Exchange-Rate Behavior? The Case of the Dollar in the Early 1980's." *J. Monetary Econ.* 23:79–100.

———. 1989b. "Changing Beliefs and Systematic Rational Forecast Errors with Evidence from Foreign Exchange." *A.E.R.* 79:621–36.

Liu, Y., and J. C. Wu. 2021. "Reconstructing the Yield Curve." *J. Financial Econ.* 142:1395–425.

Maddala, G. S. 1991. "Survey Data on Expectations: What Have We Learnt?" In *Issues in Contemporary Economics*, edited by M. Nerlove, 319–44. London: Palgrave McMillan.

Mankiw, N. G., and J. A. Miron. 1986. "The Changing Behavior of the Term Structure of Interest Rates." *Q.J.E.* 101:211–28.

Mankiw, N. G., and R. Reis. 2002. "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve." *Q.J.E.* 117:1295–328.

Mankiw, N. G., R. Reis, and J. Wolfers. 2003. "Disagreement about Inflation Expectations." *NBER Macroeconomics Ann.* 18:209–48.

Mincer, J. A., and V. Zarnowitz. 1969. "The Evaluation of Economic Forecasts." In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, edited by J. A. Mincer, 3–46. New York: NBER.

Molavi, P., A. Tahbaz-Salehi, and A. Vedolin. 2021. "Model Complexity, Expectations, and Asset Pricing." Working Paper no. 28408, NBER, Cambridge, MA.

Nagel, S., and Z. Xu. 2021. "Dynamics of Subjective Risk Premia." Working paper, Univ. Chicago.

Nordhaus, W. 1987. "Forecasting Efficiency: Concepts and Applications." *Rev. Econ. and Statis.* 69:667–74.

Patton, A. J., and A. Timmermann. 2010. "Why Do Forecasters Disagree? Lessons from the Term Structure of Cross-Sectional Dispersion." *J. Monetary Econ.* 57:803–20.

———. 2011. "Predictability of Output Growth and Inflation: A Multi-Horizon Survey Approach." *J. Bus. and Econ. Statis.* 29:397–410.

Payne, J., B. Szoke, G. J. Hall, and T. J. Sargent. 2022. "Costs of Financing U.S. Federal Debt: 1791–1933." Working paper, Brandeis Univ.

Piazzesi, M., J. Salomao, and M. Schneider. 2015. "Trend and Cycle in Bond Premia." Working paper, Stanford Univ.

Romero, J. 2013. "Treasury-Fed Accord." https://www.federalreservehistory.org/essays/treasury-fed-accord.

Sargent, T. J. 2001. *The Conquest of American Inflation.* Princeton, NJ: Princeton Univ. Press.

Schuh, S. 2001. "An Evaluation of Recent Macroeconomic Forecast Errors." *New England Econ. Rev.* 2001:35–36.

Shiller, R. J. 1979. "The Volatility of Long-Term Interest Rates and Expectations Theories of the Term Structure." *J.P.E.* 87:1190–219.

Shiller, R. J., J. Y. Campbell, and K. L. Schoenholtz. 1983. "Forward Rates and Future Policy: Interpreting the Term Structure of Interest Rates." *Brookings Papers Econ. Activity* 1983:173–223.

Sims, C. A. 2003. "Implications of Rational Inattention." *J. Monetary Econ.* 50:665–90.

Singleton, K. 2021. "How Much 'Rationality' Is There in Bond-Market Risk Premiums?" *J. Finance* 76:1611–54.

Timmermann, A. G. 1993. "How Learning in Financial Markets Generates Excess Volatility and Predictability in Stock Prices." *Q.J.E.* 108:1135–45.

van Dijk, D., S. J. Koopman, M. van der Wel, and J. H. Wright. 2014. "Forecasting Interest Rates with Shifting Endpoints." *J. Appl. Econometrics* 29:693–712.

Vayanos, D., and J.-L. Vila. 2021. "A Preferred Habitat Model of the Term Structure of Interest Rates." *Econometrica* 89:77–112.

Wachter, J. 2006. "A Consumption-Based Model of the Term Structure of Interest Rates." *J. Financial Econ.* 79:365–99.

Woodford, M. 2003. "Imperfect Common Knowledge and the Effects of Monetary Policy." In *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps,* edited by P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford, 25–58. Princeton, NJ: Princeton Univ. Press.

Xu, Z. 2019. "Expectation Formation in the Treasury Bond Market." Working paper, City Univ. Hong Kong.