Learning About the Long Run

Leland Farmer
University of Virginia

Emi Nakamura
UC Berkeley

Jón Steinsson*
UC Berkeley

May 3, 2022

Abstract

Forecasts of professional forecasters are anomalous: they are biased, forecast errors are autocorrelated, and predictable by forecast revisions. Sticky or noisy information models seem like unlikely explanations for these anomalies: professional forecasters pay attention constantly and have precise knowledge of the data in question. We propose that these anomalies arise because professional forecasters don't know the model that generates the data. We show that Bayesian agents learning about hard-to-learn features of the data generating process (low frequency behavior) can generate all the prominent aggregate anomalies emphasized in the literature. We show this for two applications: professional forecasts of nominal interest rates for the sample period 1980-2019 and CBO forecasts of GDP growth for the sample period 1976-2019. Our learning model for interest rates also provides an explanation for deviations from the expectations hypothesis of the term structure that does not rely on time-variation in risk premia.

JEL Classification: E43, E37, G12.

*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 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.

1 Introduction

For almost half a century, the assumption that people form rational expectations has dominated economic modelling in macroeconomics and 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, i.e., 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 et al., 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 don't know the correct model of the world, they use incoming data to learn about how the world works. But such learning can fundamentally change the dynamics

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

of even perfectly rational Bayesian forecasts. This idea has been recognized by researchers at least since Friedman (1979).[2]

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 can't 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.[3]

We develop two applications of these ideas, one for forecasting nominal interest rates and another for forecasting real 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 U.S. 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 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

[2]Other important papers that emphasize this idea include Caskey (1985), Lewis (1989b,a), Barsky and De Long (1993), Timmermann (1993), Lewellen and Shanken (2002), Brav and Heaton (2002), Cogley and Sargent (2005), Collin-Dufresne, Johannes, and Lochstoer (2016), Johannes, Lochstoer, and Mou (2016), Guo and Wachter (2019), Singleton (2021).

[3]Collin-Dufresne, Johannes, and Lochstoer (2016) and Kozlowski, Veldkamp, and Venkateswaran (2020) develop models with rare events in which Bayesian learning is slow.

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-WWII 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 U.S. 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. This finding echos the more general finding in the forecasting literature that allowing for unrestricted priors in complicated learning models often does not improve forecasting performance.

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 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

[4]See Fama (2006, p. 360-361) for a narrative description of these ideas.

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 over-reaction rather than under-reaction 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 in the 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 towards 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.

In our interest rate application, we focus on the mean forecast from the Survey of Professional Forecasters. An important related literature has sought to understand the behavior of individual forecasts relative to the mean forecast as well as forecast dispersion. 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 due to heterogeneity in priors rather than differences in information sets. We have chosen to abstract from such heterogeneity for simplicity. (What we do is already very computationally demanding.)

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 (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, 2) diagnostic expectations to generate overreaction of individual forecasts. We view our model as an alternative to the first

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

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 the individual level, just as Bordalo et al. (2020) combine diagnostic expectations with noisy information.[6]

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; Collin-Dufresne, Johannes, and Lochstoer, 2016; Kozlowski, Veldkamp, and Venkateswaran, 2020; Bidder and Dew-Becker, 2016). 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). Ben-David, Graham, and Harvey (2013) provide evidence for Bayesian learning among firm CFOs.

The paper proceeds as follows. Section 2 describes our data. Section 3 reviews forecasting anomalies for interest rates and real GDP data. Section 4 presents our model and results for nominal interest interest rates. Section 5 presents our model and results for real GDP growth. Section 6 presents Monte Carlo simulation exercises aimed to shed light on why our results turn out the way they do. Section 7 concludes.

2 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.

2.1 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 Reserve Bank of Philadelphia. Our sample period for these forecasts is 1981Q3 to 2019Q4. The SPF is a quarterly survey sent out to a rotating panel of forecasters. We use the mean forecast across forecasters. Figure 1 plots these forecasts.

The survey is sent out near the end of the first month of each quarter. The forecast therefore roughly coincides with the BEA's advance report of the national income and product accounts. Survey response deadlines are in the second to third week of the second month of the same quarter.

[6]Other important papers in this literature include Patton and Timmermann (2011), Andrade et al. (2016), Angeletos, Huo, and Sastry (2020), Crump et al. (2021), Cao et al. (2021), Singleton (2021), Broer and Kohlhas (2022).

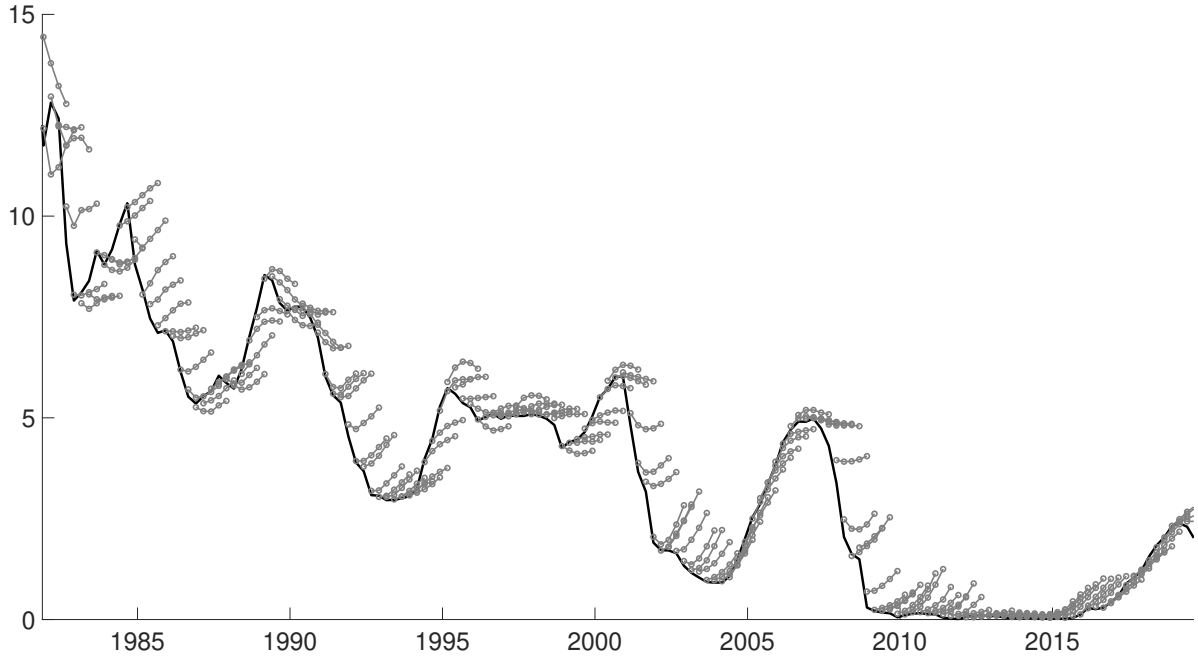

Figure 1: SPF Forecasts of the 3-Month T-Bill Rate

Note: The black solid 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).

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 is from the Board of Governors of the Federal Reserve System.[7] Our sample period for this series is 1951Q2 to 2019Q4. 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 (2020). 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 these data to quarterly data by computing the average yield in a quarter. Our sample period for these zero coupon bond yields is 1961Q3 to 2019Q4.

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

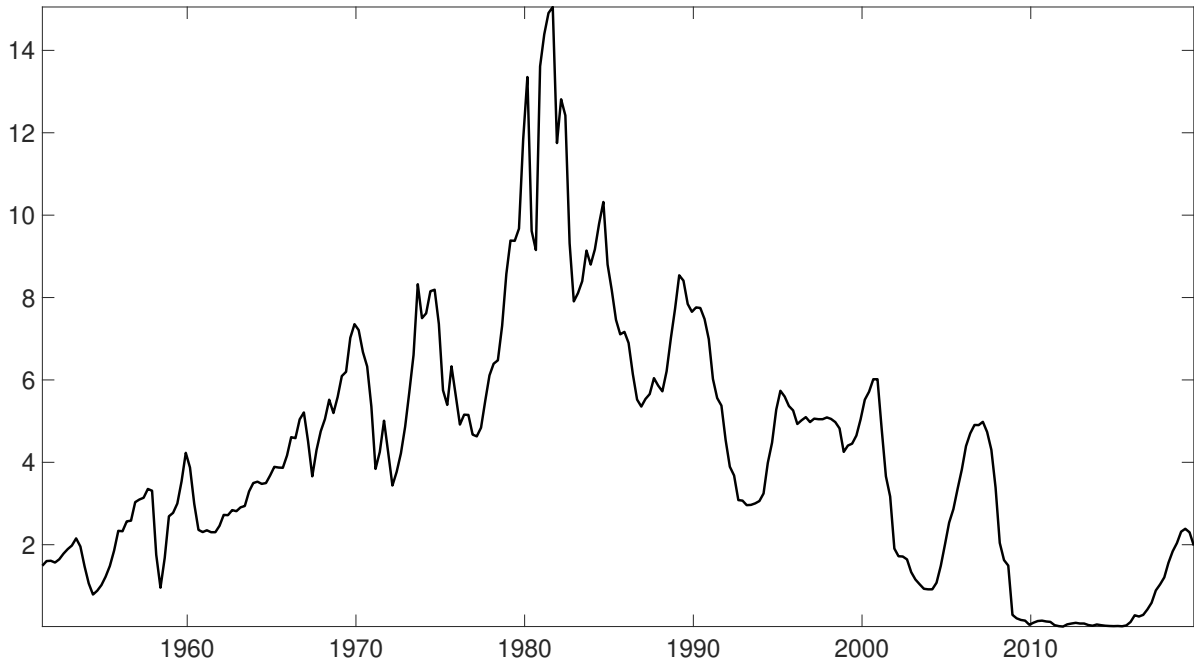

Figure 2: The 3-Month T-Bill Rate

2.2 Real GDP Growth Data and Forecasts

The real GDP growth forecasts we analyze are from the Congressional Budget Office (CBO).[8] 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 forecast real Gross National Product (GNP). Since then, they have forecast real Gross Domestic Product (GDP). For expositional simplicity, we refer to these as real GDP forecasts throughout the paper.

The data we use on actual real GDP growth is from the Philadelphia Federal Reserve Bank's Real-Time Data Set. This source publishes monthly vintages of real-time real output back to November of 1965. Most vintages contain data back to 1947Q1. However, a few vintages are missing data before 1959Q3, which limits our sample period as we discuss in greater detail in section 5.

[8]CBO forecasts of real GDP growth are very similar to analogous Blue Chip consensus real GDP growth forecasts (Congressional Budget Office, 2021).

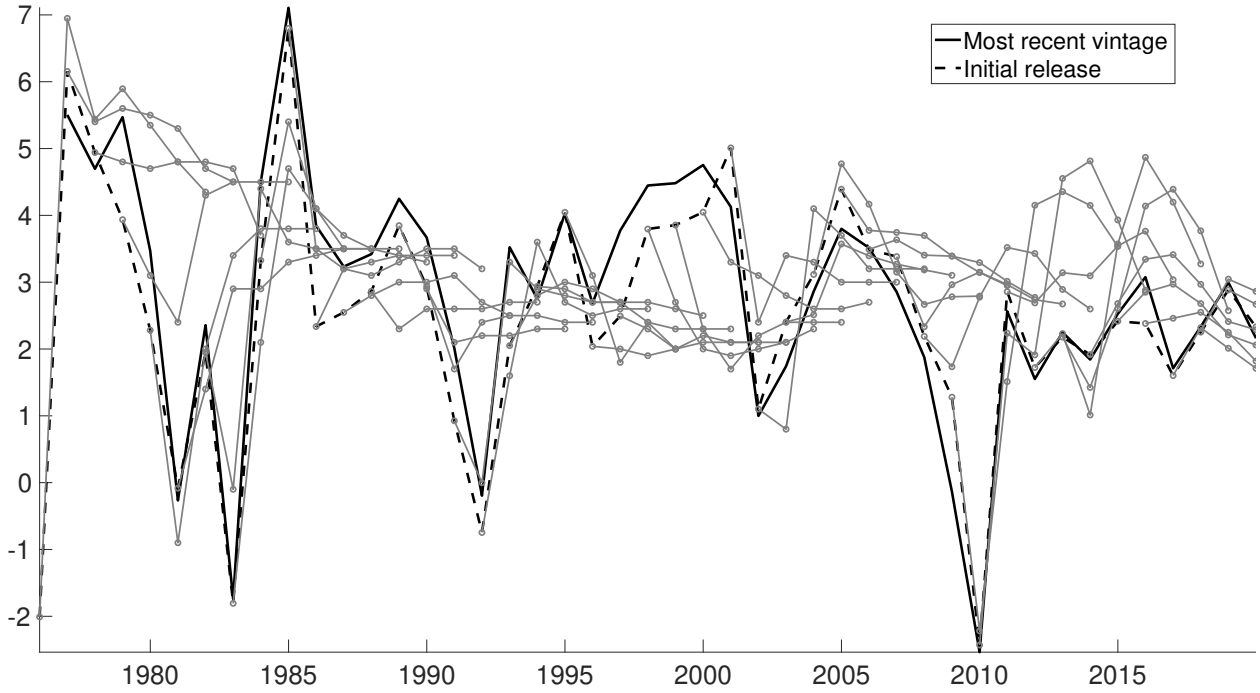

Figure 3: CBO Forecasts of Real GDP Growth

Note: The black solid line is the 2021Q1 vintage of real GDP growth from 1976 to 2019. The broken 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 six years (subsequent six circles).

3 Forecasting Anomalies

As we discuss in the introduction, the forecasts of professional forecasters exhibit a number of "anomalies"—i.e., 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 one to five *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 vs. real GDP growth.

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 4-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.

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, non-overlapping 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 $\beta = 0$, i.e., 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 1-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 two. CBO forecasts of GDP growth also display positive autocorrelation. But

Table 1: Forecast Anomalies

	Forecast Horizon				
	1	2	3	4	5
Panel A: Bias					
T-Bill	-0.18***	-0.34***	-0.52***	-0.70***	–
	(0.05)	(0.09)	(0.14)	(0.19)	
GDP Growth	0.27	-0.27	-0.54	-0.62	-0.52
	(0.25)	(0.35)	(0.50)	(0.53)	(0.49)
Panel B: Autocorrelation					
T-Bill	0.30*	0.27**	0.24*	0.13	–
	(0.14)	(0.12)	(0.12)	(0.13)	
GDP Growth	0.22	0.16	0.11	0.08	0.08
	(0.12)	(0.14)	(0.13)	(0.18)	(0.10)
Panel C: Mincer-Zarnowitz					
T-Bill	0.97*	0.94**	0.90**	0.86**	–
	(0.02)	(0.02)	(0.04)	(0.05)	
GDP Growth	0.94	0.60	0.03**	-0.42***	-0.43***
	(0.10)	(0.38)	(0.27)	(0.18)	(0.29)
Panel D: Coibion-Gorodnichenko					
T-Bill	0.23*	0.34*	0.62***	–	–
	(0.12)	(0.16)	(0.16)		
GDP Growth	0.08	0.00	0.50	-1.63**	-1.46**
	(0.08)	(0.28)	(0.58)	(0.36)	(0.40)

Note: The forecast horizons for the T-Bill are quarters, while the forecast horizons for the GDP growth are years. Stars represent significance relative to the following hypotheses: $\alpha = 0$ for bias, $\beta = 0$ for autocorrelation, $\beta = 1$ for Mincer-Zarnowitz, $\beta = 0$ 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 9 for the GDP growth regressions. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

in this case the autocorrelation is smaller and not statistically significantly different from zero.

Mincer-Zarnowitz Regressions

A classic test of forecast rationality proposed by Mincer and Zarnowitz (1969) investigates the intuitive prediction that the truth should on 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 $\beta = 1$, i.e., 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 one, 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 three years earlier than when it was forecast to be low three years earlier. For the 4-year 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 one. In contrast, our estimate of β for the T-bill forecasts are close to one. They are somewhat below one and the difference is statistically significant. But the deviation from the null of one is much less stark than in the case of GDP forecasts.

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, i.e., didn't 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 1-quarter horizon to 0.64 at the 3-quarter horizon. For GDP growth forecasts, however, we estimate neither over- nor underreaction at short horizons. At the 4-year and 5-year horizons, however, we estimate negative values of β indicating overreaction.

Table 2: Failures of the Expectations Hypothesis

	Long Horizon n						
	2	3	4	8	12	20	40
Future Short Rates	-0.01***	0.11***	0.18***	0.39**	0.57	0.74	0.71
	(0.23)	(0.23)	(0.23)	(0.23)	(0.26)	(0.23)	(0.20)
Change in Long Rate	-1.02***	-0.91***	-1.03***	-1.29***	-1.61***	-2.04***	-2.75***
	(0.45)	(0.59)	(0.62)	(0.59)	(0.57)	(0.55)	(0.87)

Note: The sample period is from 1961Q3 to 2019Q4. The top row reports estimates of β from regression (4). The bottom row reports estimates of β from regression (5). In both cases, the horizon n is listed at the top of the table. Stars 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

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 values of yields on 1-period bonds 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:

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

where $y_t^{(n)}$ denotes the yield of a 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 2 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 one, but remains below one 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.

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

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

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 negative one 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 variables 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 6 based on our model from section 4. This analysis does find evidence of some small sample biases. But the quantitative

magnitude of these biases is small.

4 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 3 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.

4.1 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:

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

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

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

Here, the short rate y_t is modelled 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 have mean 0 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 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 (UC) model.

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

To gain intuition about the model, consider the h-period forecast of the short rate assuming 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.

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.

4.2 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 Law to update their beliefs about the states and parameters over time and then in each period construct forecasts of future short rates based on their then current beliefs. More specifically, we start the agents off with initial beliefs in 1951Q2. The agents then use data on the short rate from 1951Q2 onward to update their beliefs. Starting in 1961Q3 they perform "online" forecasting of the short rate. In other words, each quarter they forecast the short rate based their beliefs at that point in time.

The world did not begin in 1951Q2. So, why do we not use data going further back in time? The reason for this is that the monetary policy regime in the U.S. was fundamentally different before 1951Q2. In March 1951, the U.S. 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 WWII and manage the massive government debt left after WWII. Before 1942, the U.S. 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 A. 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 – i.e., 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 B. We do this for each quarter starting in 1961Q3, 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 we avoid having to adopt the "anticipated utility" approach to agent forecasting (Kreps, 1998; Eusepi and Preston, 2018) which is common in research analyzing boundedly rational learning. Another advantage is that it does not matter how we write our model. For example, our UC model has an ARIMA representation. The Bayesian agents in our model see through the superficial difference between the UC 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 interest 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 only have information 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 (ZLB) towards 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 ZLB period. This corresponds to 2009Q1 to 2015Q4 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 2008Q4. Full learning then resumes in 2016Q1. This short-cut allows us to avoid substantial additional complications which we believe are unlikely to materially affect our results.[10]

4.3 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 non-trivially 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. We assume that initial beliefs about μ_t in 1951Q2 are $N(y_{1951Q12}, 1)$ and initial beliefs about x_t in 1951Q2 are $N(0, 1)$. These initial conditions are assumed to have a correlation of -1 due to the form of the observation equation (6).

For ρ, γ, and σ we specify initial beliefs in 1951Q2 of the following form:

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

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 1951Q2, based on prior history. We search over the space of initial beliefs specified above for the initial beliefs which can best rationalize the forecast anomalies we document in section 3. 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 beliefs for professional forecasters to have in 1951Q2. 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.

[10] A fuller treatment would explicitly allow for censoring of the desired short rate. This would require us to shift to non-linear 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 minimal.

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.[11] 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 3. 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 3 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 C provides more detail.

The top two 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 eight 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 1951Q2 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, i.e., 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 1951Q2. Second, the belief that γ was relatively small is arguably consistent with the history of interest rates prior to 1951Q2. Outside of war, the United States had been on a gold standard (or silver standard) almost continuously from its founding (and England had been on a gold or bimetallic standard for hundreds of years before that). Over this long time span, interest rates had been quite stable at low frequencies with most variation being rather transient (due to seasonal cycles and financial crises). Given this history, it does not seem unreasonable that forecasters' beliefs were skewed towards believing that most fluctuations in interest rates would be relatively transient. The long upward march of interest rates in the 1960s, 70s, and early 80s and subsequent downward march of interest rates since then was completely without

[11] 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.

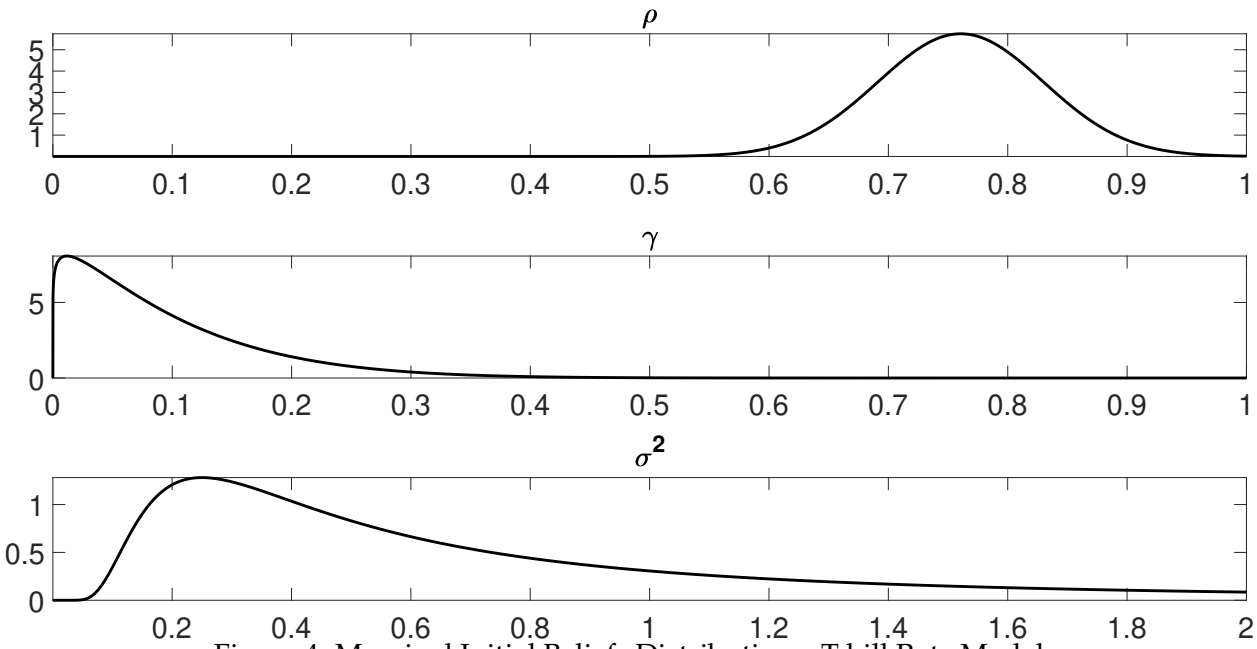

Figure 4: Marginal Initial Beliefs Distributions: T-bill Rate Model

Note: Each panel plots the initial beliefs held in 1951Q2 by agents in our T-bill rate model for each of the three model parameters: ρ, γ, and σ^2 respectively.

parallel in history. It seems unlikely that forecasters in 1951Q2 would put much weight on such an unprecedented sequence of events occurring.

While wider initial beliefs result in smaller forecast anomalies, they do not yield appreciably smaller root-mean-squared errors (RMSE). In appendix C, we compare the RMSE of our baseline model with cases where the forecasters have wider initial beliefs for ρ and γ. Moderately wider initial beliefs yields ever so slightly smaller RMSE (less than 1% difference), while still wider initial beliefs yield ever so slightly larger RMSE (again, less than 1% difference). All three cases yield RMSE that are about 4% smaller than the SPF forecasts.

4.4 Model's Fit to the Data

Figure 5 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 as is plotted in Figure 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 towards a "normal" value that is shifting over time, i.e., something close to average value of the short rate over the past business cycle. For example, in the easing cycle of 1985-1987, SPF forecasters consistently expect the short rate to rise. This leads them to be wrong in their forecast in the same

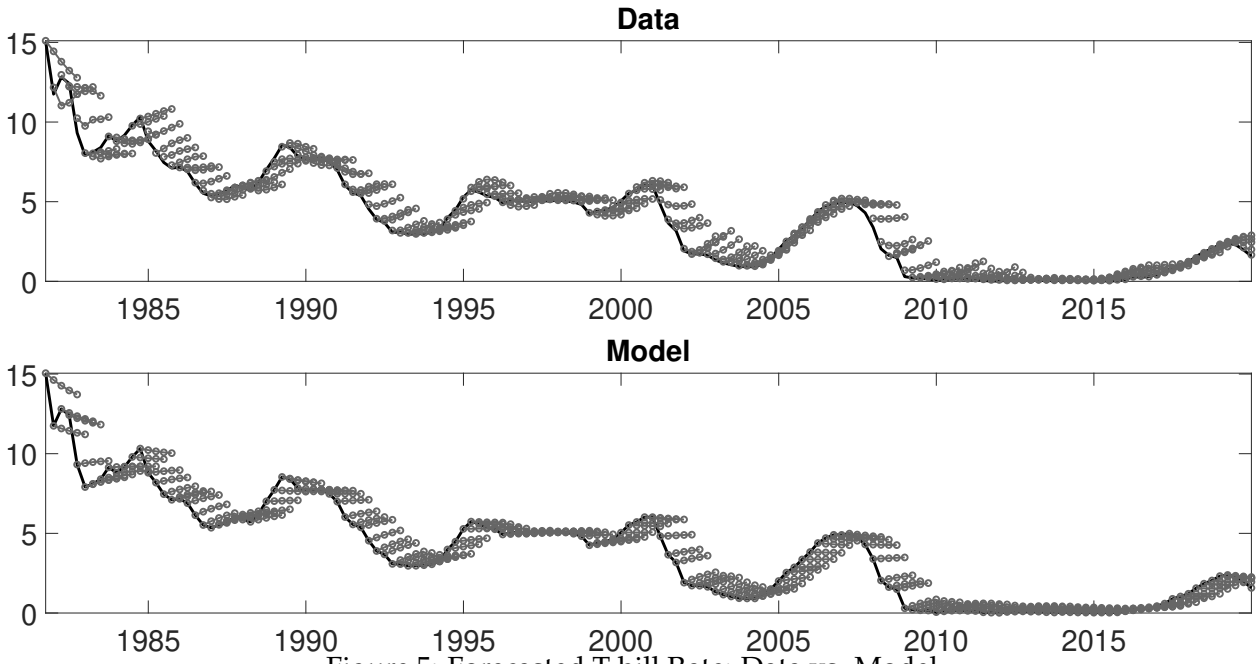

Figure 5: Forecasted T-bill Rate: Data vs. Model

Note: The black solid 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 UC model estimated in real-time.

direction over and over again. The same is true for agents in our model. This pattern repeats in later easing cycles such as 1991-1993 and 2001-2003. When rates are rising, SPF forecasters expect them to rise more slowly than they actually do. This occurs in 1988-1989, in 1994, and in 1999-2000 and leads to highly autocorrelated forecast errors. Our model matches this pattern.

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 2012-2015, 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 2004-2007 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 3 for our model-generated data (rows labelled "UC Model") and compares these with analogous results for the real-world data (rows labelled "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

Table 3: T-Bill Rate Forecast Anomalies: Model vs. Data

	Forecast Horizon			
	1	2	3	4
Panel A: Bias				
SPF	-0.18***	-0.34***	-0.52***	-0.70***
	(0.05)	(0.09)	(0.14)	(0.19)
UC Model	-0.15**	-0.27**	-0.40**	-0.51**
	(0.06)	(0.11)	(0.16)	(0.21)
Panel B: Autocorrelation				
SPF	0.30*	0.27**	0.24*	0.13
	(0.14)	(0.12)	(0.12)	(0.13)
UC Model	0.36*	0.39**	0.35**	0.23*
	(0.17)	(0.14)	(0.11)	(0.12)
Panel C: Mincer-Zarnowitz				
SPF	0.97*	0.94**	0.90**	0.86**
	(0.02)	(0.02)	(0.04)	(0.05)
UC Model	0.96*	0.93**	0.88**	0.84***
	(0.02)	(0.03)	(0.04)	(0.05)
Panel D: Coibion-Gorodnichenko				
SPF	0.23*	0.34*	0.62***	–
	(0.12)	(0.16)	(0.16)	
UC Model	0.39*	0.56	0.89*	–
	(0.18)	(0.37)	(0.42)	

Note: The forecast horizons are quarters. Stars represent significance relative to the following hypotheses: $\alpha = 0$ for bias, $\beta = 0$ for autocorrelation, $\beta = 1$ for Mincer-Zarnowitz, $\beta = 0$ 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

with the horizon, as in the data; autocorrelation in forecast errors of about 0.35 at horizons one through three and much less at horizon four, as in the data; Mincer-Zarnowitz coefficients slightly below one 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 3 based on model-generated data and compares these results with those based on real-world data. Again, our model matches the 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

Table 4: Failures of the Expectations Hypothesis: Model vs. Data

	Long Horizon n						
	2	3	4	8	12	20	40
Panel A: Future Short Rates							
Data	-0.01***	0.11***	0.18***	0.39**	0.57	0.74	0.71
	(0.23)	(0.23)	(0.23)	(0.23)	(0.26)	(0.23)	(0.20)
UC Model	-0.11***	0.08**	0.17**	0.56	0.81	0.93	0.99
	(0.32)	(0.32)	(0.33)	(0.38)	(0.37)	(0.31)	(0.36)
Panel B: Change in Long Rate							
Data	-1.02***	-0.91***	-1.03***	-1.29***	-1.61***	-2.04***	-2.75***
	(0.45)	(0.59)	(0.62)	(0.59)	(0.57)	(0.55)	(0.87)
UC Model	-1.21***	-1.25***	-1.28***	-1.40***	-1.54***	-1.84***	-2.55**
	(0.63)	(0.64)	(0.65)	(0.70)	(0.76)	(0.88)	(1.52)

Note: The sample period is from 1961Q3 to 2019Q4. The top panel reports estimates of β from regression (4). The bottom panel reports estimates of β from regression (5). In both cases, the horizon n is listed at the top of the table. Stars 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

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 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.

4.5 Parameter and State Estimates

Figure 6 plots the evolution of the mean of the posterior distributions of ρ, γ, and σ along with 90% credible intervals between 1961Q3 and 2019Q4. 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 to about 0.76 for the initial beliefs (in 1951Q2). It then gradually rises further over the sample and is around 0.9 towards 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

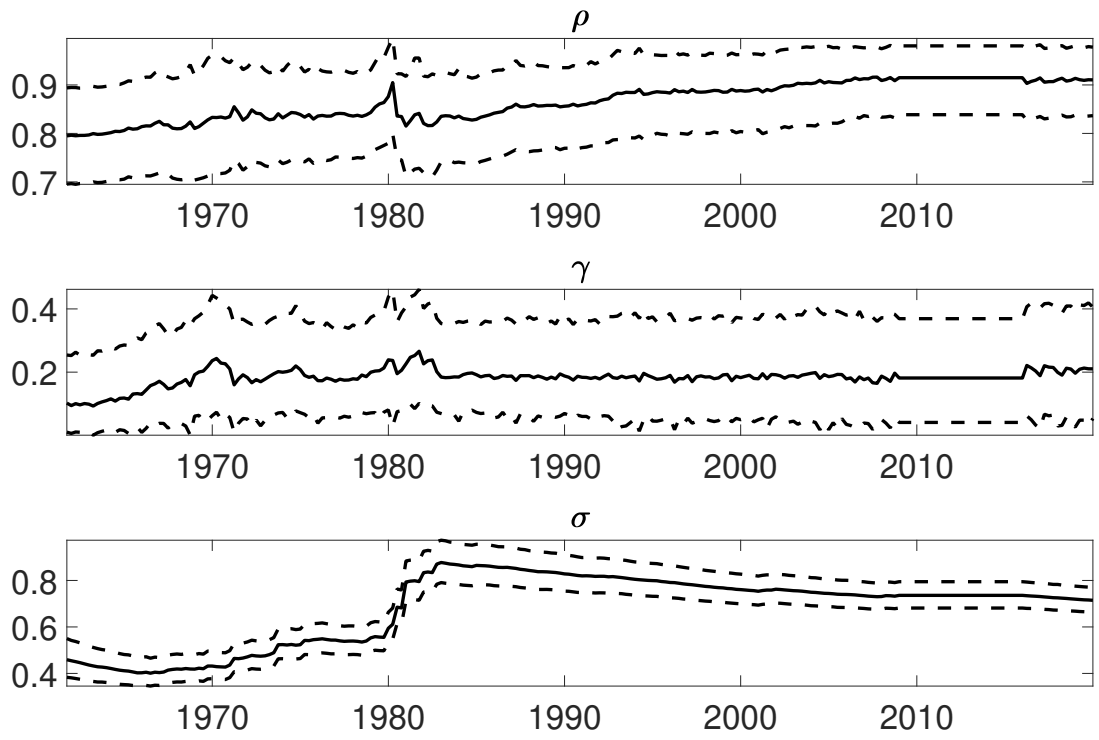

Figure 6: Parameter Estimates: T-bill Rate Model

Note: Each panel plots the evolution of beliefs about one of the three UC model parameters: ρ, γ, and σ. The black solid line is the mean and the dotted black lines are the 5th and 95th percentiles of the posterior distribution for the parameter in question. Recall that we only update beliefs about these parameters every fourth quarter.

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 7 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, i.e., 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, i.e., the belief distribution about μ_t conditional on data up to 2019Q4. The broken 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 7. 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 70s, and again underestimate the persistence of the fall in interest rates after the early 1980s. Ex post,

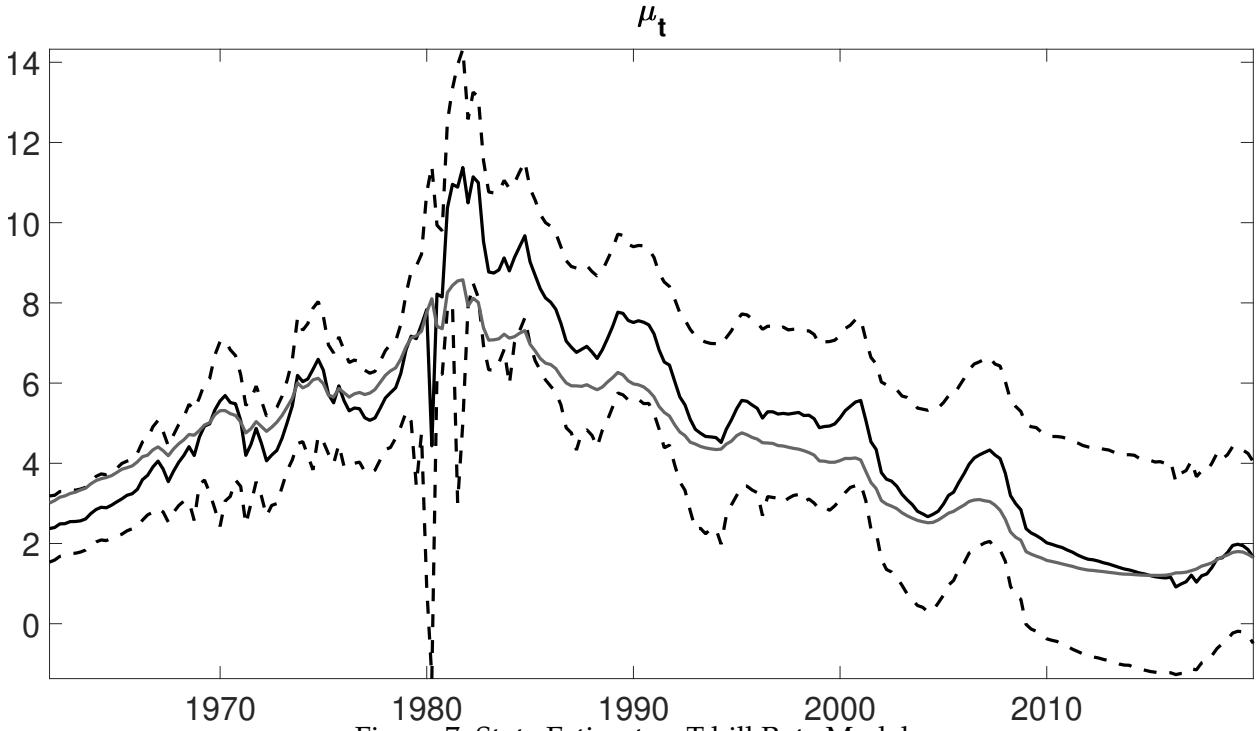

Figure 7: State Estimates: T-bill Rate Model

Note: This figure plots the evolution of beliefs about the permanent component μ_t. The black solid line is the posterior mean of the real-time filtering distributions, the dotted black lines are the 5th and 95th percentiles of the posterior real-time filtering distributions, and the solid gray line is the posterior mean of the ex-post smoothing distributions.

agents revise their view of history and conclude that both the run-up and fall in interest rates was 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.

4.6 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 market traders at some point became convinced that the change in monetary policy implemented by Paul Volcker and carried on by Alan Greenspan that 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.

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 Appendix D.)

Figure 8 plots the yield spread between a 10-year zero coupon bond and the 3-month Treasury 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 one-year excess returns on two- to five-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 one-year excess returns on 2- to 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.

Giglio and Kelly (2018) argue that long-maturity assets display excess volatility that is hard to reconcile with standard asset pricing models. They posit an affine term structure model, estimate the parameters of this model off of medium-term bond prices (1 to 7 years), and then compare the volatility of long-term bond prices implied by the estimated model with the actual volatility of long-term bond prices. Their main result is that the actual volatility is larger than the volatility implied by the affine term structure model. In appendix D, we show that our learning model can

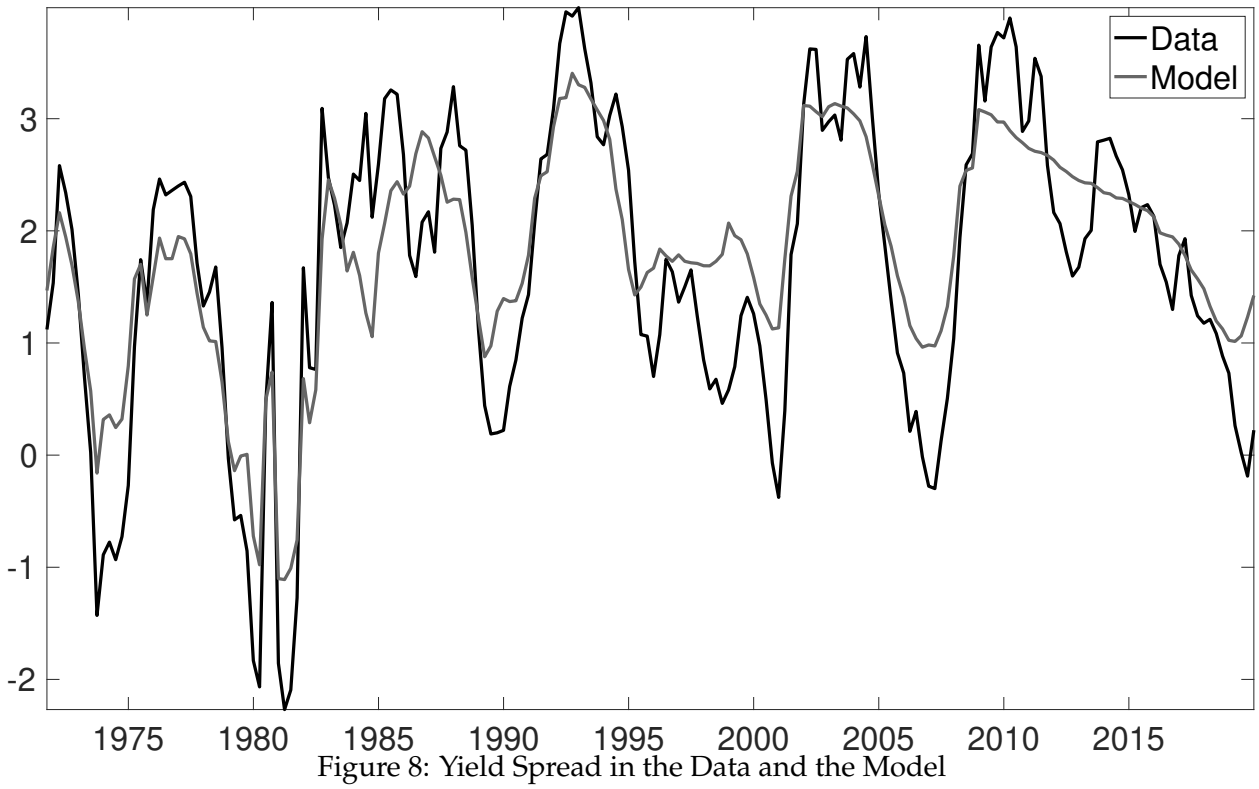

Figure 8: Yield Spread in the Data and the Model

Note: The Figure plots the spread between the yield on a 10-year zero coupon bond and the 3-month Treasury bill rate for the data and the model.

generate this type of excess volatility. However, we also show that their excess volatility result is sensitive to the details of which yields are used to estimate the affine terms structure model.

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.

5 Learning about the Real GDP Growth

Our second application is to learning about real GDP growth. As in section 4, 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 3.

5.1 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:

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

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

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

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 (UC) 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.

5.2 Bayesian Learning and Forecasting about GDP

As in the interest rate application discussed in section 4, 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 1959Q3. 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 Law. Below we plot results starting in 1976Q1. This corresponds to the first period for which CBO forecasts are available.

We assume that agents have access to the first release of Q4 GDP for the prior year (the BLS's "advance release" for that quarter) 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 Q4 GDP for the previous year, much of the underlying data that is used to construct the advance release has been made public at this point. This implies that the Q4 advance release can be predicted fairly accurately. We therefore think that endowing our model agents with the Q4 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 4 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 E. 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 F.

5.3 Initial Beliefs about GDP

As in the interest rate application in section 4, 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 is less consequential. We assume that agents' initial beliefs about z_t and x_t in 1959Q3 are $z_t \sim N(y_{1959Q3}, 0.01^2)$ and $x_t \sim N(0, 0.01^2)$.

We specify initial beliefs for the parameters in 1959Q3 of the following form

$$\rho_1 + \rho_2 \sim N(\mu_\rho, \sigma_\rho^2)\mathcal{I}(\rho_1, \rho_2), \qquad \rho_2 \sim N(\mu_{\rho_2}, \sigma_{\rho_2}^2)\mathcal{I}(\rho_1, \rho_2),$$

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

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

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 6 parameters to estimate to fit the forecast anomalies presented in section 2, 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 4. Appendix G provides details.[12]

[12]We place some bounds on the values of parameters that can be chosen in this estimation. Namely, we restrict

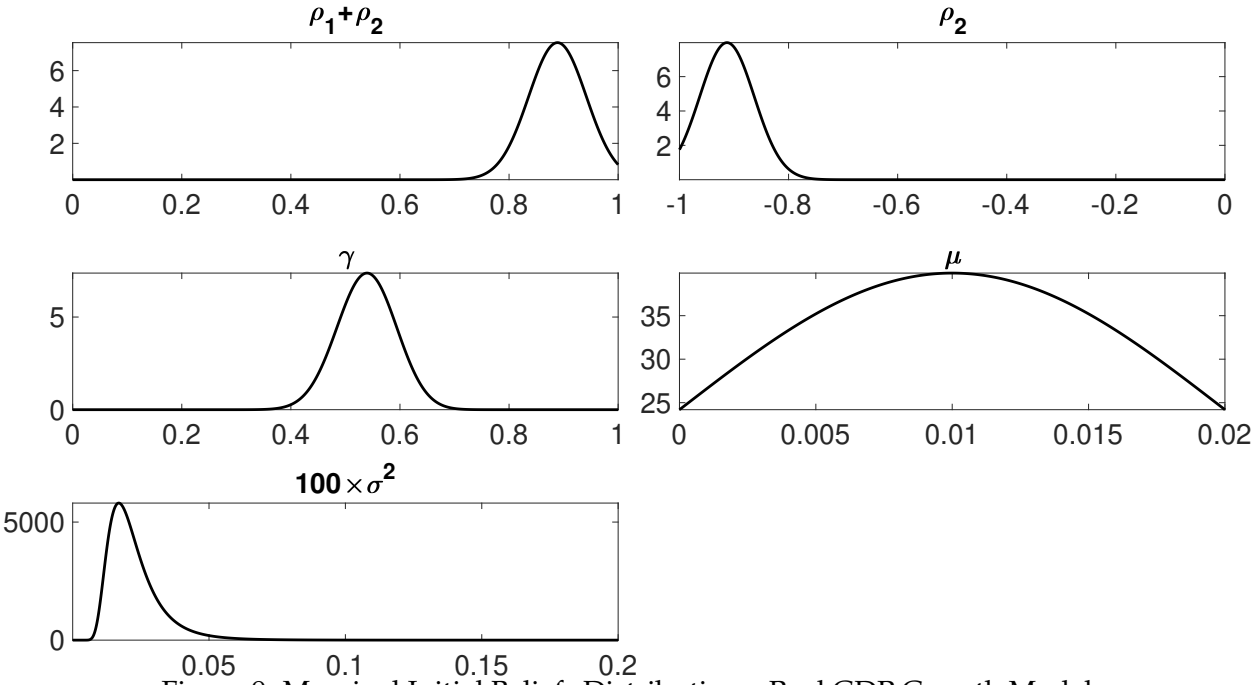

Figure 9: Marginal Initial Beliefs Distributions: Real GDP Growth Model

Note: Each panel plots the initial beliefs held in 1959Q3 by agents in our model for the following five parameter combinations: $\rho_1 + \rho_2$, ρ_2, γ, μ, and σ^2.

The resulting initial beliefs are plotted in Figure 9. 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 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.

5.4 Model's Fit to the Data

Figure 10 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 as is plotted in Figure 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 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 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.

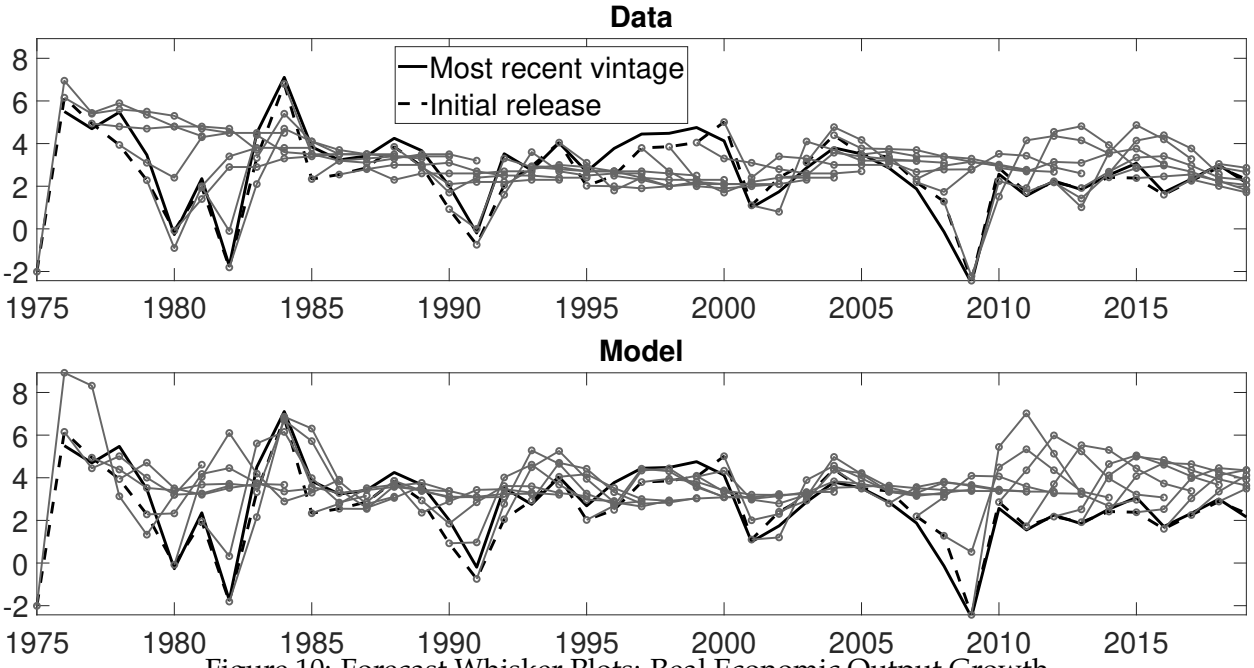

Figure 10: Forecast Whisker Plots: Real Economic Output Growth

Note: The black solid 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 following six 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 UC model estimated in real-time.

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 3 for our model-generated data (rows labelled "UC Model") and compares these with analogous results for the real-world data (rows labelled "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 one at the one-year horizon but fall to zero at the three-year horizon and to roughly -0.4 at the four and five-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 if not as impressive as in our interest rate application.

Table 5: Real GDP Forecast Anomalies: Model vs. Data

	Forecast Horizon				
	1	2	3	4	5
Panel A: Bias					
CBO	0.27	-0.27	-0.54	-0.62	-0.52
	(0.25)	(0.35)	(0.50)	(0.53)	(0.49)
UC Model	-0.65	-1.65**	-1.36**	-0.85	-0.66
	(0.32)	(0.45)	(0.45)	(0.42)	(0.40)
Panel B: Autocorrelation					
CBO	0.22	0.16	0.11	0.08	0.08
	(0.12)	(0.14)	(0.13)	(0.18)	(0.10)
UC Model	0.39*	0.31	0.23*	0.06	-0.05
	(0.17)	(0.16)	(0.10)	(0.10)	(0.05)
Panel C: Mincer-Zarnowitz					
CBO	0.94	0.60	0.03**	-0.42***	-0.43***
	(0.10)	(0.38)	(0.27)	(0.18)	(0.29)
UC Model	0.84	0.35**	0.34*	-0.38***	-0.98**
	(0.11)	(0.17)	(0.31)	(0.19)	(0.53)
Panel D: Coibion-Gorodnichenko					
CBO	0.08	0.00	0.50	-1.63**	-1.46**
	(0.08)	(0.28)	(0.58)	(0.36)	(0.40)
UC Model	0.06	-0.76	-0.11	-0.78	-1.22**
	(0.09)	(0.44)	(0.26)	(0.39)	(0.38)

Note: The forecast horizons are years. Stars represent significance relative to the following hypotheses: $\alpha = 0$ for bias, $\beta = 0$ for autocorrelation, $\beta = 1$ for Mincer-Zarnowitz, $\beta = 0$ 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

5.5 Parameter Estimates

Figure 11 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 11 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

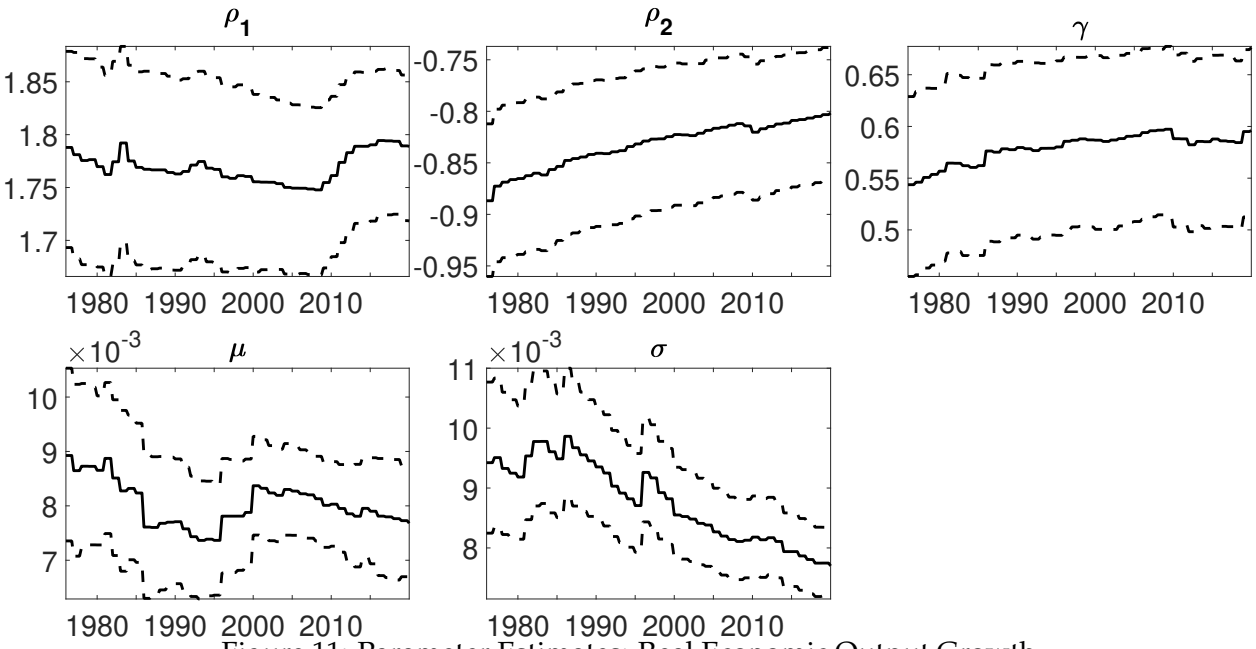

Figure 11: Parameter Estimates: Real Economic Output Growth

Note: Each panel plots the evolution of beliefs about one of the five UC model parameters: ρ_1, ρ_2, γ, μ, and σ. The black solid line is the mean and the dotted black lines are the 5th and 95th percentiles of the posterior distribution for the parameter in question. Recall that we only update beliefs about these parameters every fourth quarter.

being little changed even over a 40 year period. This is perhaps not surprising given how difficult it is to distinguish between difference-stationary time series and persistent but trend-stationary time series.

6 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 3, we now simulate data from the model we use in section 4 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 3.

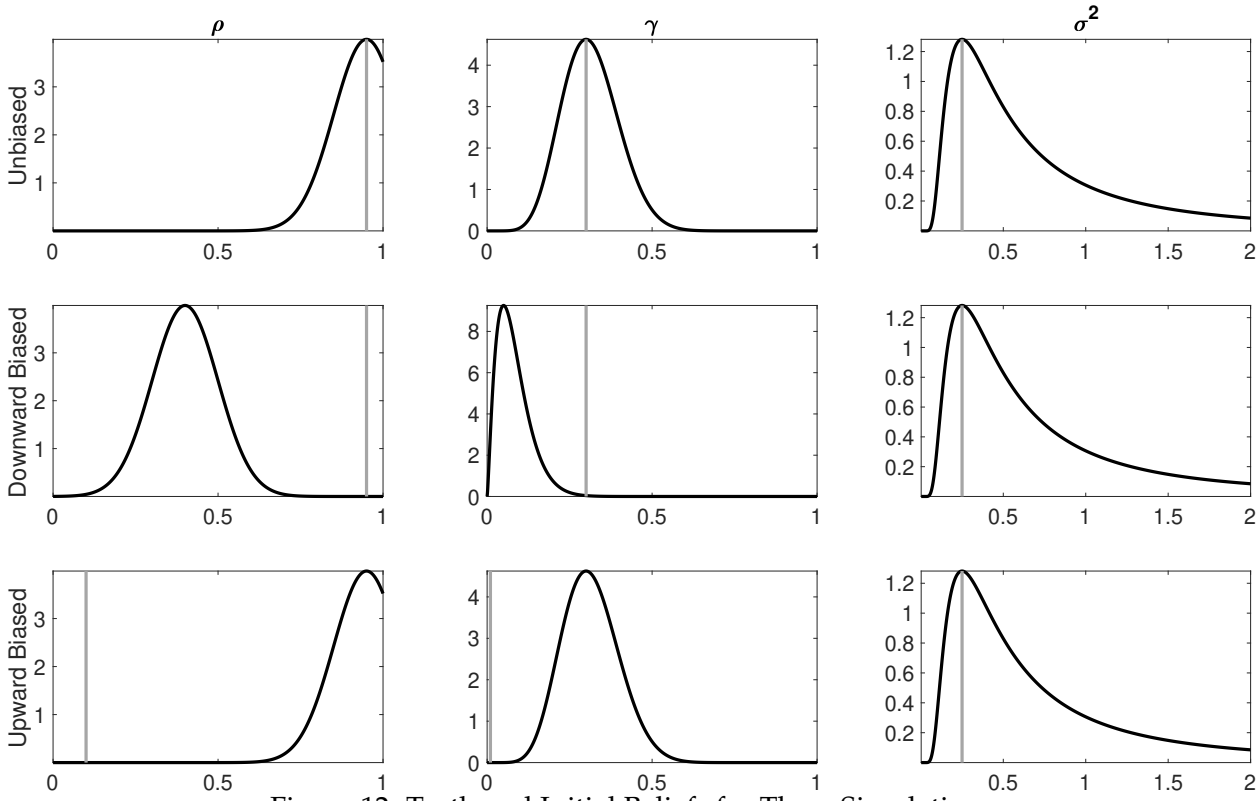

Figure 12: Truth and Initial Beliefs for Three Simulations

Note: The figure plots the truth (gray vertical line) and initial belief distribution (black line) for ρ (left column), γ (middle column), and σ (right column) for the three cases we consider. The first row of figures 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.

Recall that the model we use for the short rate in section 4 is:

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

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

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

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 12 plots the true parameter values (gray vertical 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:

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

– **Downward-Biased Initial Beliefs**: In this case, we again set the true parameters to values $\rho = 0.95$, $\gamma = 0.3$, and $\sigma = 0.5$. We however 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 smaller than the truth:

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

• **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:

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

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 4, i.e., 275 periods corresponding to the sample period from 1951Q2 to 2019Q4. For each of these simulated data series, we then perform the same exercise as we did in section 4. Given their initial beliefs, the agents in the model learn about the parameters of the model using the short rate series and Bayes 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 4.

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

Table 6: Forecast Anomalies in Simulated Data

	Forecast Horizon			
	1	2	3	4
Panel A: Autocorrelation				
Unbiased Initial Beliefs	0.01 (0.08) 1.00	0.00 (0.09) 1.00	-0.00 (0.11) 0.99	-0.01 (0.13) 0.84
Downward-Biased Initial Beliefs	0.16 (0.09) 0.93	0.19 (0.10) 0.78	0.19 (0.12) 0.61	0.18 (0.14) 0.33
Upward-Biased Initial Beliefs	-0.34 (0.06) 1.00	-0.32 (0.07) 1.00	-0.28 (0.08) 1.00	-0.26 (0.08) 1.00
Panel B: Mincer-Zarnowitz				
Unbiased Initial Beliefs	0.96 (0.03) 0.58	0.92 (0.05) 0.62	0.88 (0.08) 0.57	0.83 (0.11) 0.53
Downward-Biased Initial Beliefs	0.98 (0.03) 0.27	0.95 (0.05) 0.40	0.90 (0.08) 0.45	0.85 (0.11) 0.47
Upward-Biased Initial Beliefs	0.37 (0.17) 1.00	0.33 (0.22) 1.00	0.34 (0.25) 1.00	0.35 (0.27) 0.99
Panel C: Coibion-Gorodnichenko				
Unbiased Initial Beliefs	0.01 (0.09) 0.99	0.01 (0.12) 0.99	0.01 (0.15) 1.00	–
Downward-Biased Initial Beliefs	0.18 (0.11) 0.66	0.32 (0.19) 0.55	0.41 (0.25) 0.79	–
Upward-Biased Initial Beliefs	-0.52 (0.10) 1.00	-0.55 (0.13) 1.00	-0.53 (0.17) 1.00	–

Note: The top number for each case is the mean estimate across simulations. The middle number for each case (in parentheses) is the standard deviation across simulations. The bottom number for each case is the fraction of simulations that give a smaller estimate than the real-world data.

underreaction, while Table 7 presents results on the two tests of the expectations hypothesis we consider in section 3. In each case, we report three statistics. The first is the mean estimated coefficient across the 500 simulations; the second statistic is the standard deviation of the estimated effects across simulations (in parentheses); and the third statistic 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

Table 7: Failures of the Expectations Hypothesis in Simulated Data

	Long Horizon n						
	2	3	4	8	12	20	40
Panel A: Future Short Rates							
Unbiased Initial Beliefs	0.95	1.01	1.05	1.19	1.31	1.51	2.06
	(0.64)	(0.63)	(0.66)	(0.72)	(0.76)	(0.82)	(1.03)
	0.07	0.07	0.08	0.12	0.16	0.16	0.08
Downward-Biased Initial Beliefs	0.17	0.20	0.23	0.33	0.42	0.57	0.97
	(0.19)	(0.21)	(0.22)	(0.29)	(0.33)	(0.40)	(0.56)
	0.17	0.30	0.38	0.57	0.65	0.66	0.29
Upward-Biased Initial Beliefs	2.46	2.14	1.97	1.71	1.64	1.59	1.50
	(0.21)	(0.16)	(0.14)	(0.09)	(0.07)	(0.06)	(0.05)
	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Panel B: Change in Long Rate							
Unbiased Initial Beliefs	0.90	0.93	0.95	1.01	1.08	1.20	2.08
	(1.27)	(1.32)	(1.36)	(1.50)	(1.63)	(1.92)	(3.00)
	0.07	0.08	0.07	0.06	0.05	0.03	0.03
Downward-Biased Initial Beliefs	-0.66	-0.69	-0.74	-1.03	-1.39	-2.04	-3.62
	(0.38)	(0.40)	(0.42)	(0.52)	(0.64)	(0.91)	(1.84)
	0.17	0.28	0.24	0.32	0.34	0.51	0.66
Upward-Biased Initial Beliefs	3.91	4.13	4.38	5.59	6.90	9.56	13.77
	(0.42)	(0.42)	(0.42)	(0.51)	(0.62)	(0.86)	(1.62)
	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Note: The top number for each case is the mean estimate across simulations. The middle number for each case (in parentheses) is the standard deviation across simulations. The bottom number for each case is the fraction of simulations that give a smaller estimate than the real-world data.

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 one 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 one 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

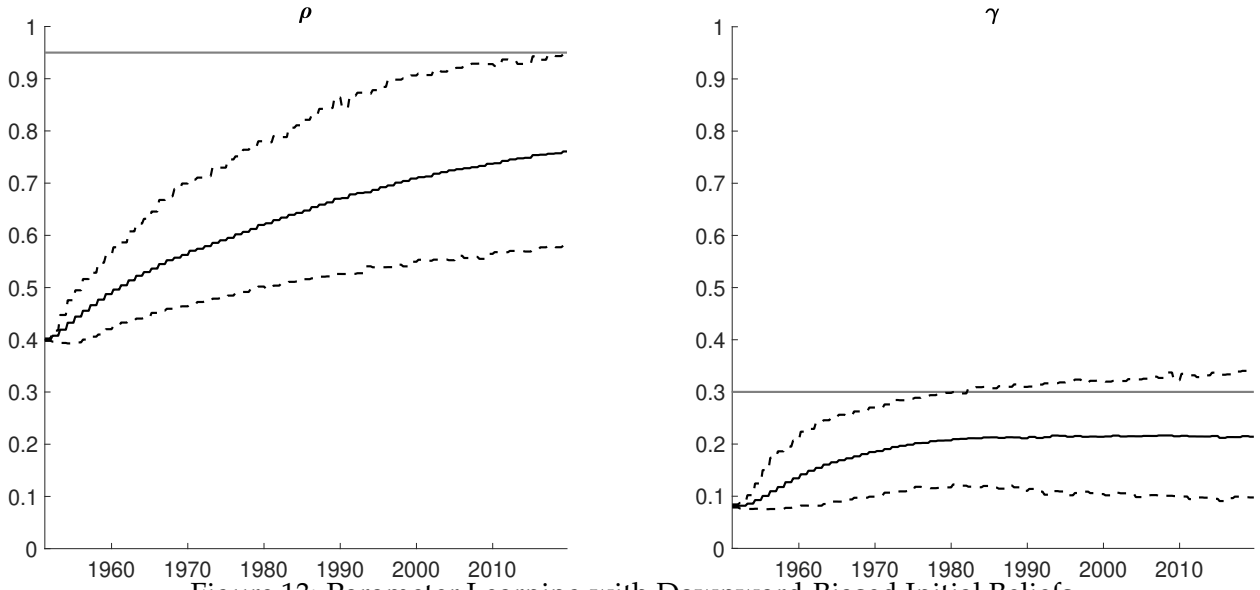

Figure 13: Parameter Learning with Downward-Biased Initial Beliefs

Note: The figure plots the evolution of beliefs about ρ (left panel) and γ (right panel) over time when agents start off with downward-biased initial beliefs. The solid gray line is the truth. The solid black line is the evolution over time of the 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 broken black lines plot the evolution of the 90% and 10% quantiles of the distribution of point estimates across simulations.

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 one at longer horizons. But the value one 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 3. As we discuss earlier in the paper, such beliefs seem reasonable given the prior history of interest rate movements. Outside of war, the U.S. 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.

It is instructive to consider the speed of learning about the key parameters ρ and γ in the sim-

ulations with downward-biased initial beliefs. Figure 13 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 1951Q2 to 2019Q4. 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 13 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 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 14 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 – i.e., we turn off variation in μ_t and learning about γ. In this case, learning about ρ is much quicker.

Figure 14 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 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 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.

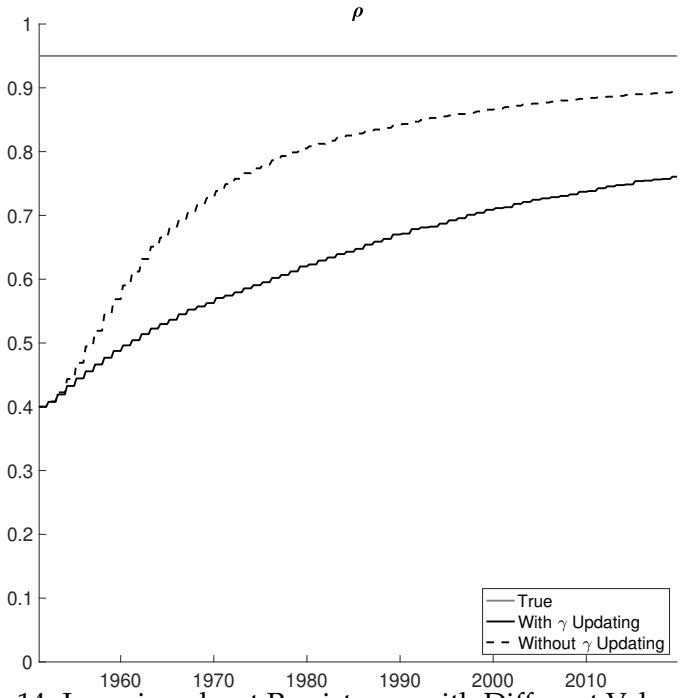

Figure 14: Learning about Persistence with Different Values for γ

Note: 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 agents have very tight initial beliefs around the true value of γ (dashed black line). The solid gray line is the truth.

7 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 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, e.g., it is hard to

reject a unit root in many macroeconomic settings.

A Bayesian Updating about Parameters and States for Interest Rates

In this appendix, we describe in more detail the Gibbs Sampling algorithm we use to sample from the joint posterior of the model parameters and latent states in the UC model for the short rate. Define the vector of the parameters and latent states of the model through date t as $\boldsymbol{\theta} \equiv (\rho, \gamma, \sigma, \boldsymbol{\mu}_{1:t}, \boldsymbol{x}_{1:t})'$. Let $p(\boldsymbol{\theta})$ denote the joint prior over the parameter vector $\boldsymbol{\theta}$. Let $L(\boldsymbol{y}_{1:t}|\boldsymbol{\theta})$ denote the likelihood function of the data through time t, given a set of parameters $\boldsymbol{\theta}$. Our goal is to sample from the posterior of the parameters given the data, $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t})$, where we know

$$p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t}) \propto L(\boldsymbol{y}_{1:t}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

We assume functional forms for the initial beliefs as follows

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

The initial beliefs for the states are given by

$$\begin{aligned}
\mu_{1951Q2} &\sim N(y_{1951Q2}, 1) \\
x_{1951Q2} &\sim N(0, 1)
\end{aligned}$$

where y_{1952Q2} denotes the 3-month Treasury bill rate in 1952Q2.

We start with an initial guess of the parameters $\boldsymbol{\theta}^{(0)} = \left(\rho^{(0)}, \gamma^{(0)}, \sigma^{(0)}, \boldsymbol{\mu}_{1:t}^{(0)}, \boldsymbol{x}_{1:t}^{(0)}\right)'$. Given a draw of the parameters $\boldsymbol{\theta}^{(b)}$, we draw $\boldsymbol{\theta}^{(b+1)}$ as follows:

1. Draw $\rho^{(b+1)}|\gamma^{(b)}, \sigma^{(b)}, \boldsymbol{\mu}_{1:t}^{(b)}, \boldsymbol{x}_{1:t}^{(b)}, \boldsymbol{y}_{1:t}$. Given the other parameters, beliefs about ρ can be updated from the autoregression

$$x_t^{(b)} = \rho x_{t-1}^{(b)} + \sqrt{1 - \gamma^{(b)}}\sigma^{(b)}\omega_t$$

Define

$$\tilde{\sigma}_\rho^2 \equiv \left[\sigma_\rho^{-2} + \frac{\sum_{s=2}^t \left(x_{s-1}^{(b)}\right)^2}{\left(1-\gamma^{(b)}\right)\left(\sigma^{(b)}\right)^2}\right]^{-1}$$
$$\tilde{\mu}_\rho \equiv \tilde{\sigma}_\rho^2 \left[\frac{\mu_\rho}{\sigma_\rho^2} + \frac{\sum_{s=2}^t x_{s-1}^{(b)} x_s^{(b)}}{\left(1-\gamma^{(b)}\right)\left(\sigma^{(b)}\right)^2}\right]$$

The posterior of ρ is $N(\tilde{\mu}_\rho, \tilde{\sigma}_\rho^2)$ and thus we draw $\rho^{(b+1)} \sim N(\tilde{\mu}_\rho, \tilde{\sigma}_\rho^2)$.

2. Draw $\gamma^{(b+1)}|\rho^{(b+1)}, \sigma^{(b)}, \boldsymbol{\mu}_{1:t}^{(b)}, \boldsymbol{x}_{1:t}^{(b)}, \boldsymbol{y}_{1:t}$. There is no closed form expression for the posterior of γ. We therefore draw it using a random walk Metropolis-Hastings step. Specifically, we draw a proposal $\tilde{\gamma}^{(b+1)} \sim N(\gamma^{(b)}, \sigma_{\gamma,prop}^2)$ where $\sigma_{\gamma,prop}^2$ is a proposal variance chosen such that this step has between a 25 and 40% acceptance rate over the burn-in period. We then set $\gamma^{(b+1)} = \tilde{\gamma}^{(b+1)}$ with probability α_{b+1}, where

$$\alpha_{b+1} \equiv \frac{L\left(\boldsymbol{y}_{1:t}|\rho^{(b+1)}, \tilde{\gamma}^{(b+1)}, \sigma^{(b)}, \boldsymbol{\mu}_{1:t}^{(b)}, \boldsymbol{x}_{1:t}^{(b)}\right) p_\gamma\left(\tilde{\gamma}^{(b+1)}\right)}{L\left(\boldsymbol{y}_{1:t}|\rho^{(b+1)}, \gamma^{(b)}, \sigma^{(b)}, \boldsymbol{\mu}_{1:t}^{(b)}, \boldsymbol{x}_{1:t}^{(b)}\right) p_\gamma\left(\gamma^{(b)}\right)}$$

Otherwise we set $\gamma^{(b+1)} = \gamma^{(b)}$.

3. Draw $\sigma^{(b+1)}|\rho^{(b+1)}, \gamma^{(b+1)}, \boldsymbol{\mu}_{1:t}^{(b)}, \boldsymbol{x}_{1:t}^{(b)}, \boldsymbol{y}_{1:t}$. Given the other parameters, beliefs about σ can be updated from the two equations

$$\mu_t^{(b)} = \mu_{t-1}^{(b)} + \sqrt{\gamma^{(b+1)}}\sigma\eta_t$$
$$x_t^{(b)} = \rho^{(b+1)} x_{t-1}^{(b)} + \sqrt{1-\gamma^{(b+1)}}\sigma\omega_t$$

Since η_t and ω_t are independent, these regression equations can be treated as two independent sources of information for σ^2. It is as if beliefs about σ^2 are first updated using information about $\{\eta_s\}_{s=2}^t$ where $\sigma\eta_s = \frac{\mu_s - \mu_{s-1}}{\sqrt{\gamma}}$ and then updated using information about $\{\omega_s\}_{s=2}^t$ where $\sigma\omega_s = \frac{x_s - \rho x_{s-1}}{\sqrt{1-\gamma}}$. These are each samples of $t-1$ observations which can be used to

learn about σ^2 using standard conjugate prior updating. Define

$$\tilde{\alpha}_{\sigma^2} \equiv \alpha_{\sigma^2} + (t-1)$$
$$\tilde{\beta}_{\sigma^2} \equiv \beta_{\sigma^2} + \frac{\sum_{s=2}^{t}\left(\mu_s^{(b)} - \mu_{s-1}^{(b)}\right)^2}{2\gamma^{(b+1)}} + \frac{\sum_{s=2}^{t}\left(x_s^{(b)} - \rho^{(b+1)} x_{s-1}^{(b)}\right)^2}{2\left(1-\gamma^{(b+1)}\right)}$$

The posterior of σ^2 is $\mathcal{IG}(\tilde{\alpha}_{\sigma^2}, \tilde{\beta}_{\sigma^2})$ and thus we draw $\left(\sigma^{(b)}\right)^2 \sim \mathcal{IG}(\tilde{\alpha}_{\sigma^2}, \tilde{\beta}_{\sigma^2})$.

4. Draw $\boldsymbol{\mu}_{1:t}^{(b+1)}, \boldsymbol{x}_{1:t}^{(b+1)} | \rho^{(b+1)}, \gamma^{(b+1)}, \sigma^{(b+1)}, \boldsymbol{y}_{1:t}$. This can be done using the standard Kalman filter and simulation smoother.

This algorithm is repeated to produce B draws from the posterior distribution of the parameters and states at each time t.

B Bayesian Forecasting of Interest Rates

The algorithm described in Appendix A yields B samples of the posterior of the states and parameters of our UC model at each point in time t. We index these samples by b as follows $\left\{\rho^{(b)}, \gamma^{(b)}, \sigma^{(b)}, \boldsymbol{\mu}_{1:t}^{(b)}, \boldsymbol{x}_{1:t}^{(b)}\right\}_{b=1}^{B}$. Draws b for which $\rho^{(b)} > 1$ are not used for forecasting as they imply explosive dynamics in interest rates. We then use the following algorithm to produce a real-time forecast distribution for the yield curve at time t:

1. For each $b = 1, \ldots, B$

   (a) Simulate a path of shocks $\left\{\eta_{t+h}^{(b)}, \omega_{t+h}^{(b)}\right\}_{h=1}^{H}$ from the standard Normal distribution.

   (b) Starting from $h = 1$, construct a simulated path of the states over H subsequent periods using equations (7)-(8):

$$\mu_{t+h}^{(b)} = \mu_{t+h-1|t}^{(b)} + \sqrt{\gamma^{(b)}}\sigma^{(b)}\eta_{t+h}^{(b)}$$
$$x_{t+h}^{(b)} = \rho^{(b)} x_{t+h-1|t}^{(b)} + \sqrt{1-\gamma^{(b)}}\sigma^{(b)}\omega_{t+h}^{(b)}$$

   (c) Use the simulated states to construct the forecast distribution of the short rate $\left\{y_{t+h|t}^{(b)}\right\}_{h=1}^{H}$ where

$$y_{t+h|t}^{(b)} = \mu_{t+h|t}^{(b)} + x_{t+h|t}^{(b)}$$

2. The forecast of y_{t+h} given time t information is computed as

$$F_t y_{t+h} = \frac{1}{B} \sum_{b=1}^{B} y_{t+h|t}^{(b)}$$

The implied yield of a bonds of maturity n is given by

$$y_t^{(n)} = c^{(n)} + \frac{1}{n} \sum_{h=0}^{n-1} F_t y_{t+h}$$

We estimate the constant $c^{(n)}$ as the average level of the corresponding n-period bond yield in the data since it is not identified from the expectations hypothesis alone. Note that this estimate of the constant does not affect the results of the expectations hypothesis regression tests we run since it only affects the level of the n-period yield.

At the end of the estimation we are left with a sequence of model-implied 1 to H-quarter ahead forecasts $\{F_t y_{t+h}\}_{h=1}^{H}$ and model-implied yields $\left\{y_t^{(h)}\right\}_{h=1}^{H}$ for every quarter t from 1961Q3 to 2019Q4.

C Search over Initial Beliefs for Nominal Short Rate

Let $\boldsymbol{\theta} = (\alpha_\rho, \beta_\rho, \alpha_\gamma, \beta_\gamma)'$. Let $\boldsymbol{\alpha} = \{\alpha_h\}_{h=1}^{H}$ and $\boldsymbol{\beta} = \{\beta_h\}_{h=1}^{H}$ denote vectors of estimated coefficients from the forecasting anomaly regressions for different horizons up through a maximum horizon of H using the SPF and yield curve data. Let $\hat{\boldsymbol{\alpha}} = \{\hat{\alpha}_h\}_{h=1}^{H}$ and $\hat{\boldsymbol{\beta}} = \left\{\hat{\beta}_h\right\}_{h=1}^{H}$ denote those same quantities estimated on the model implied forecasts and yields for a particular value of $\boldsymbol{\theta}$. Define the moment function as

$$https://www.overleaf.com/project/5fe122a5d2991bb3a6770275 \hat{m}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\alpha}_{bias} - \hat{\boldsymbol{\alpha}}_{bias} \\ \boldsymbol{\beta}_{ar} - \hat{\boldsymbol{\beta}}_{ar} \\ \boldsymbol{\beta}_{mz} - \hat{\boldsymbol{\beta}}_{mz} \\ \boldsymbol{\beta}_{cg} - \hat{\boldsymbol{\beta}}_{cg} \\ \boldsymbol{\beta}_{sr} - \hat{\boldsymbol{\beta}}_{sr} \\ \boldsymbol{\beta}_{lr} - \hat{\boldsymbol{\beta}}_{lr} \end{bmatrix} \quad (17)$$

The parameters are then estimated via the simulated method of moments (SMM) with an iden-

Table C.1: Root-Mean-Squared Errors: Ratio of Model to SPF

	Forecast Horizon				
	1	2	3	4	Mean
Baseline	0.96	0.98	0.97	0.96	0.96
Highly Dispersed	0.95	0.98	0.97	0.97	0.97

Note: The table reports the ratio of the root-mean-squared error (RMSE) of forecasts from our model for interest rates with different initial beliefs to the RMSE of SPF forecasts. The baseline initial beliefs for ρ are $N(0.6, 0.12^2)$. The initial beliefs in the'highly dispersed' case are $N(0.6, 0.31^2)$. The baseline initial beliefs for γ are $\mathcal{B}(2.3, 19.7)$. The initial beliefs in the 'highly dispersed' case are $\mathcal{B}(1.13, 2.9)$. The 'Mean' column reports the average ratio across the four horizons.

tity weighting matrix

$$\hat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}} \left\| \hat{m}(\boldsymbol{\theta}) \right\|_2 = \text{argmax}_{\boldsymbol{\theta}} \hat{m}(\boldsymbol{\theta})' \hat{m}(\boldsymbol{\theta})$$

Every evaluation of the moment function $\hat{m}(\boldsymbol{\theta})$ requires us to sample from the posterior of the UC model sequentially. Since this step is very computationally costly, we only re-estimate the model every 4 quarters rather than every quarter, and use a burn-in sample of 50,000 draws and keep the subsequent 25,000 draws rather than 75,000 for each of those quantities in our empirical specification. The global minimum is found using MATLAB's "particleswarm" optimization routine, subject to the constraint that the mean of ρ is larger than 0.5.

C.1 More Dispersed Initial Beliefs?

Table C.1 reports results on the root-mean-squared error (RMSE) of the forecasts from our model relative to the RMSE of SFP forecasters. We do this for our baseline initial beliefs – $\rho \sim N(0.6, 0.12^2)$ and $\gamma \sim \mathcal{B}(2.3, 19.7)$ – and for a case with more highly dispersed initial beliefs – $\rho \sim N(0.6, 0.31^2)$ and $\gamma \sim \mathcal{B}(1.13, 2.9)$. The first four columns of Table C.1 report the ratio of the RMSE from our model relative to the RMSE of SFP forecasters for forecast horizons of one through four quarters. The last column reports the average ratio across the four horizons.

Several facts are notable. First, the forecasts from our model are slightly 'better' than the SPF forecasts in that their RMSE is slightly lower than the RMSE of the SPF forecasts (the ratios reported in the table are all smaller than 1). Second, using more highly dispersed initial beliefs actually leads to very slightly worse forecasts from a RMSE perspective at most horizons: the 'highly dispersed' case generates slightly larger RMSE on average and for horizons 2 through 4. However, the differences are very small. Overall, we conclude that a model with more dispersed initial beliefs generates forecasts of very similar quality as measured by RMSE.

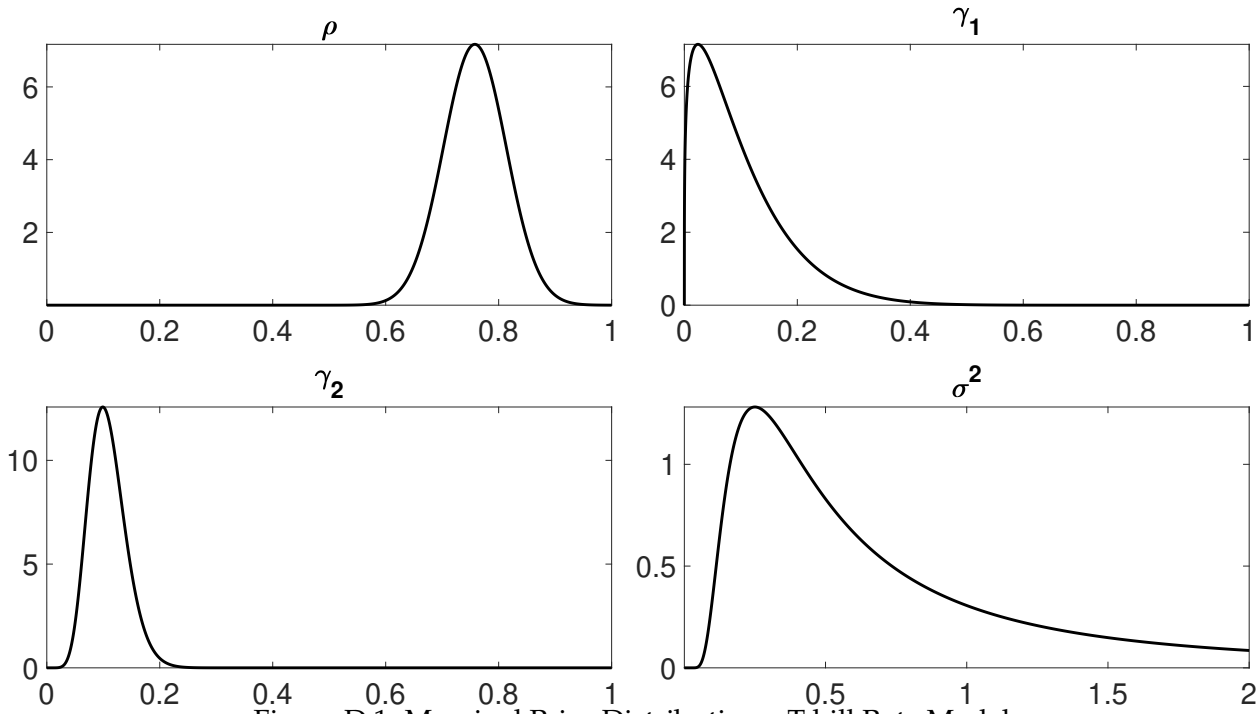

Figure D.1: Marginal Prior Distributions: T-bill Rate Model

Note: These four panels plot the initial beliefs we estimate for the four parameters of the model. The panels labelled ρ, γ_1, and σ give the initial beliefs for ρ, γ, and σ, respectively, in 1951Q2. The panel labelled γ_2 give the belief distribution for γ in 1990Q1.

D Interest Rate Results with a Break in 1990

Here we present results for a case where we allow for a break in beliefs about γ in 1990. 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.

Results analogous to those presented for our baseline model in the main body are presented in Figures D.1-D.4 and Tables D.2-D.3. We estimate a substantial decrease in the mean of the distribution of beliefs about γ (from 0.19 to 0.11) and a sharp downward shift in the standard deviation (from 0.09 to 0.03) in 1990, leading to lower posterior mean estimates in the latter part of the sample as one would expect. Other results are quite similar to in our baseline case. The extra parameters allows up to improve the fit of the model to the data modestly.

Table D.2: T-Bill Rate Forecast Anomalies: Model vs. Data

	Forecast Horizon			
	1	2	3	4
Panel A: Bias				
SPF	-0.18***	-0.34***	-0.52***	-0.70***
	(0.05)	(0.09)	(0.14)	(0.19)
UC Model	-0.18**	-0.32**	-0.47**	-0.60**
	(0.06)	(0.11)	(0.16)	(0.21)
Panel B: Autocorrelation				
SPF	0.30*	0.27**	0.24*	0.13
	(0.14)	(0.12)	(0.12)	(0.13)
UC Model	0.38*	0.41**	0.37**	0.25*
	(0.17)	(0.14)	(0.11)	(0.12)
Panel C: Mincer-Zarnowitz				
SPF	0.97*	0.94**	0.90**	0.86**
	(0.02)	(0.02)	(0.04)	(0.05)
UC Model	0.96*	0.93**	0.89**	0.84**
	(0.02)	(0.03)	(0.04)	(0.05)
Panel D: Coibion-Gorodnichenko				
SPF	0.23*	0.34*	0.62***	–
	(0.12)	(0.16)	(0.16)	
UC Model	0.41*	0.59	0.94*	–
	(0.19)	(0.39)	(0.45)	

Note: The forecast horizons are quarters. Stars represent significance relative to the following hypotheses: $\alpha = 0$ for bias, $\beta = 0$ for autocorrelation, $\beta = 1$ for Mincer-Zarnowitz, $\beta = 0$ 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. This corresponds to a bandwidth of 17. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.3: Failures of the Expectations Hypothesis: Model vs. Data

	Long Horizon n						
	2	3	4	8	12	20	40
Panel A: Future Short Rates							
Data	-0.01***	0.11***	0.18***	0.39**	0.57	0.74	0.71
	(0.23)	(0.23)	(0.23)	(0.23)	(0.26)	(0.23)	(0.20)
UC Model	-0.07***	0.09***	0.15***	0.48	0.72	0.84	0.92
	(0.25)	(0.26)	(0.27)	(0.33)	(0.33)	(0.28)	(0.33)
Panel B: Change in Long Rate							
Data	-1.02***	-0.91***	-1.03***	-1.29***	-1.61***	-2.04***	-2.75***
	(0.45)	(0.59)	(0.62)	(0.59)	(0.57)	(0.55)	(0.87)
UC Model	-1.14***	-1.17***	-1.19***	-1.31***	-1.45***	-1.79***	-2.61**
	(0.50)	(0.51)	(0.52)	(0.55)	(0.59)	(0.68)	(1.24)

Note: The sample period is from 1961Q3 to 2019Q4. The top panel reports estimates of β from regression (4). The bottom panel reports estimates of β from regression (5). In both cases, the horizon n is listed at the top of the table. Stars 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. This corresponds to a bandwidth of 19. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

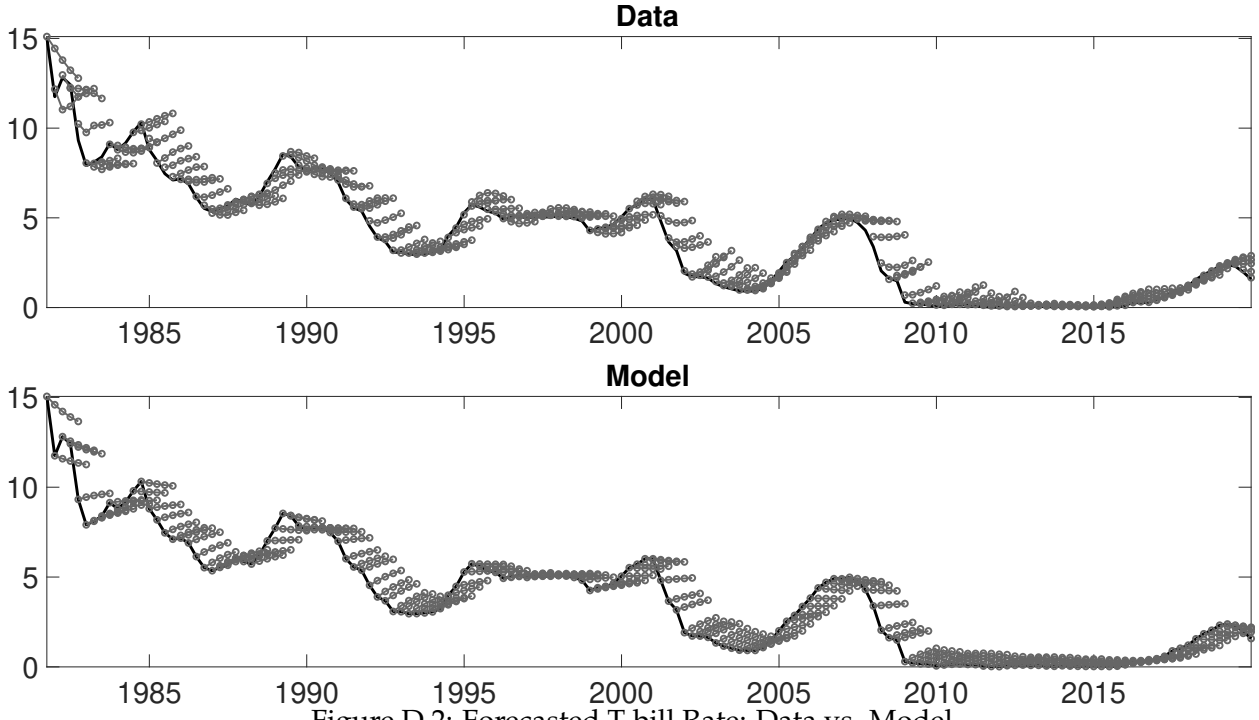

Figure D.2: Forecasted T-bill Rate: Data vs. Model

Note: The black solid 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 UC model estimated in real-time.

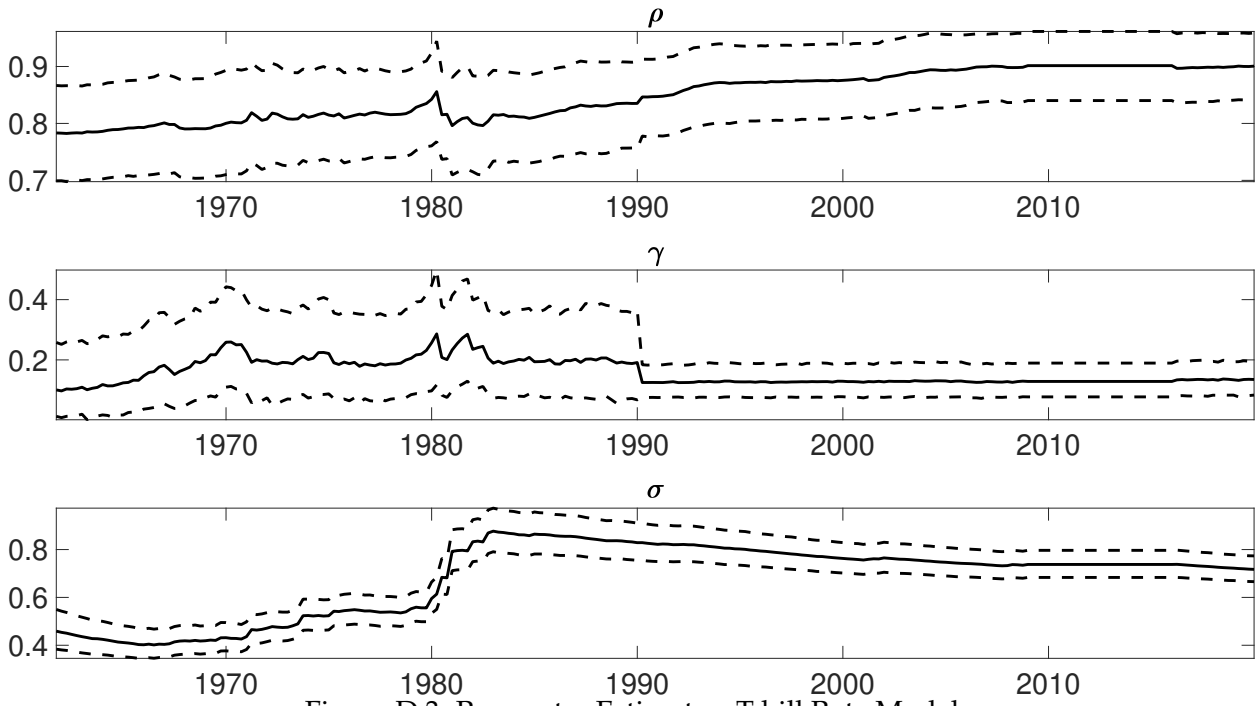

Figure D.3: Parameter Estimates: T-bill Rate Model

Note: Each panel plots the evolution of beliefs about one of the three UC model parameters: ρ, γ, and σ. The black solid line is the mean and the dotted black lines are the 5th and 95th percentiles of the posterior distribution for the parameter in question. Recall that we only update beliefs about these parameters every fourth quarter.

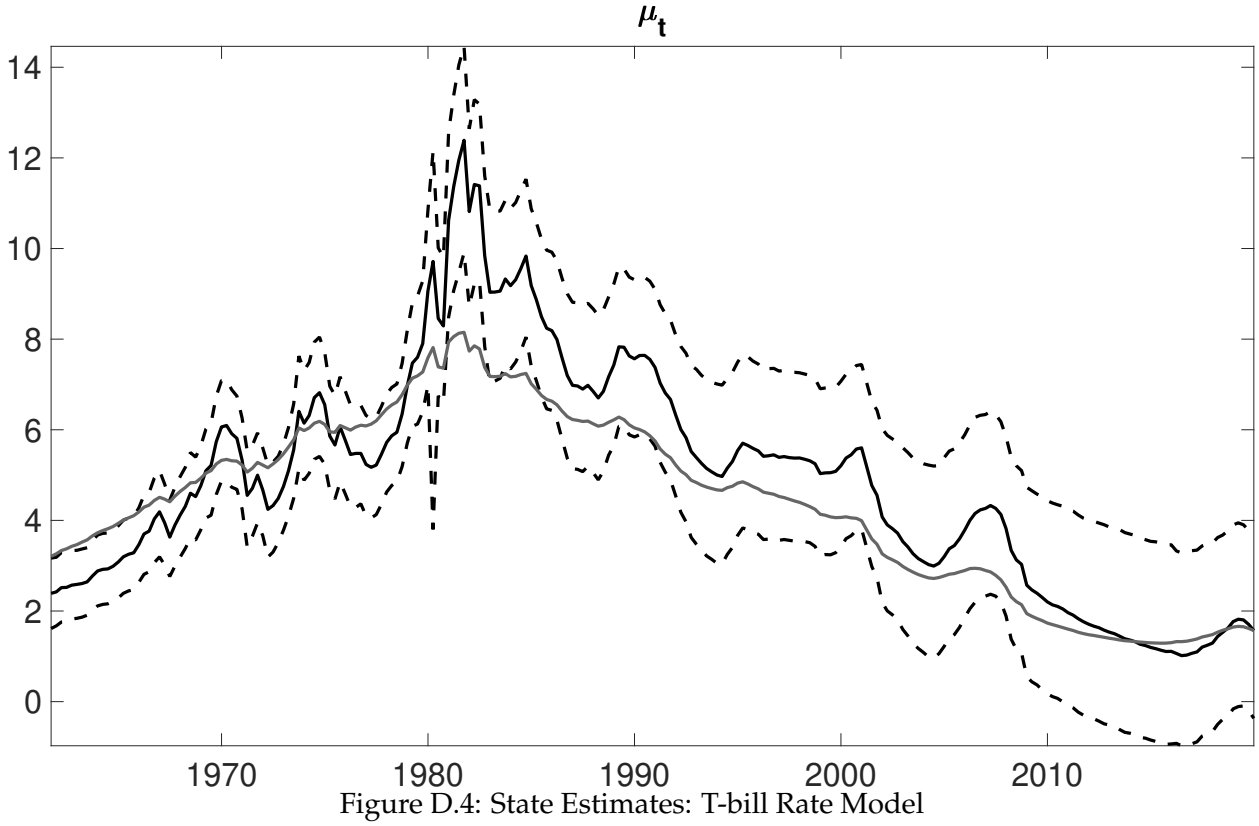

Figure D.4: State Estimates: T-bill Rate Model

Note: Each panel corresponds to one of the two UC hidden state variables: μ_t and x_t respectively. The black solid line is the posterior mean of the real-time filtering distributions, the dotted black lines are the 5th and 95th percentiles of the posterior real-time filtering distributions, and the solid gray line is the posterior mean of the ex-post smoothing distributions for the corresponding parameter.

D.1 Cochrane-Piazzesi Regressions

Cochrane and Piazzesi (2005) show that a single factor predicts one-year excess returns on one- to five-year maturity bonds with an R^2 above 0.4. Here we show that our learning model can match this return predictability. Following Cochrane and Piazzesi (2005), we consider the period length to be measured in years in this section – i.e., n and t are measured in years in this section while it is measured in quarters elsewhere in the paper. Let $p_t^{(n)}$ denote the log price of an n-year zero coupon bond at time t. The relationship between the log yield and log price of an n-year zero coupon bond is $y_t^{(n)} = -p_t^{(n)}/n$. The forward rate at time t for a loan between time $t+n-1$ and time $t+n$ is

$$f_t^{(n-1\to n)} \equiv p_t^{(n)} - p_t^{(n-1)}.$$

We refer to this as the n-year forward rate. (It might alternatively be referred to as the n-year-ahead, one-year forward rate.) The log holding period return of buying an n-year bond at time t and selling it as an $n-1$-year bond at time $t+1$ is given by

$$r_{t+1}^{(n)} \equiv p_{t+1}^{(n-1)} - p_t^{(n)}.$$

Denote the log excess return on the bond as

$$rx_{t+1}^{(n)} \equiv r_{t+1}^{(n)} - y_t^{(1)}.$$

Cochrane and Piazzesi (2005) run two sets of return predictability regressions. First, they run a set of *unrestricted* regressions:

$$rx_{t+1}^{(n)} = \beta_0^{(n)} + \beta_1^{(n)} y_t^{(1)} + \beta_2^{(n)} f_t^{(1\to 2)} + \ldots + \beta_5^{(n)} f_t^{(4\to 5)} + \varepsilon_{t+1}^{(n)} \tag{18}$$

for $n = 2, \ldots, 5$. These regressions yield a similar pattern of coefficients across the four maturities. This motivates considering the notion that a single factor may forecast excess returns at all horizons as follows:

$$rx_{t+1}^{(n)} = b_n \left(\gamma_0 + \gamma_1 y_t^{(1)} + \gamma_2 f_t^{(1\to 2)} + \ldots + \gamma_5 f_t^{(4\to 5)} \right) + \varepsilon_{t+1}^{(n)}$$

for $n = 2, \ldots, 5$. Cochrane and Piazzesi normalize the loadings b_n so they have an average value

Data						
const.	$y^{(1)}$	$f^{(1\to2)}$	$f^{(2\to3)}$	$f^{(3\to4)}$	$f^{(4\to5)}$	$\overline{R}^2$
-0.19	-0.93	-1.25	1.41	1.45	-0.71	0.38
(0.39)	(0.45)	(0.85)	(0.82)	(1.00)	(0.73)	
Model						
const.	$y^{(1)}$	$f^{(1\to2)}$	$f^{(2\to3)}$	$f^{(3\to4)}$	$f^{(4\to5)}$	$\overline{R}^2$
-2.67	-21.13	82.60	-112.64	53.76	-2.54	0.48
(1.82)	(7.31)	(34.02)	(51.08)	(37.28)	(17.71)	

Table D.4: Restricted Regression Results for γ_n

of 1:

$$\frac{1}{4}\sum_{n=2}^{5} b_n = 1$$

and estimate the b_n and γ_n coefficients in two stages. First, they estimate the γ_n coefficients from

$$\frac{1}{4}\sum_{n=2}^{5} rx_{t+1}^{(n)} = \gamma_0 + \gamma_1 y_t^{(1)} + \gamma_2 f_t^{(1\to2)} + \ldots + \gamma_5 f_t^{(4\to5)} + \overline{\varepsilon}_{t+1}. \tag{19}$$

Then they estimate the b_n coefficients from

$$rx_{t+1}^{(n)} = b_n\left(\hat{\gamma}_0 + \hat{\gamma}_1 y_t^{(1)} + \hat{\gamma}_2 f_t^{(1\to2)} + \ldots + \hat{\gamma}_5 f_t^{(4\to5)}\right) + \varepsilon_{t+1}^{(n)}, \tag{20}$$

where $\hat{\gamma}_n$ are the fitted values from regression (19). We refer to these as the *restricted* regressions.

We perform this analysis on our quarterly zero-coupon bond data from Liu and Wu (2020) for the sample period 1961Q3-2019Q4. (Cochrane and Piazzesi (2005) use monthly data for the sample period 1964-2003.) We also perform this analysis on the bond yields implied by our model (allowing for a break in γ in 1990). Before running the regressions for the model-implied data, the sample mean of our model-implied yields is made the same as the sample mean of the yields in the real-world data over our sample period.

Tables D.4 and D.5 report the coefficients [γ_1, ... γ_5] and [b_2, ... b_5] from the restricted regressions (equations (19) and (20)) respectively, along with the adjusted R^2 for these regressions. We can match the high R^2 for one-year excess returns on 2- to 5-year zero coupon bonds in data from our model: the R^2 for these predictive regressions on data from our model are between 0.46 and 0.50. This finding of high predictability is quite robust across the different variants of our model we

Data			
n	b_n	SE	$\overline{R}^2$
2	0.58	(0.07)	0.35
3	0.93	(0.08)	0.41
4	1.18	(0.09)	0.40
5	1.31	(0.10)	0.37
Model			
n	b_n	SE	$\overline{R}^2$
2	0.60	(0.05)	0.47
3	0.91	(0.07)	0.50
4	1.16	(0.08)	0.50
5	1.33	(0.10)	0.49

Table D.5: Restricted Regression Results for b_n

have considered. The shape of the factor that predicts returns is much more sensitive to model specification, due to the high correlations between yields at various maturities generated by our model.

D.2 Giglio-Kelly Volatility Ratio

Giglio and Kelly (2018) argue that long-maturity assets display excess volatility that is hard to reconcile with standard asset pricing models. One of their applications is to the term structure of interest rates. Our exposition of their methods builds on the discussion in Liu and Wu (2020). Consider an affine term structure model in which the short rate is affine in a multi-variate latent factor x_t:

$$y_t = \delta_0 + \delta_1' x_t,$$

where $\delta_1 = [1, 1, 1]'$. Assume that the latent factor evolves according to a VAR(1) under the risk-neutral measure $\mathbb{Q}$:

$$x_t = \rho^{\mathbb{Q}} x_{t-1} + \epsilon_t^{\mathbb{Q}},$$

where $\rho^{\mathbb{Q}}$ is diagonal. In this case, the log-price of an n-period zero coupon bond is affine in x_t:

$$p_t^{(n)} = a_n + b_n' x_t, \tag{21}$$

where the coefficients b_n may be calculated recursively. Hamilton and Wu (2012) provide a deriva-

tion.

Given this setup, $\rho^{\mathbb{Q}}$ can be estimated by, first, running a regression

$$p_t^{(h)} = \alpha_h + \beta_h' P_t + u_t^{(h)},$$

where P_t is a three-dimensional vector of bond prices $P_t = [p_t^{(i)}, p_t^{(j)}, p_t^{(k)}]$. The three unknown parameters in $\rho^{\mathbb{Q}}$ can then be recovered from the following three equations

$$\hat{\beta}_h = [b_i, b_j, b_k]^{-1} b_h, \tag{22}$$

where the $b's$ in this equation are the loadings on the factor x_t in equation (21). This procedure should recover $\rho^{\mathbb{Q}}$ for any i, j, k, h. We follow Giglio and Kelly (2018) and regress the 7-year zero coupon bond on the 1-year, 3-year, and 5-year zero coupon bonds.

This affine term structure model with the estimate $\hat{\rho}^{\mathbb{Q}}$ has strong implications regarding the volatility of longer term bond prices. In particular, one can calculate the β_m implied by the model for m-period zero coupon bonds. We call these $\hat{\beta}_m^R$. The model then implies that the volatility of the price of an m-period zero coupon bond should be $V_m^R = \hat{\beta}_m^{R\prime} \hat{V}(P_t) \hat{\beta}_m^R$, where $\hat{V}(P_t)$ is the sample variance-covariance matrix of P_t. This model implied volatility of the price of the m-period zero coupon bond can be compared to the volatility implied by running an unrestricted OLS regression of the price of the m-period zero coupon bond on P_t. We denote the estimated coefficients from this regression as $\hat{\beta}_m^U$ and the implied volatility $V_m^U = \hat{\beta}_m^{U\prime} \hat{V}(P_t) \hat{\beta}_m^U$. Giglio and Kelly (2018) focus attention on the volatility ratio V_m^U / V_m^R.

Figure D.5 plots the V_m^R (orange line), V_m^U (blue line) and reports the ratio between them (numbers) for zero-coupon bonds with maturity between 10 years and 30 years. The left panel reports these statistics estimated on the real-world data, while the right panel reports these statistics estimate on the zero-coupon bond prices implied by our model (with a break in 1990). We see that our model generates substantially higher V_m^U than V_m^R as does the data.

For the term structure of interest rates, the excess volatility fact that Giglio and Kelly (2018) focus on turns out to be rather sensitive to the details of how the affine term structure model is estimated. If we use a regression of the 5-year zero coupon bond on the 1-year, 2-year, and 3-year zero coupon bonds to estimate $\rho^{\mathbb{Q}}$ – rather than a regression of the 7-year zero coupon bond on the 1-year, 3-year, and 5-year zero coupon bonds – and then calculate V_m^R and V_m^U we get dramatically different results. These results are plotted in Figure D.6. In this case, V_m^R is much larger than

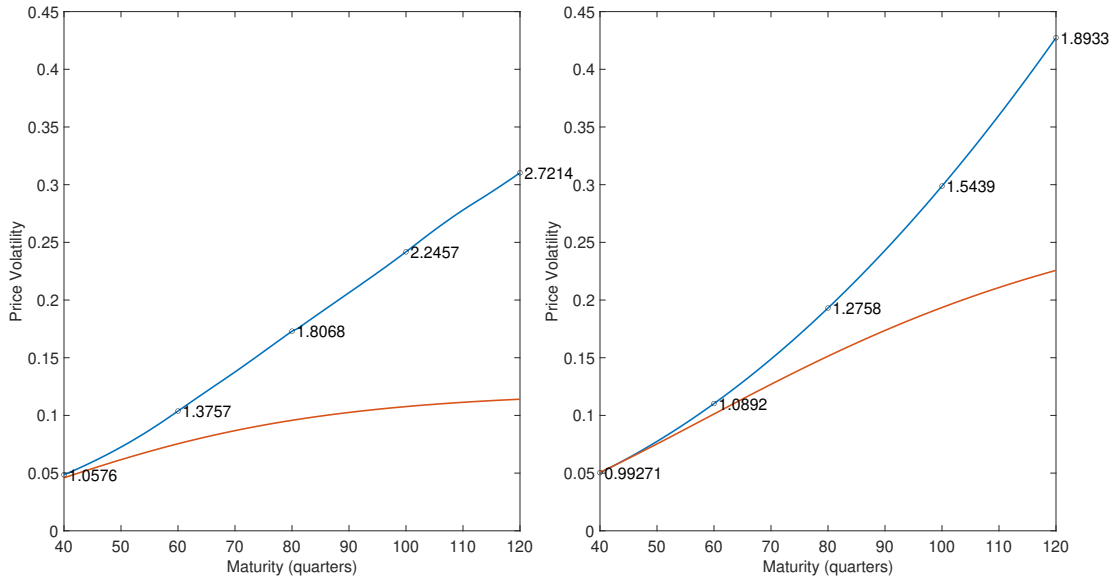

Figure D.5: Giglio-Kelly Variance Ratios: Data and Model

V_m^U and the variance ratio is therefore less than one. The reason for this is that the largest root – i.e., the largest element of $\rho^{\mathbb{Q}}$ – is estimated to be slightly larger than one (1.012 to be precise) in this case, while it is estimated to be slightly smaller than one (0.958 to be precise) in the case Giglio and Kelly (2018) focus on. The dynamics of x_t are therefore stationary in the case Giglio and Kelly (2018) focus on, but very close to being non-stationary, while they are estimated to be non-stationary in the alternative case presented in Figure D.6.

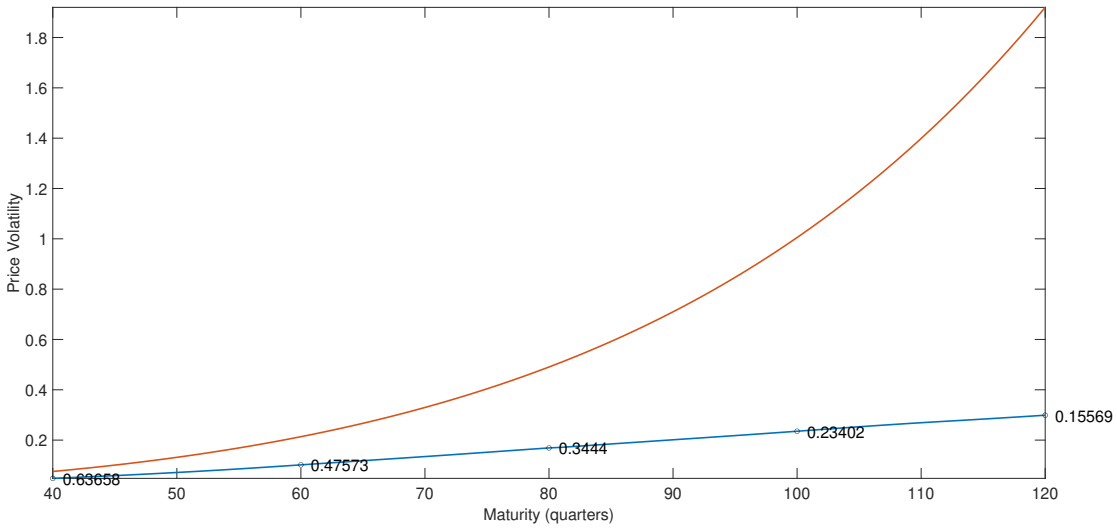

Figure D.6: Giglio-Kelly Variance Ratios for Alternative Basis

E Bayesian Updating about Parameters and States for GDP

Here we describe the initial beliefs and sampling algorithm for our GDP application. We assume that the initial belief of the CBO about the mean of the difference stationary component μ is Normal,

$$\mu \sim N(\mu_\mu, \sigma_\mu^2).$$

We assume that the CBO has independent Normal initial beliefs about the sum of the autoregressive parameters $\rho_1 + \rho_2$ and for the second autoregressive parameter ρ_2. We truncate these initial belief distributions in such a way as to put zero weight on parameter combinations that result in the x_t component being non-stationary. We can write these initial belief distributions as

$$\begin{aligned}
\rho_1 + \rho_2 &\sim N(\mu_\rho, \sigma_\rho^2)\mathcal{I}(\rho_1, \rho_2), \\
\rho_2 &\sim N(\mu_{\rho_2}, \sigma_{\rho_2}^2)\mathcal{I}(\rho_1, \rho_2).
\end{aligned}$$

where $\mathcal{I}(\rho_1, \rho_2)$ is an indicator variable which is 1 for (ρ_1, ρ_2) combinations that result x_t being stationary and 0 otherwise.

This implies a joint initial belief distribution for ρ_1, ρ_2 the moments of which are

$$\begin{aligned}
\mu_{\rho_1} &= \mu_\rho - \mu_{\rho_2}, \\
\sigma_\rho^2 &= \sigma_{\rho_1}^2 + \sigma_{\rho_2}^2 + 2\sigma_{\rho_1,\rho_2}, \\
\sigma_{\rho,\rho_2} &= \sigma_{\rho_1,\rho_2} + \sigma_{\rho_2}^2 = 0, \\
\sigma_{\rho_1,\rho_2} &= -\sigma_{\rho_2}^2, \\
\sigma_{\rho_1}^2 &= \sigma_\rho^2 + \sigma_{\rho_2}^2.
\end{aligned}$$

In other words,

$$\begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_\rho - \mu_{\rho_2} \\ \mu_{\rho_2} \end{bmatrix}, \begin{bmatrix} \sigma_\rho^2 + \sigma_{\rho_2}^2 & -\sigma_{\rho_2}^2 \\ -\sigma_{\rho_2}^2 & \sigma_{\rho_2}^2 \end{bmatrix} \right) \mathcal{I}(\rho_1, \rho_2).$$

We assume that the CBOs initial belief distribution about the the variance share γ of shocks to the trend component is a Beta distribution,

$$\gamma \sim \mathcal{B}(\alpha_\gamma, \beta_\gamma).$$

We assume that the CBOs initial belief distribution about the conditional variance σ^2 is an Inverse Gamma distribution,

$$\sigma^2 \sim \mathcal{IG}(\alpha_{\sigma^2}, \beta_{\sigma^2}).$$

Lastly, we assume that agents' initial beliefs about z_t and x_t in 1959Q3 are $z_t \sim N(y_{1959Q3}, 0.01^2)$ and $x_t \sim N(0, 0.01^2)$.

We start with an initial guess of the unknown parameters

$$\boldsymbol{\theta}^{(0)} = \left(\mu^{(0)}, \rho_1^{(0)}, \rho_2^{(0)}, \gamma^{(0)}, \sigma^{(0)}, \boldsymbol{z}_{1:t}^{(0)}, \boldsymbol{x}_{1:t}^{(0)}\right)'.$$

Given a draw of the parameters $\boldsymbol{\theta}^{(b)}$, we draw $\boldsymbol{\theta}^{(b+1)}$ as follows:

1. Draw $\mu^{(b+1)}|\rho_1^{(b)}, \rho_2^{(b)}, \gamma^{(b)}, \sigma^{(b)}, \boldsymbol{z}_{1:t}^{(b)}, \boldsymbol{x}_{1:t}^{(b)}, \boldsymbol{y}_{1:t}$. Given the other parameters, beliefs about μ can be updated from the equation for z_t:

$$\Delta z_t^{(b)} = \mu + \sqrt{\gamma^{(b)}}\sigma^{(b)} u_t.$$

   Define

$$\tilde{\sigma}_\mu^2 \equiv \left[\sigma_\mu^{-2} + \frac{t-1}{\gamma^{(b)}\left(\sigma^{(b)}\right)^2}\right]^{-1},$$
$$\tilde{\mu}_\mu \equiv \tilde{\sigma}_\mu^2 \left[\frac{\mu_\mu}{\sigma_\mu^2} + \frac{\sum_{s=2}^t \Delta z_{s-1}^{(b)}}{\gamma^{(b)}\left(\sigma^{(b)}\right)^2}\right].$$

   The posterior of μ is $N(\tilde{\mu}_\mu, \tilde{\sigma}_\mu^2)$ and thus we draw $\mu^{(b+1)} \sim N(\tilde{\mu}_\mu, \tilde{\sigma}_\mu^2)$.

2. Draw $\rho_1^{(b+1)}, \rho_2^{(b+1)}|\mu^{(b+1)}, \gamma^{(b)}, \sigma^{(b)}, \boldsymbol{z}_{1:t}^{(b)}, \boldsymbol{x}_{1:t}^{(b)}, \boldsymbol{y}_{1:t}$. Given the other parameters, beliefs about ρ_1, ρ_2 can be updated from the equation for x_t:

$$x_t^{(b)} = \rho_1 x_{t-1}^{(b)} + \rho_2 x_{t-2}^{(b)} + \sqrt{(1-\gamma^{(b)})}\sigma^{(b)} v_t.$$

   Define

$$\tilde{\boldsymbol{\Sigma}}_\rho \equiv \left[\boldsymbol{\Sigma}_\rho^{-1} + \frac{\sum_{s=3}^t [x_{s-1}^{(b)}, x_{s-2}^{(b)}]'[x_{s-1}^{(b)}, x_{s-2}^{(b)}]}{(1-\gamma^{(b)})\left(\sigma^{(b)}\right)^2}\right]^{-1},$$
$$\tilde{\boldsymbol{\mu}}_\rho \equiv \tilde{\boldsymbol{\Sigma}}_\rho \left[\boldsymbol{\Sigma}_\rho^{-1}\boldsymbol{\mu}_\rho + \frac{\sum_{s=3}^t [x_{s-1}^{(b)}, x_{s-2}^{(b)}]' x_s^{(b)}}{(1-\gamma^{(b)})\left(\sigma^{(b)}\right)^2}\right].$$

The posterior of $(\rho_1, \rho_2)'$ is $N(\tilde{\boldsymbol{\mu}}_\rho, \tilde{\boldsymbol{\Sigma}}_\rho)$ and thus we draw $(\rho_1^{(b+1)}, \rho_2^{(b+1)})' \sim N(\tilde{\boldsymbol{\mu}}_\rho, \tilde{\boldsymbol{\Sigma}}_\rho)$.

3. Draw $\gamma^{(b+1)}|\mu^{(b+1)}, \rho_1^{(b+1)}, \rho_2^{(b+1)}, \sigma^{(b)}, \boldsymbol{z}_{1:t}^{(b)}, \boldsymbol{x}_{1:t}^{(b)}, \boldsymbol{y}_{1:t}$. There is no closed form expression for the posterior of γ. We therefore draw it using a random walk Metropolis-Hastings step. Specifically, we draw a proposal $\tilde{\gamma}^{(b+1)} \sim N(\gamma^{(b)}, \sigma^2_{\gamma,prop})$ where $\sigma^2_{\gamma,prop}$ is a proposal variance chosen such that this step has between a 25 and 40% acceptance rate over the burn-in period. We then set $\gamma^{(b+1)} = \tilde{\gamma}^{(b+1)}$ with probability α_{b+1}, where

$$\alpha_{b+1} \equiv \frac{L\left(\boldsymbol{y}_{1:t}|\mu^{(b+1)}, \rho_1^{(b+1)}, \rho_2^{(b+1)}, \tilde{\gamma}^{(b+1)}, \sigma^{(b)}, \boldsymbol{z}_{1:t}^{(b)}, \boldsymbol{x}_{1:t}^{(b)}\right) p_\gamma\left(\tilde{\gamma}^{(b+1)}\right)}{L\left(\boldsymbol{y}_{1:t}|\mu^{(b+1)}, \rho_1^{(b+1)}, \rho_2^{(b+1)}, \tilde{\gamma}^{(b+1)}, \sigma^{(b)}, \boldsymbol{z}_{1:t}^{(b)}, \boldsymbol{x}_{1:t}^{(b)}\right) p_\gamma\left(\gamma^{(b)}\right)}.$$

Otherwise we set $\gamma^{(b+1)} = \gamma^{(b)}$.

4. Draw $\sigma^{(b+1)}|\mu^{(b+1)}, \rho_1^{(b+1)}, \rho_2^{(b+1)}, \gamma^{(b+1)}, \boldsymbol{z}_{1:t}^{(b)}, \boldsymbol{x}_{1:t}^{(b)}, \boldsymbol{y}_{1:t}$. Given the other parameters, beliefs about σ can be updated from the two equations

$$\begin{aligned}\Delta z_t^{(b)} &= \mu^{(b+1)} + \sqrt{\gamma^{(b+1)}}\sigma u_t. \\ x_t^{(b)} &= \rho_1^{(b+1)} x_{t-1}^{(b)} + \rho_2^{(b+1)} x_{t-2}^{(b)} + \sqrt{1-\gamma^{(b+1)}}\sigma v_t.\end{aligned}$$

Since u_t and v_t are independent, these regression equations can be treated as two independent sources of information for σ^2. It is as if beliefs about σ^2 are first updated using information about $\{u_s\}_{s=2}^t$ where $\sigma u_s = \frac{\Delta z_s - \mu}{\sqrt{\gamma}}$ and then updated using information about $\{v_s\}_{s=3}^t$ where $\sigma v_s = \frac{x_s - \rho_1 x_{s-1} - \rho_2 x_{s-2}}{\sqrt{1-\gamma}}$. These are samples of $t-1$ and $t-2$ observations respectively which can be used to learn about σ^2 using standard conjugate prior updating. Define

$$\begin{aligned}\tilde{\alpha}_{\sigma^2} &\equiv \alpha_{\sigma^2} + (2t-3)/2, \\ \tilde{\beta}_{\sigma^2} &\equiv \beta_{\sigma^2} + \frac{\sum_{s=2}^t \left(\Delta z_s^{(b)} - \mu^{(b+1)}\right)^2}{2\gamma^{(b+1)}} + \frac{\sum_{s=3}^t \left(x_s^{(b)} - \rho_1^{(b+1)} x_{s-1}^{(b)} - \rho_2^{(b+1)} x_{s-2}^{(b)}\right)^2}{2\left(1-\gamma^{(b+1)}\right)}.\end{aligned}$$

The posterior of σ^2 is $\mathcal{IG}(\tilde{\alpha}_{\sigma^2}, \tilde{\beta}_{\sigma^2})$ and thus we draw $\left(\sigma^{(b)}\right)^2 \sim \mathcal{IG}(\tilde{\alpha}_{\sigma^2}, \tilde{\beta}_{\sigma^2})$.

5. Draw $\boldsymbol{z}_{1:t}^{(b+1)}, \boldsymbol{x}_{1:t}^{(b+1)}|\mu^{(b+1)}, \rho_1^{(b+1)}, \rho_2^{(b+1)}, \gamma^{(b+1)}, \sigma^{(b+1)}, \boldsymbol{y}_{1:t}$. This can be done using the standard Kalman filter and simulation smoother.

F Bayesian Forecasting of GDP

The algorithm described in Appendix E yields B samples of the posterior of the states and parameters of our UC model for GDP at each point in time t. We index these samples by b as follows $\left\{\rho_1^{(b)}, \rho_2^{(b)}, \gamma^{(b)}, \mu^{(b)}, \sigma^{(b)}, z_{t|t}^{(b)}, x_{t|t}^{(b)}, x_{t-1|t}^{(b)}\right\}_{b=1}^{B}$. We then use the following algorithm to produce a real-time forecast distribution for the GDP at time t:

1. For each $b = 1, \ldots, B$

   (a) Simulate a path of shocks $\left\{u_{t+h}^{(b)}, v_{t+h}^{(b)}\right\}_{h=1}^{H}$ from the standard Normal distribution.

   (b) Starting from $h = 1$, construct a simulated path of the states over H subsequent periods using equations

$$z_{t+h|t}^{(b)} = \mu^{(b)} + \sqrt{\gamma^{(b)}}\sigma^{(b)} u_{t+h}^{(b)},$$
$$x_{t+h|t}^{(b)} = \rho_1^{(b)} x_{t+h-1|t}^{(b)} + \rho_2^{(b)} x_{t+h-2|t}^{(b)} + \sqrt{1-\gamma^{(b)}}\sigma^{(b)} v_{t+h}^{(b)}.$$

   (c) Use the simulated states to construct $\left\{y_{t+h|t}^{(b)}\right\}_{h=1}^{H}$ where

$$y_{t+h|t}^{(b)} = z_{t+h|t}^{(b)} + x_{t+h|t}^{(b)}.$$

2. The forecast of y_{t+h} given time t information is computed as

$$F_t y_{t+h} = \frac{1}{B}\sum_{b=1}^{B} y_{t+h|t}^{(b)}.$$

At the end of the estimation we are left with a sequence of model-implied 1 to H-quarter ahead forecasts $\{F_t y_{t+h}\}_{h=1}^{H}$ for every year t from 1976Q4 to 2019Q4.

We must perform a few additional steps to transform our forecasts to ones that are comparable to those produced by the CBO. The CBO publishes forecasts of growth in the average annual level of real output. We define the average annual level of real output over the year preceding quarter t as

$$\bar{Y}_t \equiv \frac{1}{4}\sum_{s=t-3}^{t} \exp(y_s).$$

As an example, in the CBO's economic outlook published in 1990, its 1-year ahead forecast of GDP

growth is

$$100 \times \left(\frac{\bar{Y}_{1990Q4}}{\bar{Y}_{1989Q4}} - 1 \right).$$

Thus to convert the model's forecasts of quarterly log real GDP to average annual h-year ahead level forecasts, we apply the following transformation to the simulated forecast distribution

$$F_t \bar{Y}_{t+h} \equiv \frac{1}{B} \sum_{b=1}^{B} \left[\frac{1}{4} \sum_{s=t+4h-3}^{t+4h} \exp\left(F_t y_{s|t}^{(b)} \right) \right].$$

The associated forecasts of growth in average annual levels between year $t+h-1$ and $t+h$ for $h = 1, \ldots, H$ are

$$100 \times \left(\frac{F_t \bar{Y}_{t+h}}{F_t \bar{Y}_{t+h-1}} - 1 \right).$$

G Search over Initial Beliefs for GDP Growth

We denote $\boldsymbol{\theta} = (\mu_\rho, \sigma_\rho, \mu_{\rho_2}, \sigma_{\rho_2}, \alpha_\gamma, \beta_\gamma)'$. Let $\boldsymbol{\alpha} = \{\alpha_h\}_{h=1}^H$ and $\boldsymbol{\beta} = \{\beta_h\}_{h=1}^H$ denote vectors of estimated coefficients from the forecasting anomaly regressions for different horizons up through a maximum horizon of H using the CBO data. Let $\hat{\boldsymbol{\alpha}} = \{\hat{\alpha}_h\}_{h=1}^H$ and $\hat{\boldsymbol{\beta}} = \left\{\hat{\beta}_h\right\}_{h=1}^H$ denote those same quantities estimated on the model implied forecasts and yields. Additionally, denote the $t-statistics$ associated with these coefficients as $\{\boldsymbol{t}_\alpha, \boldsymbol{t}_\beta\} = \{t_{\alpha,h}, t_{\beta,h}\}_{h=1}^H$ for the data and $\left\{\boldsymbol{t}_{\hat{\alpha}}, \boldsymbol{t}_{\hat{\beta}}\right\} = \left\{t_{\hat{\alpha},h}, t_{\hat{\beta},h}\right\}_{h=1}^H$ for the model. Define the moment function as

$$\hat{m}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\alpha}_{bias} - \hat{\boldsymbol{\alpha}}_{bias} \\ \boldsymbol{t}_{\alpha,bias} - \boldsymbol{t}_{\hat{\alpha},bias} \\ \boldsymbol{\beta}_{ar} - \hat{\boldsymbol{\beta}}_{ar} \\ \boldsymbol{t}_{\beta,ar} - \boldsymbol{t}_{\hat{\beta},ar} \\ \boldsymbol{\beta}_{mz} - \hat{\boldsymbol{\beta}}_{mz} \\ \boldsymbol{t}_{\beta,mz} - \boldsymbol{t}_{\hat{\beta},mz} \\ \boldsymbol{\beta}_{cg} - \hat{\boldsymbol{\beta}}_{cg} \\ \boldsymbol{t}_{\beta,cg} - \boldsymbol{t}_{\hat{\beta},cg} \end{bmatrix} \tag{23}$$

The parameters are then estimated via SMM with the following objective function

$$\hat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}} \hat{m}(\boldsymbol{\theta})' \boldsymbol{W} \hat{m}(\boldsymbol{\theta})$$

where the elements of the objective function associated with the Mincer-Zarnowitz and Coibion-Gorodnichenko coefficients are given 3 times the weight of all other elements in $\boldsymbol{W}$. We also place bounds on the estimated parameters as described in footnote 12 in the main text. The estimated initial belief distributions are plotted in Figure 9.

Every evaluation of the moment function $\hat{m}(\boldsymbol{\theta})$ requires us to sample from the posterior of the UC model sequentially. Since this step is very computationally costly, we only re-estimate the model every 4 quarters rather than every quarter, and use a burn-in sample of 15,000 draws and keep the subsequent 15,000 draws rather than 50,000 for each of those quantities in our empirical specification. The global minimum is found using MATLAB's "particleswarm" optimization routine.

References

ANDRADE, P., R. K. CRUMP, S. EUSEPI, AND E. MOENCH (2016): "Fundamental Disagreement," *Journal of Monetary Economics*, 83, 106–128.

ANGELETOS, G.-M., Z. HUO, AND K. A. SASTRY (2020): "Imperfect Macroeconomic Expectations: Evidence and Theory," *NBER Macroeconomics Annual*, 35.

BANSAL, R. AND I. SHALIASTOVICH (2013): "A Long-run Risks Explanation of Predictability Puzzles in Bond and Currency Markets," *The Review of Financial Studies*, 26, 1–33.

BARSKY, R. B. AND J. B. DE LONG (1993): "Why Does the Stock Market Fluctuate?" *Quarterly Journal of Economics*, 108, 291–311.

BAUER, M. D. AND G. D. RUDEBUSCH (2020): "Interest Rates under Falling Stars," *American Economic Review*, 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," *Journal of Financial Economics*, 42, 309–348.

——— (2001): "Peso Problem Explanations for Term Structure Anomalies," *Journal of Monetary Economics*, 48, 241–270.

BEN-DAVID, I., J. R. GRAHAM, AND C. R. HARVEY (2013): "Managerial Miscalibration," *Quarterly Journal of Economics*, 128, 1547–1584.

BIANCHI, F., M. LETTAU, AND S. C. LUDVIGSON (2020): "Monetary policy and Asset Valuation," *Journal of Finance*.

BIDDER, R. AND I. DEW-BECKER (2016): "Long-Run Risk Is the Worst-Case Scenario," *American Economic Review*, 106, 2494–2527.

BORDALO, P., N. GENNAIOLI, Y. MA, AND A. SHLEIFER (2020): "Overreaction in Macroeconomic Expectations," *American Economic Review*, 110, 2748–2782.

BRAV, A. AND J. B. HEATON (2002): "Competing Theories of Financial Anomalies," *Review of Financial Studies*, 15, 575–606.

BROER, T. AND A. N. KOHLHAS (2022): "Forecaster (Mis-)Behavior," *Review of Economics and Statistics*, forthcoming.

CAMPBELL, J. Y. AND R. SHILLER (1991): "Yield Spreads and Interest Rate Movements: A Bird's Eye View," *Review of Economic 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, University of Texas, Austin.

CASKEY, J. (1985): "Modelling the Formation of Price Expectations: A Bayesian Approach," *American Economic Review*, 75, 768–776.

CIESLAK, A. (2018): "Short-rate expectations and unexpected returns in treasury bonds," *The Review of Financial Studies*, 31, 3265–3306.

CIESLAK, A. AND P. POVALA (2015): "Expected Returns in Treasury Bonds," *Review of Financial Studies*, 28, 2859–2901.

COCHRANE, J. (1988): "How Big Is the Random Walk in GNP?" *Journal of Political Economics*, 96, 893–920.

COCHRANE, J. H. AND M. PIAZZESI (2005): "Bond Risk Premia," *American Economic Review*, 95, 138–160.

COGLEY, T. AND T. J. SARGENT (2005): "The Conquest of US Inflation: Learning and Robustness

to Model Uncertainty," *Review of Economic Dynamics*, 8, 528–563.

——— (2008): "The Market Price of Risk and the Equity Premium: A Legacy of the Great Depression?" *Journal of Monetary Economics*, 55, 454–476.

COIBION, O. AND Y. GORODNICHENKO (2012): "What Can Survey Forecasts Tell Us about Information Rigidities?" *Journal of Political Economy*, 120, 116–159.

——— (2015): "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts," *American Economic Review*, 105, 2644–2678.

COLLIN-DUFRESNE, P., M. JOHANNES, AND L. A. LOCHSTOER (2016): "Parameter Learning in General Equilibrium: The Asset Pricing Implications," *American Economic Review*, 106, 664–98.

CONGRESSIONAL BUDGET OFFICE (2021): "CBO's Economic Forecasting Record: 2021 Update," Https://www.cbo.gov/publication/57579.

CROCE, M. M., M. LETTAU, AND S. C. LUDVIGSON (2015): "Investor Information, Long-run Risk, and the Term Structure of Equity," *Review of Financial Studies*, 28, 706–742.

CROUSHORE, D. (1998): "Evaluating Inflation Forecasts," Working Paper, Federal Reserve Bank of Philadelphia.

CRUMP, R. K., S. EUSEPI, E. MOENCH, AND B. PRESTON (2021): "The Term Structure of Expectations," in *Handbook of Economic Expectations*, ed. by R. Bachmann, G. Topa, and W. van der Klaauw, Amsterdam, Holland: Elsevier.

EUSEPI, S. AND B. PRESTON (2011): "Expectation, Learning, and Business Cycle Fluctuations," *American Economic Review*, 101, 2844–2872.

——— (2018): "The Science of Monetary Policy: An Imperfect Knowledge Perspective," *Journal of Economic Literature*, 56, 3–59.

EVANS, G. W. AND S. HONKAPOHJA (2001): *Learning and Expectations in Macroeconomics*, Princeton, NJ: Princeton University Press.

FAMA, E. (1984): "The Information in the Term Structure," *Journal of Financial Economics*, 13, 509–528.

FAMA, E. F. (2006): "The Behavior of Interest Rates," *Review of Finacial Studies*, 19, 359–379.

FAMA, E. F. AND R. R. BLISS (1987): "The Information in Long-Maturity Forward Rates," *American Economic Review*, 77, 680–692.

FRIEDMAN, B. M. (1979): "Optimal Expectations and the Extreme Information Assumptions of 'Rational Expectations' Macromodels," *Journal of Monetary Economics*, 5, 23–41.

——— (1980): "Survey Evidence on the 'Rationality' of Interest Rate Expectations," *Journal of Monetary Economics*, 6, 453–465.

FROOT, K. A. (1989): "New Hope for the Expectations Hypothesis of the Term Structure of Interest Rates," *Journal of 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 University.

GIGLIO, S. AND B. KELLY (2018): "Excess Volatility: Beyond Discount Rates," *Quarterly Journal of Economics*, 133, 71–127.

GOURINCHAS, P.-O. AND A. TORNELL (2004): "Exchange Rate Puzzles and Distorted Beliefs," *Journal of International Economics*, 64, 303–333.

GUO, H. AND J. WACHTER (2019): ""Superstitious" Investors," NBER Working Paper No. 25603.

HAMILTON, J. D. AND C. WU (2012): "Identification and Estimation of Gaussian Affine Term

Structure Models," *Journal of Econometrics*, 168, 315–331.

JOHANNES, M., L. A. LOCHSTOER, AND Y. MOU (2016): "Learning about Consumption Dynamics," *Journal of Finance*, 71, 551–600.

KOZICKI, S. AND P. TINSLEY (2001): "Shifting endpoints in the term structure of interest rates," *Journal of Monetary Economics*, 47, 613–652.

KOZLOWSKI, J., L. VELDKAMP, AND V. VENKATESWARAN (2020): "The Tail that Wags the Economy: Beliefs and Persistent Stagnation," *Journal of Political Economy*, 128, 2839–2879.

KREPS, D. M. (1998): "Anticipated Utility and Dynamic Choice," in *Frontiers of Research in Economic Theory: The Nancy L. Schwartz Memorial Lectures, 1983–1997*, ed. by D. P. Jacobs, E. Kalai, and M. I. Kamien, Cambridge, UK: Cambridge University Press, 242–274.

LAZARUS, E., D. J. LEWIS, J. H. STOCK, AND M. W. WATSON (2018): "HAR Inference: Recommendations for Practice," *Journal of Business & Economic Statistics*, 36, 541–559.

LEWELLEN, J. AND J. SHANKEN (2002): "Learning, Asset-pricing Tests, and Market Efficiency," *Journal of Finance*, 57, 1113–1145.

LEWIS, K. K. (1989a): "Can Learning Affect Exchange-Rate Behavior?: The Case of the Dollar in the Early 1980's," *Journal of Monetary Economics*, 23, 79–100.

——— (1989b): "Changing Beliefs and Systematic Rational Forecast Errors with Evidence from Foreign Exchange," *American Economic Review*, 621–636.

LIU, Y. AND J. C. WU (2020): "Reconstructing the Yield Curve," *Journal of Financial Economics*, forthcoming.

MADDALA, G. S. (1991): "Survey data on Expectations: What Have We Learnt?" in *Issues in Contemporary Economics*, ed. by M. Nerlove, London, UK: Palgrave McMillan, 319–344.

MANKIW, N. G. AND R. REIS (2002): "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve," *Quarterly Journal of Economics*, 117, 1295–1328.

MANKIW, N. G., R. REIS, AND J. WOLFERS (2003): "Disagreement about Inflation Expectations," *NBER Macroeconomics Annual*, 18, 209–248.

MINCER, J. A. AND V. ZARNOWITZ (1969): "The Evaluation of Economic Forecasts," in *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, ed. by J. A. Mincer, New York, NY: NBER, 3–46.

MOLAVI, P., A. TAHBAZ-SALEHI, AND A. VEDOLIN (2021): "Model Complexity, Expectations, and Asset Pricing," NBER Working Paper No. 28408.

NAGEL, S. AND Z. XU (2021): "Dynamics of Subjective Risk Premia," Working Paper, University of Chicago.

NORDHAUS, W. (1987): "Forecasting Efficiency: Concepts and Applications," *Review of Economics and Statistics*, 69, 667–674.

PATTON, A. J. AND A. TIMMERMANN (2010): "Why do forecasters disagree? Lessons from the term structure of cross-sectional dispersion," *Journal of Monetary Economics*, 57, 803–820.

——— (2011): "Predictability of output growth and inflation: A multi-horizon survey approach," *Journal of Business & Economic Statistics*, 29, 397–410.

PIAZZESI, M., J. SALOMAO, AND M. SCHNEIDER (2015): "Trend and Cycle in Bond Premia," Working Paper, Stanford University.

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 University Press.

SCHUH, S. (2001): "An Evaluation of Recent Macroeconomic Forecast Errors," *New England Economic Review*, 2001, 35–36.

SHILLER, R. J. (1979): "The Volatility of Long-Term Interest Rates and Expectations Theories of the Term Structure," *Journal of Political Economics*, 87, 1190–1219.

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 on Economic Activity*, 1983, 173–223.

SIMS, C. A. (2003): "Implications of Rational Inattention," *Journal of Monetary Economics*, 50, 665–690.

SINGLETON, K. (2021): "How Much "Rationality" Is There in Bond-Market Risk Premiums?" *Journal of Finance*, 76, 1611–1654.

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

VAN DIJK, D., S. J. KOOPMAN, M. VAN DER WEL, AND J. H. WRIGHT (2014): "Forecasting Interest Rates with Shifting Endpoints," *Journal of Applied 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," *Journal of Financial Economics*, 79, 365–399.

WOODFORD, M. (2003): "Imperfect Common Knowledge and the Effects of Monetary Policy," *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, 25.

XU, Z. (2019): "Expectation Formation in the Treasury Bond Market," Working Paper, City University of Hong Kong.