WORKING paper

BANQUE DE FRANCE
EUROSYSTÈME

Information in (and not in) Interest Rate Surveys

Laura Coroneo[1] and Adam Goliński[2]

November 2026, WP #1070

ABSTRACT

Central banks and investors routinely decompose long-term interest rates into expected policy rates and term premia, using either bond prices or surveys to pin down expectations. We show that the two approaches measure different objects. We develop a joint shadow rate term structure model for yields and surveys in which survey participants may misperceive both how the policy rate responds to yield-curve factors and how those factors evolve. Using U.S. data from 1987 to 2024, we formally test for these two biases and find evidence of both: misperceptions of the policy rate's response dominate at short maturities, and persistent forecast errors at long maturities. As a result, survey-implied 10-year term premia are on average 1.2 percentage points lower than those inferred from bond prices, with important implications for interpreting movements in long-term interest rates. Surveys are therefore essential to identify the risk compensation investors require, but imposing survey expectations on historical dynamics worsens interest rate forecasts.

Keywords: Term Premium, Surveys, Probability Measures, Term Structure Model

JEL classification: E43, E52, G12, C58, E44

[1] University of York, laura.coroneo@york.ac.uk.

[2] Banque de France and University of York, adam.golinski@banque-france.fr.

We thank Michael Bauer, Don Kim, Sarah Mouabbi, Jean-Paul Renne and Stefania D'Amico for discussing the paper, and Jens Christensen, John Cochrane, Wolfgang Lemke, Athanasios Orphanides, Raman Uppal, and seminar participants at the ECB Monetary Policy Conference, the Federal Reserve Board, the Banque de France, the University of Milan, the University of Liverpool, the University of East Anglia, the Workshop "Non-Linearities in Macro" (Queen Mary University of London), the 10th Asset Pricing Workshop (University of York), the 2023 Qatar Centre for Global Banking & Finance Annual Conference (King's Business School), the IFABS 2023 (Saïd Business School), the EEA 2023 (Barcelona School of Economics), the MMF 2023 (University of Portsmouth), the University of Lancaster, the 12th Term Structure Workshop (Bundesbank), the University of Glasgow, the York-Tohoku Symposium and the 2026 RCEA International Conference in Madrid for useful comments.

Working Papers reflect the opinions of the authors and do not necessarily express the views of the Banque de France. This document is available on publications.banque-france.fr/en

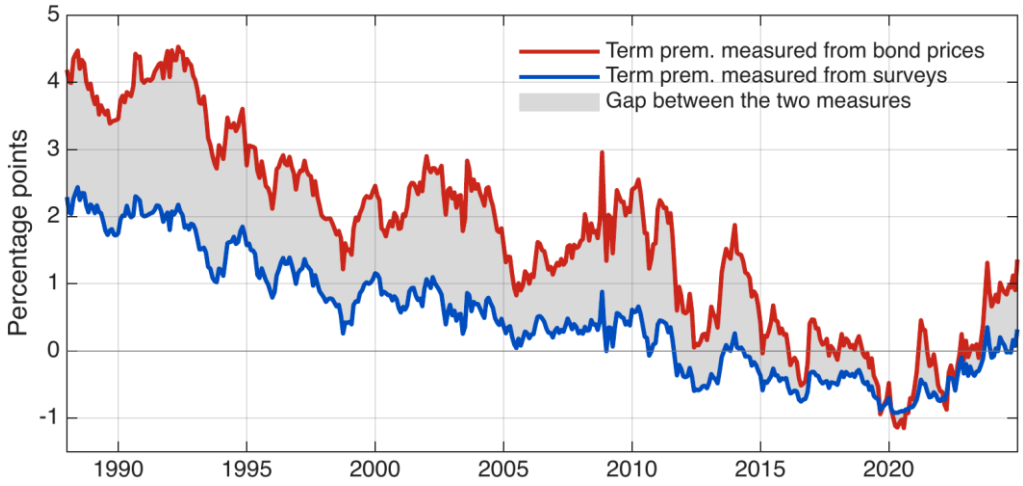
NON-TECHNICAL SUMMARY

Long-term interest rates reflect two things: what investors expect short-term policy rates to be in the future, and the compensation for the risk of holding long-term bonds – the term premium. Separating the two is important for central banks. A fall in long-term rates driven by lower expected policy rates says something different about market views than one driven by a lower term premium. Expectations are not observed directly, however. They are usually inferred either from the past behaviour of bond prices or from surveys of professional forecasters, and the two sources are often treated as interchangeable.

We show that they are not. We build a model of the US yield curve that uses both bond yields and the Blue Chip Financial Forecasts survey of interest rates from 1987 to 2024, including the period in which policy rates were at their lower bound. The model allows forecasters to make two kinds of systematic error: misjudging how the policy rate responds to the underlying drivers of interest rates, and misjudging how these drivers evolve over time.

We find that both errors are present. Misperceptions of how the policy rate responds matter mostly for forecasts of short-term interest rates, while persistent forecast errors matter most for long-term rates. As a result, the two approaches, based on bond prices and surveys, give different measures of term premia. The survey-implied 10-year term premium is on average 1.2 pp lower than the one inferred from bond prices, and the gap has at times exceeded 2 percentage points (see Figure 1). The two approaches also tell different stories about the decline in long-term rates since the late 1980s: the survey-based measure attributes about half of it to lower expected policy rates, against less than a third for the measure based on bond prices.

Figure 1. 10-year term premium measured from bond prices and from surveys



Note: Monthly data, December 1987 to December 2024. The shaded area is the gap between the two measures.
Source: authors' calculations.

Our findings highlight the distinction between measuring perceived risk compensation and forecasting interest rates. Surveys are essential to gauge the risk compensation investors subjectively perceive, which cannot be recovered from bond prices alone. However, imposing survey expectations on the statistical behaviour of interest rates worsens forecasts. Thus, ultimately, the relevant measure of the term premium depends on the question being asked: assessing, for instance, how much of the response of long-term rates to an asset purchase announcement reflects a change in perceived risk compensation calls for survey-based measures, while forecasting the path of interest rates should instead draw on their historical behaviour.

Ce que révèlent (et ne révèlent pas) les enquêtes sur les taux d'intérêt

RÉSUMÉ

Les banques centrales et les investisseurs décomposent couramment les taux d'intérêt de long terme en taux directeurs anticipés et en primes de terme, en s'appuyant soit sur les prix des obligations, soit sur des enquêtes pour mesurer les anticipations. Nous montrons que ces deux approches ne mesurent pas la même chose. Nous développons un modèle de structure par terme des taux d'intérêt de type « shadow rate », estimé conjointement sur les rendements obligataires et les enquêtes, dans lequel les participants aux enquêtes peuvent mal percevoir à la fois la réaction du taux directeur aux facteurs de la courbe des taux et la dynamique de ces facteurs. À partir de données américaines de 1987 à 2024, nous testons formellement ces deux biais et trouvons des éléments en faveur de chacun d'eux : les erreurs de perception de la réaction du taux directeur dominent aux échéances courtes, et les erreurs de prévision persistantes aux échéances longues. En conséquence, les primes de terme à 10 ans implicites dans les enquêtes sont en moyenne inférieures de 1,2 point de pourcentage à celles déduites des prix des obligations, ce qui a des implications importantes pour l'interprétation des mouvements des taux d'intérêt de long terme. Les enquêtes sont donc indispensables pour identifier la compensation du risque exigée par les investisseurs, mais imposer les anticipations des enquêtes à la dynamique historique des taux détériore les prévisions de taux d'intérêt.

Mots-clés : prime de terme, enquêtes de prévision, mesures de probabilité, modèle de structure par terme des taux

Les Documents de travail reflètent les idées personnelles de leurs auteurs et n'expriment pas nécessairement la position de la Banque de France. Ils sont disponibles sur publications.banque-france.fr

1 Introduction

Central banks, governments and market participants closely monitor long-term interest rates and routinely use term structure models to separate the information about investors' expectations of the economy from the risks they perceive. However, term structure models can suffer from identification issues (see Ang and Piazzesi 2003, Hamilton and Wu 2012) that can lead to a misestimation of the risk premium and a misperception of the market's outlook.

Survey forecasts have been used to address these identification problems (see Kim and Wright 2005, Kim and Orphanides 2012, Chun 2011, Chernov and Mueller 2012, d'Amico, Kim and Wei 2018, Crump, Eusepi and Moench 2018) relying on two implicit assumptions. The first is that the perceived policy reaction function in surveys coincides with the true reaction function (equal reaction assumption), and the second is that the probability measure underlying survey expectations is the same as the statistical probability measure that governs observed bond prices (equal dynamics assumption). Yet, evidence on short-term interest rates points to systematic violations of both: biases in perceived policy rules (see Cieslak 2018, Schmeling, Schrimpf and Steffensen 2022, Bauer, Pflueger and Sunderam 2024, Sastry 2026) and deviations from rational expectations (see Piazzesi, Salomao and Schneider 2015, Farmer, Nakamura and Steinsson 2024). This raises the question of how such biases affect the decomposition of long-term interest rates into expectation and term premium components. In particular, if survey expectations are biased, then survey-augmented term structure models may identify neither the historical dynamics of interest rates nor the term premium required by market participants, unless those biases are modeled.

In this paper, we address this question by developing a joint shadow rate term structure model for zero-coupon yields and survey expectations. Our framework allows for separate

objective and subjective policy reaction functions and for three distinct probability measures (risk-neutral, objective, and subjective), while explicitly accounting for the zero lower bound on both yields and survey expectations. We estimate the resulting joint shadow-rate model by maximum likelihood on monthly U.S. interest rate data from December 1987 to December 2024 and Blue Chip Financial Forecasts. We then use the flexibility of this framework to test the equal reaction function and equal dynamics assumptions. Existing shadow rate models estimated with survey data use forecasts of the short rate only (e.g. Priebsch 2023, Kim and Priebsch 2020, Connolly and Struby 2026), whose model counterpart is the mean of a censored normal variable. To incorporate survey expectations of yields of any maturity, we derive a closed-form approximation in the spirit of Wu and Xia (2016), which accounts jointly for the nonlinearity of the pricing function and for the uncertainty about the future state variables. To the best of our knowledge, this extension is novel.

Our results show that although surveys identify subjective beliefs and subjective risk compensation, they do not, in general, identify the physical law of motion of interest rates. In particular, we find that the equal reaction function and equal dynamics assumptions are both rejected in the data. Survey participants make systematic reaction function errors at short maturities and probability forecast errors at long maturities. These findings have three important implications. First, ignoring survey biases leads to a distorted characterization of the true dynamics of interest rates, with important consequences for inference and forecasting. Second, incorporating surveys into a term structure model is essential to recover the subjective term premium, which captures investors' perceived risks even when their expectations are biased. Third, survey-implied term premia are systematically lower than those inferred from bond prices, indicating that term premia extracted solely from yields tend to overstate the compensation investors require given their subjective expectations.[1] For the

[1]We call the "subjective" (or survey) term premium the compensation investors perceive given their be-

10-year yield, the wedge is on average 121 bp (and exceeds 200 bp at times), thus potentially materially changing the interpretation of monetary policy episodes. For example, in November 2008, the term premium accounts for 74% of the decline in the 10-year yield under the physical decomposition, but only 42% under the subjective decomposition.

Contrary to earlier approaches in the term structure literature, we find that surveys do not aid in identifying the physical dynamics of interest rates. Under the full-information rational expectations hypothesis, agents' subjective probability measure coincides with the real-world measure and their perceived policy reaction function matches the true one, so forecast errors should not be biased or persistent.[2] In contrast, our formal tests reject both the equal reaction and the equal dynamics assumptions. Specifically, imposing equal reaction functions and equal dynamics forces the estimated coefficients of the model to collapse towards survey-based expectations, but these fail to forecast observed interest rates. Only by allowing for reaction function errors and probability forecast errors do we obtain the dynamics and reaction function that are consistent with those from observed interest rates. A key implication is that the term premium perceived by survey participants is not determined solely by the real-world dynamics of observed interest rates and, therefore, can only be identified using survey expectations. This evidence against rational survey expectations aligns with earlier findings in macroeconomics (see, among others Mankiw, Reis and Wolfers 2003, Coibion and Gorodnichenko 2015, Bianchi, Ludvigson and Ma 2022) and finance (see Greenwood and Shleifer 2014, Adam, Marcet and Beutel 2017, De la O and Myers 2021).

Thus, our results prompt a fundamental question: When, exactly, are surveys necessary? The answer depends on what we intend to measure. If the objective is to forecast interest

liefs, while the "physical" term premium is the econometrician's decomposition under the historical measure.

[2]This definition of rationality is different from the one used by Adam and Marcet (2011) and Singleton (2021), in which agents do not know the data-generating process but form rational beliefs.

rates or expected returns, a yields-only model may suffice. In contrast, if the goal is to quantify the premium that investors (or survey respondents) demand for holding long-maturity bonds, then survey data become essential, as such premiums cannot be reliably identified from yields alone. These two purposes—forecasting and measuring risk compensation—are often treated as though they require the same informational inputs, but they do not. Our paper makes this distinction explicit and shows why it matters.

A common motivation for incorporating surveys into dynamic term structure model estimation is that, given the high persistence of interest rates, survey expectations can mitigate the short-sample bias that otherwise leads to underestimated yield persistence and distorted inferences about interest rate dynamics (see, e.g., Bauer, Rudebusch and Wu 2012, Duffee and Stanton 2012, Kim and Orphanides 2012). For example, Kim and Orphanides (2012) augment observed yields with survey forecasts and argue that this approach helps alleviate short-sample problems and improves identification of interest rate dynamics. In our framework, however, this identification benefit arises only under restrictions that tightly link subjective survey expectations to the historical law of motion of observed interest rates. Because our empirical analysis uses a long U.S. sample starting in 1987, it is not primarily designed to use surveys as a substitute for sample length. Nevertheless, to quantify the extent to which finite-sample estimation bias affects our estimated forecast errors, we implement a bootstrap procedure to assess the impact of sample length on the small-sample estimation bias. Our results show that a long sample is important not only for sharpening inference on interest rate dynamics, but also for improving the identification and quantification of biases embedded in survey expectations.

In closely related work, Piazzesi et al. (2015) also construct subjective bond risk premia from survey forecasts of short-term interest rates, while Chernov and Mueller (2012) apply a similar approach to survey expectations of inflation. In these frameworks, the subjective

measure can differ from the objective one only through the dynamics of the state variables. We develop a more general model that also accommodates a subjective policy reaction function, and we formally test the assumptions of an equal reaction function, equal dynamics and an equal long-run mean of the short rate under the two measures. We reject the hypothesis that the subjective and objective measures coincide for interest rate expectations, in contrast with the findings of Chernov and Mueller (2012) for inflation expectations.[3] Piazzesi et al. (2015) impose that subjective expectations are on average correct, which in their framework also implies that the long-run mean of the short rate is the same under physical and subjective measures. Our results suggest that the assumption of equal long-run mean of the short rate is not supported by the data either. Although this restriction is not formally rejected by the likelihood ratio test, the evidence against it accumulates across other model validation checks: imposing it increases the Akaike information criterion relative to the unrestricted model, worsens both in-sample and out-of-sample forecasts, and implies counterfactually high long-run means of yields, particularly at the short end of the yield curve.

More recently, Zeng and Zhao (2026) attribute movements in bond and equity yields to subjective expectations using an equilibrium model whose learning gains are estimated from survey forecasts of macroeconomic variables. However, their framework implies a constant subjective risk premium, so that all variation in yields is expectational, whereas our no-arbitrage approach leaves the subjective term premium unrestricted and estimates it jointly with subjective expectations.

The paper is organized as follows. In Section 2 we describe the data and some preliminary evidence. In Section 3 we present the general modelling framework and the shadow rate term structure model. In Section 4 we describe the estimation approach. In Section 5 we report

[3]This difference may reflect the different variables being forecast, or the different sample periods: the sample of Chernov and Mueller (2012) ends in 2008, whereas ours ends in 2024 and thus spans the post-2008 zero-lower-bound period and the associated shifts in the monetary policy regime.

the empirical results, and, finally, in Section 6 we conclude.

2 Data and preliminary evidence

We use monthly observations for zero-coupon yields and surveys from December 1987 to December 2024. For interest rates, we use end-of-month rates for maturities 3 and 6 months, 1, 2, 5, 7 and 10 years from Liu and Wu (2021).[4]

For surveys, we use Blue Chip Financial Forecasts (BCFF) consensus (mean) forecasts of 3-month, 6-month, and 1, 2, 5 and 10-year yields at 3, 6, 9 and 12 months ahead, obtained by linearly interpolating adjacent end of quarter horizons. This gives a total of 24 interest rate survey forecasts.[5] More details about the BCFF can be found in Appendix C.

As explained in the introduction, the standard approach in term structure modelling assumes that the survey reaction function is the same as the historical reaction function and that the probability measure used by economic agents is the same as the statistical probability measure.[6] It follows that the model-implied term premium is interpreted as the risk premium that investors require for holding long-maturity bonds. In this conceptual framework, investors' expectations are filtered out from the physical dynamics of interest rates using the physical reaction function. Then, a valid question is how the expectations implied by the observed interest rates are related to market expectations.

[4]The data is available from: https://sites.google.com/view/jingcynthiawu/yield-data.

[5]The consensus is the average of the point forecasts of individual respondents, which we match to the model-implied subjective mean of future yields, as is standard in the literature (see, e.g., Kim and Orphanides 2012). Near the lower bound, the distribution of future interest rates is skewed, so its mean, median and mode need not coincide, and individual point forecasts may refer to any of them; Kim and Priebsch (2020), for instance, treat the consensus forecast of the short rate as a noisy observation of its modal expectation during the lower-bound period. In our framework, such discrepancies are absorbed by the survey measurement errors, and they are likely to be small except for the shortest maturities during the lower-bound period.

[6]In the literature, the term "statistical measure" is also called "real-world measure", "physical measure" or "objective measure". In this paper, we use these terms interchangeably. Similarly, we use interchangeably the terms "subjective measure" and "survey measure".

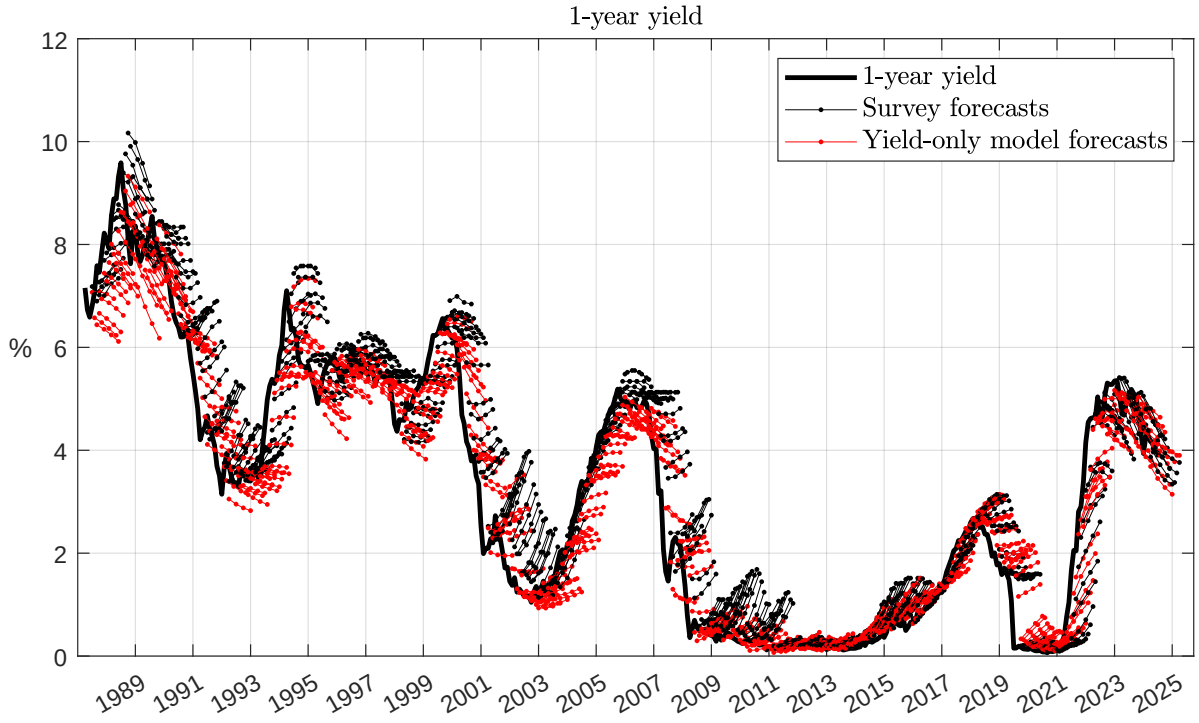


Figure 1. Market expectations measured by surveys (black-dotted lines) and expectations extracted from a yield-only term structure model of interest rates (red-dotted lines), along with the observed 1-year yield (black bold line).

To address this question, in Figure 1 we plot the consensus forecasts of the Blue Chip Financial Forecasts for the 1-year yield over the period December 1987 to December 2024 for horizons 1 through 4 quarters ahead.[7] In the same figure, we plot the forecasts implied by a term structure model estimated using only yields (in the rest of the paper we refer to this specification as Case 0). Since a large part of the sample period covers the zero lower bound period, we use a shadow rate model that prohibits negative interest rates (for more details please see Section 3.2).[8] Visual inspection reveals that forecasts from the yield-

[7]Figures for other maturities look fundamentally the same and they are available from the authors upon request.

[8]The figure reports in-sample expectations extracted from a yield-only term structure model of interest rates. The real-time expectations are in the top plot of Figure G-1 in Appendix G.

only model seem to have little relation to market-based expectations. Even more striking is the pattern that indicates a different direction of predicted interest rates. This preliminary evidence indicates that survey expectations of interest rates are not driven only by historical interest rates, and thus in order to correctly estimate the term premium and interest rate expectations, any term structure model that incorporates survey expectations must allow for survey biases.

3 Framework

In this section, we present our term structure modeling framework which involves both the standard specification of the risk-neutral and the physical dynamics, and also the subjective dynamics. In addition, we allow for distinct objective and subjective policy reaction functions and we embed the underlying Gaussian factor dynamics within a shadow rate model that enforces a lower bound on interest rates.

3.1 Bond prices

We assume that bond prices are driven by a K-dimensional state vector $\mathbf{x}_t$, and that there exists a risk-neutral probability measure $\mathbb{Q}$ that prices all financial assets, under which the state variable follows a first-order Gaussian vector autoregression ($VAR(1)$):

$$\mathbf{x}_t = \boldsymbol{\mu}^{\mathbb{Q}} + \boldsymbol{\Phi}^{\mathbb{Q}}\mathbf{x}_{t-1} + \mathbf{u}_t^{\mathbb{Q}}, \tag{1}$$

where $\mathbf{u}_t^{\mathbb{Q}} \overset{\mathbb{Q}}{\sim} i.i.d.\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$.

We denote the logarithm of the price of a zero-coupon bond at time t with remaining maturity n by $p_{n,t}$ and the corresponding yield by $y_{n,t} = -p_{n,t}/n$. In the absence of arbitrage

opportunities, the price of a zero-coupon bond can be expressed as

$$p_{n,t} = \log E_t^{\mathbb{Q}} \left[\exp \left(- \sum_{j=0}^{n-1} r_{t+j} \right) \right], \tag{2}$$

where r_t is the short rate driven solely by the state vector $\mathbf{x}_t$, that is $r_t = r(\mathbf{x}_t)$.[9] It follows that bond yields are a function of the state vector, which we denote as

$$y_{n,t} \equiv y(\mathbf{x}_t, n; \boldsymbol{\Psi}), \tag{3}$$

where $\boldsymbol{\Psi}$ is a vector of relevant risk-neutral parameters and $y(.)$ is the (yield) pricing function.

3.2 The shadow rate term structure model

To enforce a lower bound on interest rates, we assume that there exists a shadow short rate that is linear in the Gaussian state variables

$$ssr_t = \delta_0 + \boldsymbol{\delta}_1' \mathbf{x}_t \tag{4}$$

and that the observed short-term interest rate r_t is equal to the shadow short rate only when the latter is above the lower bound, and equal to the lower bound otherwise, i.e.,

$$r_t = \max\{ssr_t, \underline{r}\}. \tag{5}$$

The advantage of the shadow rate model is that it allows us to operate within the Gaussian framework described in Section 3.1 but, at the same time, it imposes the lower bound on interest rates, which is essential for the consistency of the model estimates if the sample

[9] If the pricing function includes lags of $\mathbf{x}_t$, we can always redefine the state vector as $\tilde{\mathbf{x}}_t = [\mathbf{x}_t', \mathbf{x}_{t-1}', \ldots]'$.

period includes such episodes. The idea of a shadow rate was first introduced by Black (1995) and is based on the observation that negative interest rates are hard to enforce if investors have the option to convert to currency. As a result of this option, all term rates and forward rates are bounded and do not have a reflecting boundary condition, as opposed to affine models with factors that follow square-root processes or quadratic Gaussian models. Another feature of the shadow rate term structure model is that, when short-term interest rates are far from the lower bound, interest rates behave approximately as in a Gaussian affine term structure model.

The pricing function for the shadow rate model is not available in closed form when there are more than two factors. Wu and Xia (2016), however, provide a convenient approximation in discrete time. In particular, denote by $f_{n,t}$ the time t one-period forward rate for a loan starting at $t+n$. Wu and Xia (2016) show that, under Eq.(1) and Eq.(5), the forward rate $f_{n,t}$ is approximately equal to

$$f_{n,t} \approx \underline{r} + \sigma_n g\left(\frac{a_n + \mathbf{b}_n' \mathbf{x}_t - \underline{r}}{\sigma_n}\right), \tag{6}$$

where the function $g(z) = z\mathtt{N}(z) + \mathtt{n}(z)$ with $\mathtt{N}(z)$ and $\mathtt{n}(z)$ the standard normal CDF and the PDF of z, respectively, and

$$\begin{aligned} a_n &= a(\delta_0, \boldsymbol{\delta}_1, \boldsymbol{\mu}^{\mathbb{Q}}, \boldsymbol{\Phi}^{\mathbb{Q}}, \boldsymbol{\Sigma}, n), \\ \mathbf{b_n} &= b(\boldsymbol{\delta}_1, \boldsymbol{\Phi}^{\mathbb{Q}}, n), \\ \sigma_n &= \sigma(\boldsymbol{\delta}_1, \boldsymbol{\Phi}^{\mathbb{Q}}, \boldsymbol{\Sigma}, n). \end{aligned}$$

For more details and definitions of the functions a, b and σ, please refer to Wu and Xia (2016), Appendix A.

3.3 Surveys

Assume now that, in addition to the zero-coupon bond prices, we also observe market agents' expectations about future zero-coupon yields. Denote the h-period ahead expectation at time t of a yield on a zero-coupon bond with an n-period tenor by $y^s_{n,t,h}$, such that $y^s_{n,t,h} \equiv E^{\mathbb{S}}_t[y^{\mathbb{S}}(\mathbf{x}_{t+h}, n; \mathbf{\Psi})]$, where $E^{\mathbb{S}}$ is the *subjective* expectation operator and $y^{\mathbb{S}}(.)$ is the *perceived* yield pricing function.[10]

To account for *subjective* expectations, we allow survey expectations of the state variables $E^{\mathbb{S}}_t[\mathbf{x}_{t+h}]$ to derive from the subjective state dynamics

$$\mathbf{x}_t = \boldsymbol{\mu}^{\mathbb{S}} + \mathbf{\Phi}^{\mathbb{S}}\mathbf{x}_{t-1} + \mathbf{u}^{\mathbb{S}}_t, \quad \mathbf{u}^{\mathbb{S}}_t \overset{\mathbb{S}}{\sim} i.i.d.\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}), \tag{7}$$

which do not necessarily coincide with the physical dynamics observed in the data

$$\mathbf{x}_t = \boldsymbol{\mu}^{\mathbb{P}} + \mathbf{\Phi}^{\mathbb{P}}\mathbf{x}_{t-1} + \mathbf{u}^{\mathbb{P}}_t, \ \mathbf{u}^{\mathbb{P}}_t \overset{\mathbb{P}}{\sim} i.i.d.\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}). \tag{8}$$

Note that, in this case, by the Girsanov theorem, the conditional variance matrix of the factor innovations matrix, $\mathbf{\Sigma}$, defined in Eq.(1) is the same under different probability measures.

The *perceived* pricing function is determined by allowing the relation between the survey-expected shadow short rate and the survey-expected state variables to be driven by the perceived policy rule

$$E^{\mathbb{S}}_t[ssr_{t+h}] = \delta^{\mathbb{S}}_0 + \boldsymbol{\delta}^{\mathbb{S}'}_1 E^{\mathbb{S}}_t[\mathbf{x}_{t+h}], \tag{9}$$

[10]Note that, in our model, surveys and yields are spanned by the same set of state variables. Unaccounted unspanned risk factors can potentially account for different dynamics under the survey and the physical measure. Indeed, Duffee (2011) shows that much of the variation in expected bond returns is driven by a factor that is largely hidden from the current cross-section of yields, so that the term structure alone need not reveal all the information relevant for expectations and risk premia. Spanning tests reported in Appendix A do not find evidence of survey factors unspanned by observed yields.

which does not necessarily coincide with the historical policy rule determined by δ_0 and $\boldsymbol{\delta}_1$. For example, Sastry (2026) provides evidence that market participants perceive a policy rule that differs from the one the central bank follows.

Using the Wu and Xia (2016) shadow rate approximation of the one-period forward rate, the *perceived* one-period forward-rate pricing rule at $t+h$ is

$$f^{\mathbb{S}}(\mathbf{x}_{t+h}, n; \boldsymbol{\Psi}) \approx \underline{r} + \sigma_n^{\mathbb{S}} g\left(\frac{a_n^{\mathbb{S}} + \mathbf{b}_n^{\mathbb{S}\prime}\mathbf{x}_{t+h} - \underline{r}}{\sigma_n^{\mathbb{S}}}\right), \tag{10}$$

where, as in Eq. (6), $g(z) = z\mathtt{N}(z) + \mathtt{n}(z)$ with $\mathtt{N}(z)$ and $\mathtt{n}(z)$ the standard normal CDF and the PDF of z, respectively, and

$$\begin{aligned} a_n^{\mathbb{S}} &= a(\delta_0^{\mathbb{S}}, \boldsymbol{\delta}_1^{\mathbb{S}}, \boldsymbol{\mu}^{\mathbb{Q}}, \boldsymbol{\Phi}^{\mathbb{Q}}, \boldsymbol{\Sigma}, n) \\ \mathbf{b}_n^{\mathbb{S}} &= b(\boldsymbol{\delta}_1^{\mathbb{S}}, \boldsymbol{\Phi}^{\mathbb{Q}}, n) \\ \sigma_n^{\mathbb{S}} &= \sigma(\boldsymbol{\delta}_1^{\mathbb{S}}, \boldsymbol{\Phi}^{\mathbb{Q}}, \boldsymbol{\Sigma}, n). \end{aligned}$$

Thus, the time t survey expectation of the one-period forward rate at time $t+h$ for holding from $t+h+n$ is $E_t^{\mathbb{S}}[f^{\mathbb{S}}(\mathbf{x}_{t+h}, n; \boldsymbol{\Psi})]$, which is non-trivial because $f^{\mathbb{S}}(\mathbf{x}_{t+h}, n; \boldsymbol{\Psi})$ is a nonlinear function of $\mathbf{x}_{t+h}$. However, since $\mathbf{x}_{t+h} \mid \mathcal{F}_t$ is Gaussian under $\mathbb{S}$, the survey expectation at time t of the perceived forward rate at $t+h$ has a closed-form expression:

$$E_t^{\mathbb{S}}\left[f^{\mathbb{S}}(\mathbf{x}_{t+h}, n; \boldsymbol{\Psi})\right] \approx \underline{r} + \sigma_n^{\mathbb{S}}\sqrt{1 + s_{n,h}^2}\, g\left(\frac{a_n^{\mathbb{S}} + \mathbf{b}_n^{\mathbb{S}\prime} E_t^{\mathbb{S}}[\mathbf{x}_{t+h}] - \underline{r}}{\sigma_n^{\mathbb{S}}\sqrt{1 + s_{n,h}^2}}\right), \tag{11}$$

with

$$
\begin{aligned}
s^2_{n,h} &= \frac{\mathbf{b}_n^{\mathbb{S}\prime}\boldsymbol{\Sigma}_h^{\mathbb{S}}\mathbf{b}_n^{\mathbb{S}}}{(\sigma_n^{\mathbb{S}})^2}, \\
\boldsymbol{\Sigma}_h^{\mathbb{S}} &= \sum_{j=0}^{h-1}(\boldsymbol{\Phi}^{\mathbb{S}})^j\boldsymbol{\Sigma}(\boldsymbol{\Phi}^{\mathbb{S}\prime})^j.
\end{aligned}
$$

See Appendix B for more details. Also, notice that Eq. (11) has the same functional form as the original Wu and Xia (2016) formula, but with adjusted arguments.

The survey expectation of an n-period yield h-periods ahead within the shadow rate model is then

$$
y^s_{n,t,h} = \frac{1}{n}\sum_{j=0}^{n-1} E_t^{\mathbb{S}}[f^{\mathbb{S}}(\mathbf{x}_{t+h}, j; \boldsymbol{\Psi})]. \tag{12}
$$

Two remarks on Eq. (9) are in order. First, it does not describe the subjective expectation of the physically-defined shadow short rate $ssr_{t+h} = \delta_0 + \boldsymbol{\delta}_1'\mathbf{x}_{t+h}$, which would trivially preserve δ_0 and $\boldsymbol{\delta}_1$, since a change of measure re-weights the distribution of the states without altering the coefficients that define the short rate as a function of $\mathbf{x}$. Rather, $(\delta_0^{\mathbb{S}}, \boldsymbol{\delta}_1^{\mathbb{S}})$ is a separate primitive—the policy rule that agents *perceive*—so that the equality $(\delta_0^{\mathbb{S}}, \boldsymbol{\delta}_1^{\mathbb{S}}) = (\delta_0, \boldsymbol{\delta}_1)$ is a testable restriction (our Cases 1 and 3 below) rather than an identity. Second, the perceived reaction function is identified separately from the subjective dynamics: in Eq. (10), $\boldsymbol{\delta}_1^{\mathbb{S}}$ enters the cross-sectional loadings $\mathbf{b}_n^{\mathbb{S}} = b(\boldsymbol{\delta}_1^{\mathbb{S}}, \boldsymbol{\Phi}^{\mathbb{Q}}, n)$, jointly with the risk-neutral persistence $\boldsymbol{\Phi}^{\mathbb{Q}}$, whereas the subjective dynamics $(\boldsymbol{\mu}^{\mathbb{S}}, \boldsymbol{\Phi}^{\mathbb{S}})$ enter only through the conditional mean $E_t^{\mathbb{S}}[\mathbf{x}_{t+h}]$. Hence the cross-maturity shape of the survey forecasts identifies the perceived reaction function, while their evolution across horizons identifies the subjective dynamics.

3.4 Model specifications

Typically, in the literature, it is assumed that the perceived pricing function is the same as the historical pricing function, i.e. $\delta_0^{\mathbb{S}} = \delta_0$ and $\boldsymbol{\delta}_1^{\mathbb{S}} = \boldsymbol{\delta}_1$, and that the subjective dynamics are the same as the dynamics under the real-world probability measure, i.e. $\boldsymbol{\mu}^{\mathbb{S}} = \boldsymbol{\mu}^{\mathbb{P}}$ and $\boldsymbol{\Phi}^{\mathbb{S}} = \boldsymbol{\Phi}^{\mathbb{P}}$. These assumptions rest on the premise that market expectations can be "measured" from the history of bond prices, which in turn allows us to determine the risk premium. However, if the perceived pricing function differs from the historical pricing function and/or the subjective dynamics differ from the dynamics under the real-world probability measure, then two alternative decompositions of long-term interest rates emerge:

$$\begin{aligned} y_{n,t} &= \frac{1}{n}\sum_{j=0}^{n-1} E_t^{\mathbb{S}}[\max\{\underline{r}; \delta_0^{\mathbb{S}} + \boldsymbol{\delta}_1^{\mathbb{S}\prime}\mathbf{x}_{t+j}\}] + tp_{n,t}^{\mathbb{S}} \qquad (13) \\ &= \frac{1}{n}\sum_{j=0}^{n-1} E_t^{\mathbb{S}}[\max\{\underline{r}; \delta_0^{\mathbb{P}} + \boldsymbol{\delta}_1^{\mathbb{P}\prime}\mathbf{x}_{t+j}\}] + RFE_{n,t} + tp_{n,t}^{\mathbb{S}} \\ &= \frac{1}{n}\sum_{j=0}^{n-1} E_t^{\mathbb{P}}[\max\{\underline{r}; \delta_0^{\mathbb{P}} + \boldsymbol{\delta}_1^{\mathbb{P}\prime}\mathbf{x}_{t+j}\}] + PFE_{n,t} + RFE_{n,t} + tp_{n,t}^{\mathbb{S}} \\ &= \frac{1}{n}\sum_{j=0}^{n-1} E_t^{\mathbb{P}}[\max\{\underline{r}; \delta_0^{\mathbb{P}} + \boldsymbol{\delta}_1^{\mathbb{P}\prime}\mathbf{x}_{t+j}\}] + tp_{n,t}^{\mathbb{P}} \qquad (14) \end{aligned}$$

where the reaction function error, $RFE_{n,t}$, is defined as

$$RFE_{n,t} = \frac{1}{n}\sum_{j=0}^{n-1}\left[E_t^{\mathbb{S}}[\max\{\underline{r}; \delta_0^{\mathbb{S}} + \boldsymbol{\delta}_1^{\mathbb{S}\prime}\mathbf{x}_{t+j}\}] - E_t^{\mathbb{S}}[\max\{\underline{r}; \delta_0^{\mathbb{P}} + \boldsymbol{\delta}_1^{\mathbb{P}\prime}\mathbf{x}_{t+j}\}]\right], \qquad (15)$$

the probability forecast error, $PFE_{n,t}$, is defined as

$$PFE_{n,t} = \frac{1}{n}\sum_{j=0}^{n-1}\left[E_t^{\mathbb{S}}[\max\{\underline{r}; \delta_0^{\mathbb{P}} + \boldsymbol{\delta}_1^{\mathbb{P}\prime}\mathbf{x}_{t+j}\}] - E_t^{\mathbb{P}}[\max\{\underline{r}; \delta_0^{\mathbb{P}} + \boldsymbol{\delta}_1^{\mathbb{P}\prime}\mathbf{x}_{t+j}\}]\right], \qquad (16)$$

	Equal dynamics ($\mathbb{S} \sim \mathbb{P}$) $\boldsymbol{\mu}^{\mathbb{P}} = \boldsymbol{\mu}^{\mathbb{S}},\ \boldsymbol{\Phi}^{\mathbb{P}} = \boldsymbol{\Phi}^{\mathbb{S}}$	Different dynamics ($\mathbb{S} \nsim \mathbb{P}$) $\boldsymbol{\mu}^{\mathbb{P}} \neq \boldsymbol{\mu}^{\mathbb{S}},\ \boldsymbol{\Phi}^{\mathbb{P}} \neq \boldsymbol{\Phi}^{\mathbb{S}}$
Equal reaction $\delta_0 = \delta_0^{\mathbb{S}},\ \boldsymbol{\delta}_1 = \boldsymbol{\delta}_1^{\mathbb{S}}$	**Case 1** $tp_{n,t}^{\mathbb{P}} = tp_{n,t}^{\mathbb{S}}$	**Case 3** $tp_{n,t}^{\mathbb{P}} = tp_{n,t}^{\mathbb{S}} + PFE_{n,t}$
Different reaction $\delta_0 \neq \delta_0^{\mathbb{S}},\ \boldsymbol{\delta}_1 \neq \boldsymbol{\delta}_1^{\mathbb{S}}$	**Case 2** $tp_{n,t}^{\mathbb{P}} = tp_{n,t}^{\mathbb{S}} + RFE_{n,t}$	**Case 4** $tp_{n,t}^{\mathbb{P}} = tp_{n,t}^{\mathbb{S}} + RFE_{n,t} + PFE_{n,t}$

Case 5
$\mu_{\infty}^{\mathbb{P}} = \mu_{\infty}^{\mathbb{S}}$

Table 1. Model specifications.

and the term premium under the physical measure is given by

$$tp_{n,t}^{\mathbb{P}} = RFE_{n,t} + PFE_{n,t} + tp_{n,t}^{\mathbb{S}}. \tag{17}$$

Equations (13) and (14) are two alternative decompositions of observed long-term interest rates. Equation (14) is the standard decomposition of observed long-term interest rates into physical term premium and physical expectations, while Equation (13) is the decomposition of observed long-term interest rates into the survey term premium and survey expectations. These two decompositions coincide only when both the reaction function errors and the probability forecast errors are zero, i.e. under the assumptions of equal pricing function and equal probability measures.

In our framework, both these assumptions can be tested. To facilitate the interpretation of these assumptions and their implications, in Table 1 we tabulate the four different cases that arise, together with an intermediate specification, Case 5, described below. In Case 1,

the perceived pricing function coincides with the historical pricing function and the subjective dynamics coincide with the dynamics under the real-world probability measure. In this case, reaction function errors and probability forecast errors are zero, so the physical term premium and the survey term premium coincide.

In Case 2, the pricing functions are not equal but we maintain the assumption of equal dynamics. Survey participants misestimate the response of the short rate to the state variables, and thus make an error when making their predictions of future interest rates. This implies that the objective term premium does not coincide with the survey term premium due to the presence of survey reaction function errors. In Case 3, we maintain the assumption that the perceived pricing function coincides with the historical pricing function but the subjective and the objective probability measures are not equal. In this case, survey participants make an error when predicting the state variables, which then translates into an error in their predicted interest rates. This implies that the objective term premium does not coincide with the survey term premium due to the presence of survey probability forecast errors. Finally, in Case 4, the pricing functions are different and also the subjective and the objective probability measures are different. This implies that agents make errors both due to misestimating the response of the short rate to the state variables and in the predictions of the state variables. This implies that the objective term premium does not coincide with the survey term premium due to both the presence of survey reaction function and survey probability forecast errors. This specification corresponds to the general model presented in the previous section, in which all parameters are freely estimated. As such, it nests all the other cases and allows for statistical testing of the associated restrictions.

We also consider an intermediate specification, Case 5, in which both the reaction function and the dynamics are allowed to differ between the survey and the real-world measures, but the long-run mean of the shadow short rate is restricted to be the same under the two

measures, i.e. $\mu_\infty^{\mathbb{P}} = \mu_\infty^{\mathbb{S}}$, where

$$\mu_\infty^{\mathbb{P}} = \delta_0 + \boldsymbol{\delta}_1' \left(\mathbf{I} - \boldsymbol{\Phi}^{\mathbb{P}}\right)^{-1} \boldsymbol{\mu}^{\mathbb{P}} \quad \text{and} \quad \mu_\infty^{\mathbb{S}} = \delta_0^{\mathbb{S}} + \boldsymbol{\delta}_1^{\mathbb{S}\prime} \left(\mathbf{I} - \boldsymbol{\Phi}^{\mathbb{S}}\right)^{-1} \boldsymbol{\mu}^{\mathbb{S}}.$$

In this case, survey participants can misestimate both the response of the short rate to the state variables and the dynamics of the state variables, but they correctly anticipate the level to which the short rate reverts in the long run. Their errors in predicting the shadow short rate are therefore transitory, and vanish as the forecast horizon increases. The restriction $\mu_\infty^{\mathbb{P}} = \mu_\infty^{\mathbb{S}}$ is a single restriction that jointly involves the parameters of the dynamics and of the reaction functions, and it does not require either of them to be equal across the two measures, so Case 5 does not belong to any single cell of the table. As such, the representation of Case 5 in Table 1 is only schematic. We place it at the intersection of the four cells to indicate that it lies between Case 1 and Case 4: it nests Case 1, in which the long-run means coincide because both the dynamics and the reaction functions are equal, and it is nested in Case 4, from which it differs by one restriction. It neither nests nor is nested in Case 2 or Case 3, since under either of these specifications the long-run means generally differ. The framework of Piazzesi et al. (2015), in which $\boldsymbol{\mu}^{\mathbb{P}} = \boldsymbol{\mu}^{\mathbb{S}} = \mathbf{0}$ and the reaction function is common to both measures, is a special case of Case 5, while that of Connolly and Struby (2026), in which $\boldsymbol{\mu}^{\mathbb{P}} = \boldsymbol{\mu}^{\mathbb{S}}$ but the persistence of the state variables differs, is a special case of Case 3.

Notice that the physical term premium and the survey term premium coincide only in Case 1, when reaction function errors and probability forecast errors are zero. In all the other cases, there is a wedge between the two. This implies that the subjective term premium represents the compensation required by market participants, even if their forecasts are erroneous, whereas the objective term premium corresponds to the measure an econo-

metrician infers from realised bond prices.

4 Estimation and identification

We denote the vector of model-implied zero-coupon yields at time t by $\mathbf{y}_t$ and the associated vector of pricing errors by $\mathbf{v}_t \sim N(\mathbf{0}, \mathbf{\Sigma_y})$, such that the observed yields are $\widetilde{\mathbf{y}}_t = \mathbf{y}_t + \mathbf{v}_t$. Similarly, we denote the vector of observable surveys with different maturities and forecasting horizons as $\widetilde{\mathbf{y}}_t^s = \mathbf{y}_t^s + \mathbf{v}_t^s$, where entries of the model-implied surveys $\mathbf{y}_t^s$ are given in Eq.(12) and $\mathbf{v}_t^s$ are the associated zero mean measurement errors with variance $\mathbf{\Sigma_s}$.

Following Joslin, Singleton and Zhu (2011), we assume that the first three principal components of yields are measured without error. If the eigenvectors associated with the three largest eigenvalues of the covariance matrix of observed yields are collected in an $N \times 3$ matrix $\mathbf{W}$ (where N is the number of yields), the first three principal components of observed yields are given by $\mathbf{q}_{y,t} \equiv \mathbf{W}'\widetilde{\mathbf{y}}_t = \mathbf{W}'\mathbf{y}_t$, where the last equality follows from the observability assumption ($\mathbf{W}'\mathbf{v}_t = \mathbf{0}$). As shown in Golinski and Spencer (2024), this observability assumption allows us to extract the underlying state vector $\mathbf{x}_t$ from the nonlinear system, conditionally on the risk-neutral parameters. Accordingly, $\mathbf{v}_t$ is restricted to have zero variance in the $\mathbf{W}$ directions, and the density for $\mathbf{v}_t$ is understood as the Gaussian density for the $N-3$ noisy yield combinations, conditional on the exactly measured portfolios $\mathbf{q}_{y,t} = \mathbf{W}'\widetilde{\mathbf{y}}_t$.

We adopt the parameterization of the model of Joslin et al. (2011) in which the $\mathbb{Q}$ dynamics depend on $\{\mu_1^{\mathbb{Q}}, \boldsymbol{\lambda}^{\mathbb{Q}}, \mathbf{\Sigma}\}$, where $\mu_1^{\mathbb{Q}}$ is the first element of $\boldsymbol{\mu}^{\mathbb{Q}}$, other elements being zeros, and $\boldsymbol{\lambda}^{\mathbb{Q}}$ is the vector of eigenvalues of $\mathbf{\Phi}^{\mathbb{Q}}$. The shadow short rate Eq. (4) is thus identified with $\delta_0 = 0$ and $\boldsymbol{\delta}_1 = \mathbf{1}$, where $\mathbf{1}$ denotes a vector of ones.

We estimate the joint model of zero-coupon yields and interest rate survey expectations by maximum likelihood. We implement it using the factor extraction method proposed

by Golinski and Spencer (2024), which generalizes the estimation technique with observable factors introduced by Joslin et al. (2011) to non-linear models. This method does not require any filtering approximation (such as, e.g. that required by the extended Kalman filter) and it is computationally efficient since it allows for concentrating the likelihood function with respect to the $\mathbb{P}$ parameters.[11] Further details on the estimation procedure can be found in Appendix D.

5 Results

In this section, we report the empirical results focusing on the comparison between the five different cases in Table 1, and also Case 0 (the yield-only model). After reporting the model fit for each case, we perform formal model specification testing and inspect the model estimates. We then turn to the general model (Case 4) and assess the estimates of the reaction function and probability forecast errors, and their implications for long-term rates decompositions. Finally, we assess the ability of the different model specifications to predict realized returns and rates.

5.1 Fit to data

In Tables 2 and 3, we report root mean-square fitting errors for yields and surveys across the cases tabulated in Table 1, as well as for the yield-only model (Case 0). The first observation we can make is that the fit for yields reported in Table 2 is very similar across models, regardless of whether surveys are included or not in the model. In contrast, Table 3

[11] The identification strategy relies on treating the first three yield PCs as error-free and solving a nonlinear inversion for $\mathbf{x}_t$. In a shadow-rate setting, the derivative $\partial \mathbf{q}_t / \partial \mathbf{x}_t'$ might be ill-conditioned, potentially undermining numerical stability. To assess this risk, we examine the condition number of the corresponding Jacobian term. Across all specifications, the highest observed condition number is 686 (Case 2), a value that remains well within the range considered numerically stable.

RMSE

Model	3m	6m	1y	2y	5y	7y	10y	Av
Case 0 (yield-only)	5.03	2.30	6.14	3.75	5.55	3.69	4.97	4.49
Case 1 ($RFE = 0$, $PFE = 0$)	5.05	2.31	6.09	3.83	5.62	3.76	5.09	4.53
Case 2 ($PFE = 0$)	5.03	2.30	6.07	3.83	5.62	3.74	5.08	4.53
Case 3 ($RFE = 0$)	5.05	2.31	6.10	3.82	5.63	3.76	5.09	4.54
Case 4 (unrestricted)	5.08	2.31	6.11	3.81	5.67	3.73	5.11	4.55
Case 5 ($\mu_\infty^{\mathbb{P}} = \mu_\infty^{\mathbb{S}}$)	5.08	2.31	6.11	3.80	5.67	3.73	5.11	4.54

Table 2. Model fit for yields measured by the root mean-square error ($RMSE$). The numbers are reported in basis points. The specifications for Case 1 to Case 5 are detailed in Table 1.

shows that the yield-only model (Case 0) is unable to fit the surveys, implying that including surveys improves the fit of surveys, regardless of model specification. Among the models that include surveys, the specifications that allow for a perceived reaction function (Cases 2, 4 and 5) fit surveys better than those that do not (Cases 1 and 3), especially at short maturities and horizons, whereas allowing for different dynamics improves the fit only marginally (compare Case 1 with Case 3, and Case 2 with Case 4). Imposing equal long-run means under the two measures (Case 5) has virtually no effect on the fit: its root mean-square errors for both yields and surveys are almost identical to those of Case 4.

In Figure 2 we report survey expectations for the 1-year yield, the subjective expectations from Case 4, and the realized 1-year yield. In contrast to Figure 1, the maximally flexible model is able to replicate survey expectations, further supporting the evidence that survey expectations are different from those extrapolated from historical yields. Furthermore, we verify that the model's strong fit to survey expectations is not driven by the use of full-sample information in forecasting. To this end, we re-estimate the model using a rolling-window scheme described in Section 5.6. As shown in Figure G-1 (see Appendix G), the resulting forecasts are very similar to those reported in Figure 2.

Model	h	3m	6m	1y	2y	5y	10y	Av
Case 0	1q	19.95	26.34	31.81	23.43	24.53	23.21	24.88
(yield-only)	2q	29.29	35.95	41.85	34.13	34.77	31.43	34.57
	3q	40.42	47.37	54.29	47.53	47.05	42.56	46.54
	4q	55.01	62.32	69.32	63.01	60.54	54.99	60.87
Case 1	1q	15.15	19.69	24.56	20.44	22.94	24.02	21.13
(RFE=PFE=0)	2q	20.34	23.24	27.09	25.11	26.52	26.48	24.80
	3q	24.66	25.52	29.02	28.85	28.75	28.06	27.48
	4q	30.79	30.61	32.23	32.08	30.56	30.17	31.07
Case 2	1q	14.49	17.98	22.70	20.55	22.61	23.93	20.38
(PFE=0)	2q	19.95	22.39	26.07	24.91	26.27	26.50	24.35
	3q	24.64	25.53	28.94	28.83	28.74	28.06	27.46
	4q	29.74	30.04	32.13	31.74	30.58	30.03	30.71
Case 3	1q	15.11	19.66	24.54	20.44	22.94	24.02	21.12
(RFE=0)	2q	20.31	23.20	27.06	25.11	26.53	26.47	24.78
	3q	24.70	25.54	29.01	28.86	28.76	28.05	27.49
	4q	30.92	30.70	32.26	32.10	30.58	30.16	31.12
Case 4	1q	14.46	17.89	22.63	20.67	22.62	23.88	20.36
(unrestricted)	2q	19.93	22.32	25.99	24.92	26.26	26.51	24.32
	3q	24.51	25.42	28.89	28.83	28.73	28.08	27.41
	4q	29.40	29.79	32.04	31.72	30.56	30.05	30.59
Case 5	1q	14.47	17.86	22.57	20.69	22.63	23.91	20.36
($\mu_\infty^{\mathbb{P}} = \mu_\infty^{\mathbb{S}}$)	2q	19.93	22.31	25.97	24.94	26.26	26.51	24.32
	3q	24.50	25.42	28.89	28.83	28.72	28.09	27.41
	4q	29.38	29.79	32.04	31.71	30.56	30.08	30.59

Table 3. Model fit for surveys measured by the root mean-square error ($RMSE$). The numbers are reported in basis points. The specifications for Case 1 to Case 5 are detailed in Table 1.

5.2 Testing the survey restrictions

From the cross-sectional fit statistics presented in Tables 2 and 3 it is clear that as long as we have surveys in the model, the fitting errors do not provide sufficient information to discriminate among different hypotheses. Thus, for statistical comparison, we use the likelihood ratio test of each restricted model (Cases 1, 2, 3 and 5) against the unrestricted model (Case 4), complemented by information criteria.

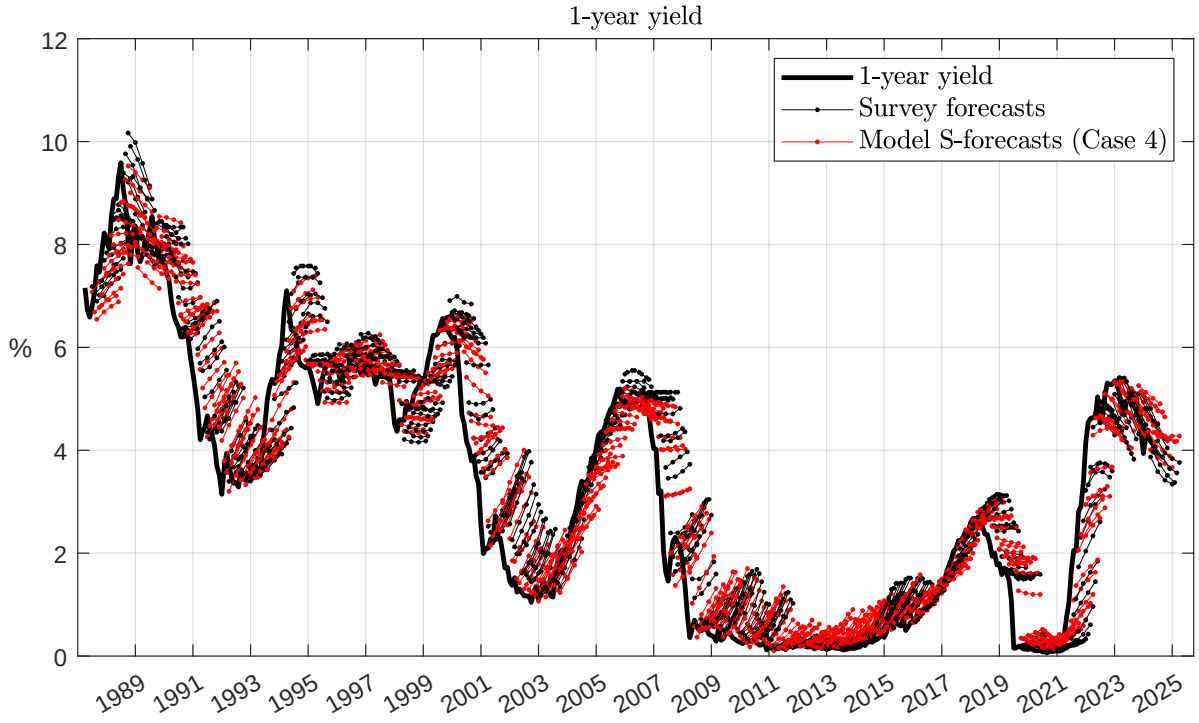


Figure 2. Market expectations measured by surveys (black dotted lines) and in-sample subjective expectations extracted from the Case 4 model which allows for different pricing functions and probability measures (red dotted lines), along with the observed 1-year yield (black bold line).

In Table 4, we report the value of the log-likelihood function at the estimated maximum as well as its components as in Eq.(D-7). We also report the number of restrictions that each restricted model, Cases 1, 2, 3 and 5, imposes relative to the unrestricted alternative, Case 4, together with the bootstrap p-values of the corresponding likelihood ratio statistics. The bootstrap is carried out under the corresponding null model in each case (Case 1, 2, 3 or 5), as described in Appendix E.

The likelihood ratio tests reject Cases 1, 2 and 3 at the 5% level or smaller in favour of the unrestricted model. The inspection of the likelihood decomposition allows us to explore the reasons for these rejections. Confirming our previous informal analysis of the fit of the model

Model	Total L	$L^{\mathbb{Q}}$	$L^{\mathbb{S}}$	$L^{\mathbb{P}}$	# Restr.	LR test	ΔAIC	ΔBIC
Case 1 (RFE=PFE=0)	64,886.32	10,595.00	48,116.68	6,174.64	16	0.0030	285.4	254.3
Case 2 (PFE=0)	65,025.67	10,597.71	48,264.15	6,163.81	12	0.0390	14.7	0.0
Case 3 (RFE=0)	64,901.60	10,594.02	48,112.67	6,194.91	4	0.0030	278.8	296.9
Case 5 ($\mu_\infty^{\mathbb{P}} = \mu_\infty^{\mathbb{S}}$)	65,042.41	10,589.44	48,286.66	6,166.31	1	0.3900	3.2	33.6
Case 4 (unrestricted)	65,045.00	10,588.62	48,286.01	6,170.37	-	-	0.0	34.5

Table 4. Likelihood ratio test and information criteria. Total likelihood values (first column), and its components as in Equation (D-7), the number of restrictions relative to Case 4 (Restr.), and likelihood ratio bootstrap p-values for the Likelihood Ratio test statistics with respect to Case 4 as the unrestricted model. The last two columns report the Akaike and Bayesian information criteria, each expressed as the difference from its minimum across the models. The specifications for Case 1 to Case 5 are detailed in Table 1.

in Table 2, we can see that the parts of the likelihood corresponding to the cross-sectional fit of yields ($L^{\mathbb{Q}}$) are similar for all cases. However, allowing for reaction function errors (Cases 2, 4 and 5) increases the log-likelihood along the survey-fit dimension. Although Case 3 achieves the highest $L^{\mathbb{P}}$, the restriction RFE=0 limits its ability to fit survey expectations, leading to its rejection in favor of the unrestricted Case 4.

In contrast, the likelihood ratio test does not reject Case 5, which imposes a single restriction, the equality of the long-run means of the shadow short rate under the $\mathbb{P}$ and $\mathbb{S}$ measures (p-value of 0.39).[12] The restriction reduces the total log-likelihood by only 2.6. The loss is concentrated in the $\mathbb{P}$ component, and is partly offset by marginal gains in the fit of yields and surveys. This result, however, does not imply that the data support the restriction. The long-run mean of a highly persistent process is only weakly identified by the likelihood, which is therefore nearly flat along this dimension. As we discuss in Section 5.3

[12] As for the other restricted models, the bootstrap distribution of the likelihood ratio statistic under Case 5 is considerably more dispersed than its asymptotic χ^2 counterpart: the asymptotic $\chi^2(1)$ p-value would be 0.02.

(see Table 7), imposing the restriction pulls the long-run mean of the short rate under the $\mathbb{P}$ measure up to the level implied by surveys, well above the average level of short-term rates in our sample.

The information criteria reported in the last two columns of Table 4 provide a complementary comparison, which penalises model size and does not require the models to be nested. Hence, they also allow us to rank the non-nested Cases 2, 3 and 5. Both criteria decisively discard Cases 1 and 3, the two specifications without a perceived reaction function. The Akaike criterion selects the unrestricted Case 4, closely followed by Case 5, in line with the likelihood ratio tests. The Bayesian criterion, instead, selects Case 2: allowing the survey dynamics to differ from the physical ones raises the log-likelihood by about 19 relative to Case 2, which is less than the Bayesian penalty for the 12 additional parameters. The evidence for a perceived reaction function is therefore robust across all criteria, while the evidence for different dynamics depends on how model size is penalised: it is supported by the bootstrap likelihood ratio test, which accounts for the finite-sample distribution of the test statistic, and by the Akaike criterion, but not by the more parsimonious Bayesian criterion.

5.3 Model estimates

We now inspect the factor dynamics in each of the cases considered. In Table 5, we present the estimates of factor dynamics under the real-world measure.[13] The last column reports the eigenvalues of the autoregressive matrix, $\mathbf{\Phi}^{\mathbb{P}}$. In all cases, these eigenvalues lie within the unit circle, confirming stationarity. Across models, the largest eigenvalue is approximately 0.98, a value commonly observed in the term structure literature. The lowest persistence, corresponding to a modulus of the maximum eigenvalue equal to 0.9797, is found in the fully

[13] The estimates of the risk-neutral parameters are reported in Table F-1 in Appendix F.

Model	$\boldsymbol{\mu}^{\mathbb{P}}$		$\boldsymbol{\Phi}^{\mathbb{P}}$		$Eig(\Phi^{\mathbb{P}})$
Case 0	0.0003	0.9920	0.0160	0.0161	0.9815
(yield-only model)	-0.0003	0.0215	0.7834	-0.1551	0.9730
	0.0006	-0.0379	0.1446	1.0882	0.9090
Case 1	0.0007	0.9820	-0.0066	-0.0058	0.9324
(RFE=PFE=0)	0.0018	-0.2104	0.9095	-0.0286	0.9797+
	-0.0020	0.2156	0.0605	1.0003	$\pm\, 0.0051\, i$
Case 2	0.0006	0.9830	-0.0093	-0.0076	0.9813
(PFE=0)	-0.0006	-0.0932	0.8588	-0.0637	0.9716
	0.0004	0.0984	0.0991	1.0255	0.9145
Case 3	0.0003	0.9918	0.0161	0.0162	0.9816
(RFE=0)	-0.0001	0.0129	0.8480	-0.0942	0.9732
	0.0004	-0.0296	0.0760	1.0260	0.9111
Case 4	0.0002	0.9925	0.0169	0.0171	0.9817
(unrestricted)	-0.0004	0.0156	0.8676	-0.0767	0.9732
	0.0008	-0.0324	0.0571	1.0095	0.9147
Case 5	0.0002	0.9938	0.0187	0.0188	0.9876
($\mu_\infty^{\mathbb{P}} = \mu_\infty^{\mathbb{S}}$)	-0.0008	-0.0211	0.8165	-0.1259	0.9759
	0.0012	0.0084	0.1143	1.0644	0.9112

Table 5. Estimates of physical parameters. The table reports the estimates of the intercept vector $\boldsymbol{\mu}^{\mathbb{P}}$, the autoregressive matrix $\boldsymbol{\Phi}^{\mathbb{P}}$ and the eigenvalues of the autoregressive matrix.

restricted specification (Case 1). In contrast, the highest persistence occurs in the models that allow for distinct dynamics under the $\mathbb{P}$ and $\mathbb{S}$ probability measures (Cases 3, 4 and 5), as well as in the yield-only model (Case 0), with a maximum eigenvalue of about 0.982 in Cases 0, 3 and 4, and of 0.988 in Case 5. The higher persistence in Case 5 is consistent with the restriction on the long-run mean: slower mean reversion allows the model to reconcile a higher long-run mean with the prolonged period of low interest rates in our sample.

Table 6 reports the estimates of the survey parameters. Two features stand out. First, whenever it is estimated (Cases 2, 4 and 5), the perceived reaction function differs from the physical one, $\delta_0 = 0$ and $\boldsymbol{\delta}_1 = \mathbf{1}$, mainly through the loadings on the second and third factors, which are about 10% larger than under the physical measure, while the loading on the first factor is essentially one and the intercept is close to zero. Survey participants thus perceive

Model	$\delta_0^{\mathbb{S}} \times 100$	$\boldsymbol{\delta}_1^{\mathbb{S}}$	$\boldsymbol{\mu}^{\mathbb{S}}$		$\boldsymbol{\Phi}^{\mathbb{S}}$		$Eig(\Phi^{\mathbb{S}})$
Case 2	0.0317	0.9999	0.0006	0.9830	-0.0093	-0.0076	0.9813
(PFE=0)		1.1121	-0.0006	-0.0932	0.8588	-0.0637	0.9716
		1.1032	0.0004	0.0984	0.0991	1.0255	0.9145
Case 3	0.0000	1.0000	0.0007	0.9819	-0.0067	-0.0053	0.9337
(RFE=0)		1.0000	0.0011	-0.1213	0.9388	0.0007	0.9796+
		1.0000	-0.0012	0.1269	0.0311	0.9721	$\pm 0.0053\,i$
Case 4	0.0144	1.0019	0.0006	0.9828	-0.0089	-0.0074	0.9813
(unrestricted)		1.1123	-0.0003	-0.0919	0.8744	-0.0534	0.9714
		1.1042	0.0002	0.0967	0.0836	1.0152	0.9198
Case 5	0.0153	1.0021	0.0006	0.9828	-0.0092	-0.0076	0.9810
($\mu_\infty^{\mathbb{P}} = \mu_\infty^{\mathbb{S}}$)		1.1130	-0.0004	-0.0914	0.8716	-0.0559	0.9719
		1.1047	0.0003	0.0962	0.0865	1.0179	0.9193

Table 6. Estimates of survey parameters. The table reports the estimates of the intercept and loadings of the reaction function, $\delta_0^{\mathbb{S}}$ and $\boldsymbol{\delta}_1^{\mathbb{S}}$, respectively, the intercept vector and the autoregressive matrix of the $\mathbb{S}$ dynamics, $\boldsymbol{\mu}^{\mathbb{S}}$ and $\boldsymbol{\Phi}^{\mathbb{S}}$, respectively, and the eigenvalues of the autoregressive matrix.

the short rate to respond more strongly to the factors that are less persistent under the risk-neutral measure, which is consistent with reaction function errors being concentrated at short maturities (Section 5.4). Second, the estimated survey dynamics are very similar across specifications: Cases 2, 4 and 5 deliver nearly identical estimates of $\boldsymbol{\mu}^{\mathbb{S}}$ and $\boldsymbol{\Phi}^{\mathbb{S}}$, and the survey dynamics in Case 3 nearly coincide with the common dynamics estimated in Case 1 (Table 5). Surveys therefore pin down the subjective dynamics robustly, which, in Cases 1 and 2, are imposed on the physical measure as well.

We now turn to the long-run means implied by these estimates, which are reported in Table 7. The first column of Panel A reports the long-run mean of the shadow short rate under the real-world probability measure, computed as

$$\mu_\infty^{\mathbb{P}} = \mathbf{1}' \left(\mathbf{I} - \boldsymbol{\Phi}^{\mathbb{P}}\right)^{-1} \boldsymbol{\mu}^{\mathbb{P}}. \tag{18}$$

When surveys are excluded from the estimation (Case 0), the physical long-run mean is estimated at 2.47%. Incorporating surveys and restricting the $\mathbb{P}$ and $\mathbb{S}$ dynamics substantially raises this estimate to 5.13% and 4.45% for Case 1 and Case 2, respectively. In contrast, when allowing for probability forecast errors, the estimates revert to 2.48% and 2.49% for Case 3 and Case 4, respectively, closely aligning with those obtained from the yield-only model. Finally, imposing equal long-run means under the two measures (Case 5) raises the physical long-run mean to 4.47%, close to the values obtained in Cases 1 and 2.

The long-run mean of the shadow short rate under the survey probability measure is computed as

$$\mu_\infty^{\mathbb{S}} = \delta_0^{\mathbb{S}} + \boldsymbol{\delta}_1^{\mathbb{S}\prime} \left(\mathbf{I} - \boldsymbol{\Phi}^{\mathbb{S}}\right)^{-1} \boldsymbol{\mu}^{\mathbb{S}} \tag{19}$$

and is reported in the first column of Panel B of Table 7. The differences between these estimates and those under the $\mathbb{P}$ measure in Panel A reflect reaction function errors in Case 2, probability forecast errors in Case 3, and both sources of error in Cases 4 and 5. When the dynamics are allowed to differ, the long-run mean under the $\mathbb{S}$ measure is substantially higher than under the $\mathbb{P}$ measure: 5.13% against 2.48% in Case 3, and 4.52% against 2.49% in Case 4. In Case 2, where the dynamics are equal, the two long-run means are close (4.51% and 4.45%), as the physical dynamics inherit the high long-run mean implied by surveys. In Case 5, the long-run means under the two measures are equal by construction, at 4.47%.

These long-run means refer to the shadow short rate. To assess what they imply for observed interest rates, the remaining columns of Table 7 report the long-run means of yields across maturities under the two measures, which account for the lower bound, together with the sample averages of observed yields and of survey forecasts. Under the physical measure, the models in which the dynamics are not tied to surveys (Cases 0, 3 and 4) imply long-run means between 2.7% and 2.8% for the 3-month yield and between 4.0% and 4.1% for the 10-

Panel A: real-world measure ($\mathbb{P}$)

	$\mu_\infty^{\mathbb{P}}$	3m	6m	1y	2y	5y	10y
Data		2.99	3.10	3.26	3.46	4.00	4.58
Case 0 (yield-only)	2.47	2.67	2.72	2.82	2.99	3.43	3.98
Case 1 ($RFE = 0$, $PFE = 0$)	5.13	5.18	5.22	5.27	5.32	5.38	5.57
Case 2 ($PFE = 0$)	4.45	4.53	4.57	4.64	4.73	4.87	5.14
Case 3 ($RFE = 0$)	2.48	2.72	2.77	2.87	3.04	3.47	4.01
Case 4 (unrestricted)	2.49	2.78	2.84	2.94	3.12	3.55	4.07
Case 5 ($\mu_\infty^{\mathbb{P}} = \mu_\infty^{\mathbb{S}}$)	4.47	4.66	4.72	4.83	4.98	5.27	5.64

Panel B: survey measure ($\mathbb{S}$)

	$\mu_\infty^{\mathbb{S}}$	3m	6m	1y	2y	5y	10y
Surveys (4q-ahead)		3.33	3.49	3.69	3.83	4.32	4.83
Case 1 ($RFE = 0$, $PFE = 0$)	5.13	5.18	5.22	5.27	5.32	5.38	5.57
Case 2 ($PFE = 0$)	4.51	4.59	4.64	4.71	4.80	4.94	5.20
Case 3 ($RFE = 0$)	5.13	5.18	5.21	5.27	5.32	5.37	5.57
Case 4 (unrestricted)	4.52	4.60	4.65	4.73	4.82	4.96	5.22
Case 5 ($\mu_\infty^{\mathbb{P}} = \mu_\infty^{\mathbb{S}}$)	4.47	4.56	4.61	4.69	4.78	4.92	5.17

Table 7. Long-run means. The table reports the long-run means, in percent, implied by each model under the real-world ($\mathbb{P}$, Panel A) and the survey ($\mathbb{S}$, Panel B) probability measures. The first column reports the long-run mean of the shadow short rate, $\mu_\infty^{\mathbb{P}}$ in Panel A and $\mu_\infty^{\mathbb{S}}$ in Panel B. The remaining columns report the long-run means of yields, computed as the model-implied forecasts of each yield at a 200-year horizon, evaluated at the full-sample parameter estimates; these forecasts account for the lower bound on interest rates and do not depend on the forecast origin. In Case 0 only the $\mathbb{P}$ measure is defined, and in Case 1 the two measures coincide. For comparison, the first row of Panel A reports the sample mean of observed yields, and the first row of Panel B the sample mean of the four-quarter-ahead survey forecasts, both over December 1987 to December 2024.

year yield, somewhat below the sample averages of 3.0% and 4.6%, respectively. The survey measure implies considerably higher long-run means, between 4.6% and 5.2% for the 3-month yield and between 5.2% and 5.6% for the 10-year yield across Cases 2 to 5. The long-range forecasts published by the Blue Chip Financial Forecasts each June and December, which refer to horizons of 5 to 10 years ahead, provide a direct survey measure of long-horizon expectations. Over December 1996 to December 2024, they average 3.7% for the 3-month

yield and 4.8% for the 10-year yield, also above the physical long-run means, confirming that survey participants expect interest rates to revert to higher levels than those implied by the physical dynamics. The model-implied survey long-run means are higher still, although this comparison is only indicative, since the long-range forecasts refer to finite horizons and cover a shorter sample that excludes the high interest rates of the late 1980s and early 1990s.

Table 7 also illustrates the consequences of restricting the physical dynamics with survey information. In Cases 1, 2 and 5, the physical long-run means inherit the high level of the survey measure. In Case 5, for instance, the long-run mean of the 3-month yield under $\mathbb{P}$ is 4.7%, about 1.7 percentage points above its sample average, and that of the 10-year yield is 5.6%. Forecasts from these specifications therefore revert towards levels well above those observed over most of our sample, which contributes to their weaker forecasting performance in Section 5.6. At the same time, the restriction barely affects the fit of the model at the monthly frequency, which explains why the likelihood ratio test in Section 5.2 does not reject it.

5.4 Reaction function and probability forecast errors

To assess the extent to which reaction function errors and probability forecast errors contribute to the dynamics of survey expectations, we report their summary statistics in Panel A of Table 8. The table indicates that although average reaction function errors are close to zero for all survey maturities, they display substantial variability for short-term rates, as reflected in a standard deviation of about 15 bp, and the minimum and maximum larger than $\pm$30 bp. The average probability forecast errors instead increase with the maturity of the bond up to more than one percentage point for the 10-year rate. This indicates that while both errors are important, the reaction function error is dominant for short maturities

Panel A: Estimated Survey Forecast Errors														
	RFE							PFE						
Maturity	3m	6m	1y	2y	5y	7y	10y	3m	6m	1y	2y	5y	7y	10y
Mean	-0.02	-0.02	-0.02	-0.01	0.01	0.02	0.03	0.03	0.07	0.16	0.34	0.79	0.99	1.18
p-val (f.s.)	34.00	32.40	32.50	43.00	96.70	74.70	48.40	1.00	1.00	0.90	0.70	2.60	4.20	7.00
p-val (s.s.)	47.30	46.00	47.90	58.80	94.50	85.60	68.90	13.00	12.80	10.80	11.10	14.20	19.00	24.40
Std	0.16	0.15	0.13	0.12	0.08	0.06	0.04	0.04	0.10	0.19	0.36	0.61	0.65	0.62
Min	-0.40	-0.38	-0.34	-0.27	-0.14	-0.09	-0.05	-0.13	-0.30	-0.56	-0.83	-0.83	-0.63	-0.29
Max	0.36	0.35	0.32	0.27	0.18	0.15	0.12	0.11	0.28	0.61	1.17	2.15	2.41	2.52
Panel B: Bias (Full Sample)														
	RFE Bias							PFE Bias						
Maturity	3m	6m	1y	2y	5y	7y	10y	3m	6m	1y	2y	5y	7y	10y
Mean	0.01	0.01	0.01	0.01	0.01	0.01	0.01	-0.01	-0.01	-0.03	-0.04	-0.04	-0.02	0.00
Std	0.06	0.06	0.05	0.04	0.03	0.02	0.02	0.02	0.05	0.10	0.18	0.35	0.39	0.41
Min	-0.08	-0.08	-0.08	-0.07	-0.07	-0.08	-0.09	-0.06	-0.13	-0.26	-0.47	-1.16	-1.52	-1.99
Max	0.14	0.13	0.13	0.14	0.14	0.14	0.15	0.04	0.09	0.21	0.47	1.26	1.71	2.21
Panel C: Bias (Short Sample)														
	RFE Bias							PFE Bias						
Maturity	3m	6m	1y	2y	5y	7y	10y	3m	6m	1y	2y	5y	7y	10y
Mean	0.02	0.02	0.02	0.02	0.01	0.01	0.01	-0.01	-0.02	-0.03	-0.04	-0.01	0.02	0.06
Std	0.08	0.08	0.07	0.06	0.04	0.04	0.03	0.03	0.08	0.15	0.27	0.49	0.54	0.55
Min	-0.12	-0.11	-0.11	-0.11	-0.18	-0.27	-0.42	-0.11	-0.22	-0.39	-0.69	-1.85	-2.73	-4.29
Max	0.19	0.19	0.18	0.18	0.23	0.29	0.38	0.07	0.17	0.35	0.79	2.04	2.80	4.61

Table 8. Summary statistics of RFE and PFE and their estimation bias. Panel A reports summary statistics for RFE and PFE across maturities, along with p-values for the null hypothesis of zero errors. The reaction function error (RFE) is defined in Eq. (15) and the probability forecast error (PFE) is defined in Eq. (16). Panel B and Panel C report summary statistics for the distribution of spurious RFE and PFE estimates generated by small-sample estimation bias under the data-generating process of Case 1 ($RFE = PFE = 0$). For each of the $1,000$ simulated data sets, we estimate the unrestricted model, Case 4. Mean, Min and Max are calculated as statistics of sample averages of the spuriously estimated RFE and PFE. Std is calculated as the average, across 1,000 bootstrap samples, of time-series standard deviations of the bootstrap biases. The length of each simulated time series is $T = 445$ in Panel B and $T = 200$ in Panel C. The p-values in Panel A are computed from these bootstrap distributions. In particular, p-value (f.s) and p-value (s.s) denote, respectively, the p-values from the full sample and the small sample. Numbers are reported in percent.

and the probability forecast error is dominant on the long end of the term structure.

A well-known issue in the estimation of term structure models is that, due to the high persistence of interest rates, maximum likelihood estimates are prone to small-sample estimation bias (Kim and Orphanides 2012, Bauer et al. 2012). The implication is that, in small samples, the state variables can be spuriously estimated to be less persistent than they truly

are. In our framework, the presence of such small-sample estimation bias could distort the yield decomposition in Eq. (14).

In order to assess whether the estimated RFEs and PFEs are large relative to the spurious components that would be obtained if the true data-generating process had no survey biases, we simulate 1,000 data samples with the same properties as the original dataset under the null hypothesis of Case 1, thus ensuring that they are free of probability forecast errors and reaction function errors.[14] Notably, since under Case 1 the persistence of the physical and survey dynamics is the same, and since the survey dynamics are strongly identified, further small-sample bias correction, such as the one proposed in (Bauer et al. 2012), is unwarranted. For each simulated sample, we estimate the unrestricted model (Case 4) and compute the RFE and PFE as in Eqs. (15)–(16). Since the true data-generating process is free of PFE and RFE, any estimated values of these components reflect estimation bias and small-sample variation.

Panel B of Table 8 reports the resulting distribution of spurious RFEs and PFEs using 1,000 simulated samples of the same length as the original data, i.e. $T = 445$. The table indicates that the small-sample estimation biases are close to zero on average. Their dispersion is small for RFE, while PFE displays larger finite-sample variation, especially at long maturities. Importantly, the average PFE estimated in the actual data is substantially larger than the average bias generated under the no-bias data generating process. For example, the 10-year PFE is 1.18% in Panel A, while the mean full-sample bias is essentially zero and its standard deviation is 0.41% in Panel B. In addition, the second line of Panel A (p-value f.s.) reports the p-values for the null hypothesis of zero mean error obtained from this bootstrap distribution calculated as $p_n = \frac{1}{1{,}000} \sum_b \mathbf{1}\{|\bar{X}_{b,n} - c_n| \geq |\bar{X}_n^{data} - c_n|\}$, where $X \in \{RFE, PFE\}$ and c_n is the bootstrap mean for maturity n. In the full sample, results

[14] Full details of the simulation procedure are provided in Appendix E.

in the table indicate that while we cannot reject the null hypothesis of zero mean true reaction function errors for any maturity, the null of zero mean true probability forecast error is always rejected at the 10% level, which provides evidence that mean probability forecast errors are not driven by estimation bias.

Kim and Orphanides (2012) argue that incorporating surveys into term structure models can help eliminate small-sample bias in estimates of interest rate dynamics. Our results suggest that survey biases may prevent this benefit from being fully realized. Nonetheless, it is instructive to examine the impact of sample length on bias magnitude. To this end, we estimate the unrestricted model (Case 4) using only the first 200 observations (16 years and 8 months) of the simulated samples (generated under Case 1) and compute the distribution statistics. Summary statistics for the small-sample estimation bias using this short sample of 200 observations are reported in Panel C of Table 8. The results indicate that, in the shorter sample, the average small-sample estimation bias remains close to zero, but the dispersion increases substantially. This is especially visible for long-maturity PFEs, whose simulated average bias can reach several percentage points in absolute value, while the RFE bias remains smaller but is still more dispersed than in the full-sample exercise. The third line of Panel A (p-value s.s.) reports the p-values for the null hypothesis of zero mean survey error obtained from this small sample bootstrap. Results indicate that, in the short sample, we cannot reject the null that the average survey bias is equal to the small-sample estimation bias.

Figure 3 reports the time series estimates of the PFEs and the RFEs, together with the bootstrap reference bands for small-sample estimation bias. The bands, which represent 95% confidence intervals for the null hypothesis that the RFE and PFE are zero, are calculated as the time-series standard errors of the bootstrap biases reported in Panel B and Panel C of Table 8, centered around the average estimation bias. As such, they can be interpreted as

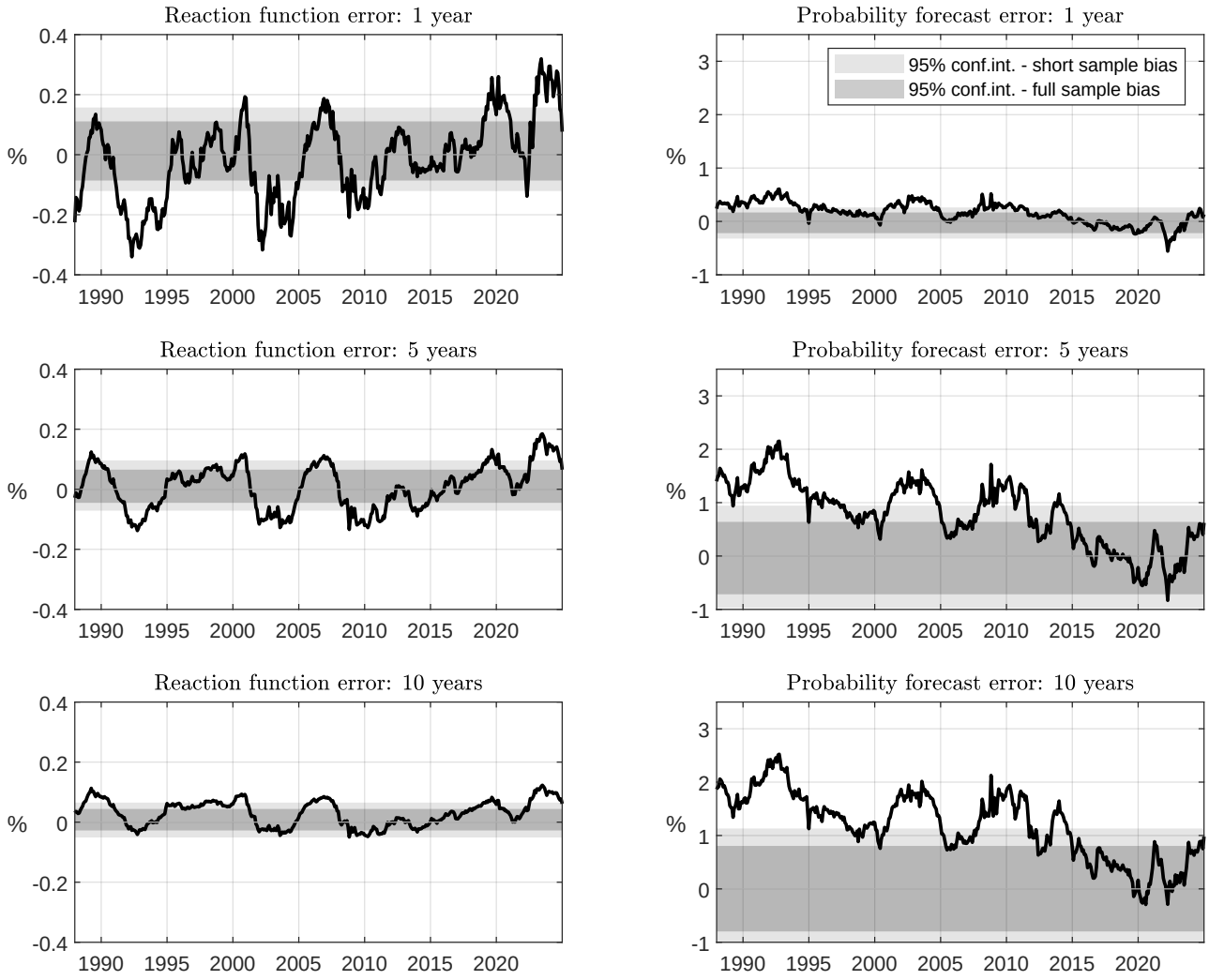


Figure 3. Survey forecast errors. The figure presents reaction function errors (left-hand column) and probability forecast errors (right-hand column) for maturities 1 year, 5 years and 10 years. The shaded areas are bootstrap reference bands for small-sample estimation bias under the data-generating process of Case 1 ($RFE = PFE = 0$). They are calculated as ± 1.96 times the bootstrap average standard deviation of the spurious RFE or PFE estimates around mean bias (see Table 8), using 1,000 simulations with the same size as the original sample ($T = 445$, dark shaded area) and with a shorter sample ($T = 200$, light shaded area).

bootstrap reference bands under the no-bias null for the realized RFE or PFE at each date.

Probability forecast errors for the 1-year yield range between -0.6% and $+0.6\%$ and remain close to zero during the latter part of the sample. For longer maturities, however, these errors are larger. For instance, for the 10-year yield, probability forecast errors oscillate between 0.7% and 2.5% until 2015, then decline to about zero by 2020, before rising again to around 1% in 2024. These long-maturity PFEs are frequently outside the full-sample bias band, indicating that they are too large to be explained by the small-sample bias generated under Case 1. Reaction function errors, by contrast, generally oscillate around zero but exhibit greater magnitude and significance for shorter maturities. For the 1-year yield, they fluctuate between -0.35% and $+0.35\%$, whereas for the 10-year yield they remain approximately within $\pm 0.1\%$ and are often within the small-sample bias bands throughout most of the sample period. This supports the interpretation that RFEs are economically relevant mainly at the short end, while PFEs dominate the long-maturity wedge between subjective and physical expectations. To further illustrate this, Figure 3 includes the short-sample bias reference band as a light-shaded area. These results demonstrate that a long sample is important, not merely to enhance the precision of interest rate dynamics estimates beyond survey-based inference, but also to improve the identification and quantification of biases embedded in survey expectations.

5.5 Subjective and physical term premia

We now examine the model's implications for the decomposition of long-term interest rates into expected future rates and risk premium, with reference to the two decompositions given in Eq. (13) and Eq. (14).

Figure 4 reports the 10-year yield along with the estimated expectation component under

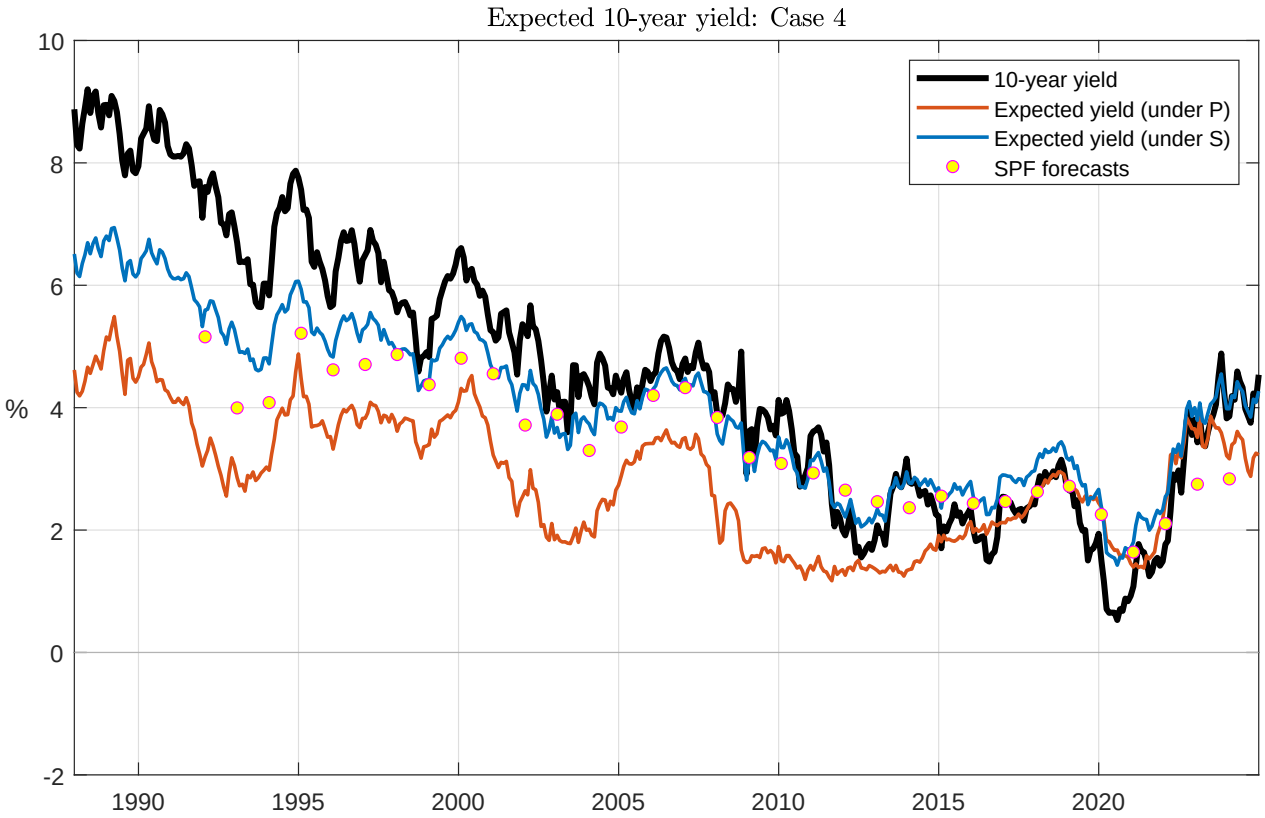


Figure 4. Expectation component. 10-year yield along with the expectation component under the physical measure (first term in (14)) and under the survey measure (first term in (13)). For reference, we also plot the average expected three-month T-bill rate over the next 10 years from Philadelphia Fed's Survey of Professional Forecasters.

the physical measure (first term in Eq. (14)) and under the survey measure (first term in Eq. (13)). The figure shows that the estimated expectation component under the survey measure is (almost) always higher than under the physical measure. The gap between the two narrows over the sample period, almost closing around 2022, before widening again towards the end of the sample.

Notably, the two decompositions attribute the secular decline in long-term interest rates to different components. Over our sample, the 10-year yield declined by about 4.6 percentage points. Under the physical measure, less than a third of this decline (29%) is attributed to the expectation component, with the remaining 71% absorbed by the term premium. Under

the survey measure, instead, the expectation component accounts for about half of the decline (52%).[15] This distinction is economically meaningful: Bauer and Rudebusch (2020) argue that the downward trend in long-term interest rates is driven by shifting trends in the equilibrium real rate and expected inflation, that is, by expectations rather than by risk compensation. In this respect, the decomposition based on survey expectations is more consistent with the macroeconomic evidence on the drivers of the secular decline in interest rates.

As indicated in Eq. (14), this gap between the expectation component under the physical and the survey measure implies a gap between the survey and the physical term premium. Indeed, in Figure 5, we can see that the survey term premium is lower than implied from historical bond prices. Similarly to the expectations components, the gap starts at about 2% and decreases over the sample, but it widens again to 1% in 2024. This discrepancy is economically large and implies that term premia inferred solely from bond prices would overstate the compensation investors require based on their subjective expectations. In particular, while the survey term premium turns negative in 2012 — a pattern also observed in the reduced-form SPF-based term premium — the physical term premium remains positive until 2016. The figure also indicates that the divergence between the survey and the physical term premium is mainly driven by the probability forecast error, as the contribution of reaction function errors at the long end of the yield curve is negligible.

This divergence between the survey and the physical term premium can have implications for the assessment of the effects of monetary policy measures. To illustrate these differences, consider the 1.16% decline in the 10-year yield in November 2008, when the first round

[15] These figures compare calendar-year averages of the first and last full years of the sample, which smooth out short-run fluctuations in the endpoints. The decomposition is not sensitive to this choice. Comparing instead the first and last monthly observations, the 10-year yield declines by 4.34 percentage points, of which the expectation component accounts for 32% and 52% under physical and survey measures, respectively.

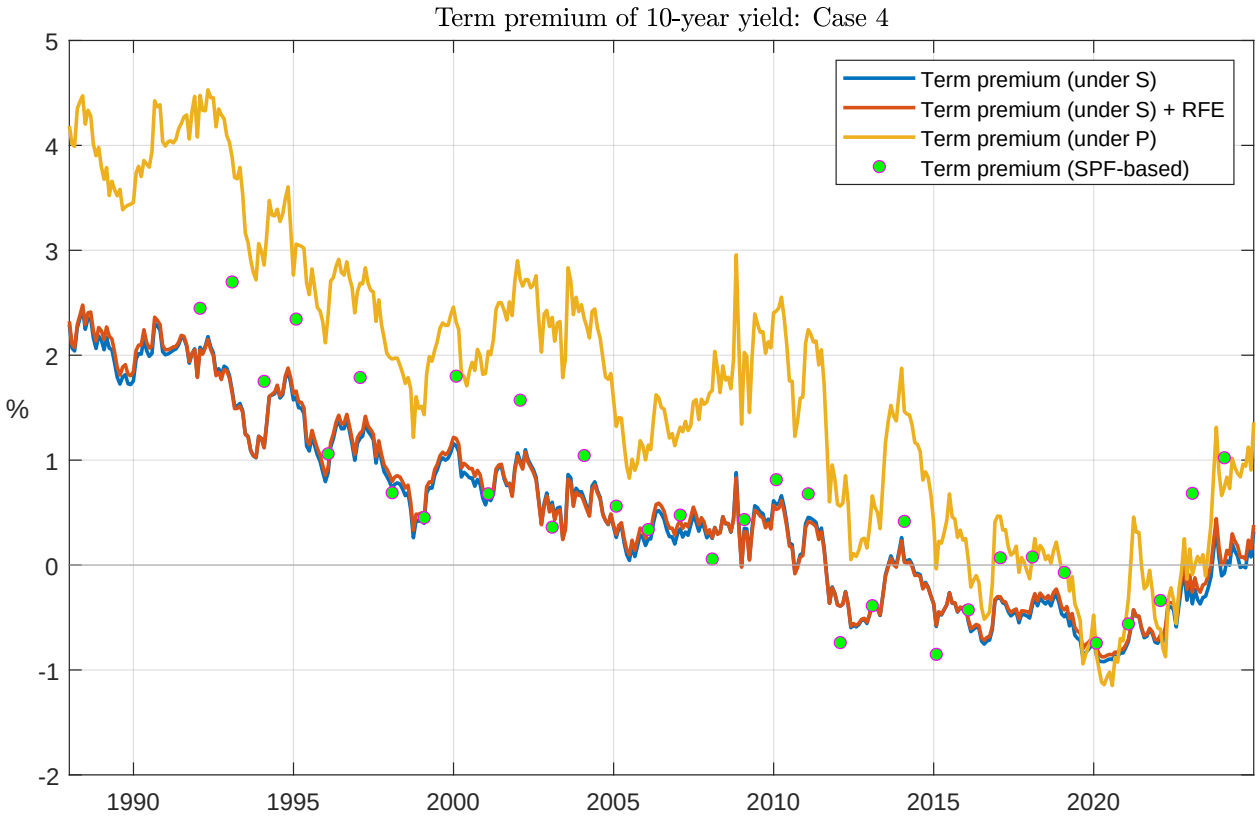


Figure 5. Term premium component. The figure shows the 10-year survey term premium and the 10-year physical term premium along with the decomposition in (17). For reference, we also plot the survey term premium computed as the difference between the 10-year yield and the average expected T-bill rate over the next 10 years from Philadelphia Fed's Survey of Professional Forecasters.

of quantitative easing was announced in response to the great financial crisis. The term premium accounts for 74% of this change under the physical measure, but only for 42% under the subjective measure.[16] Similarly, if we look at the one-month change in March 2020 (-0.46%), when the Fed announced the restart of quantitative easing measures, the return to near-zero rates and extensive new measures in the face of the COVID-19 pandemic, the term premium accounts for 29% of this change under the physical measure, while it accounts for only 1% under the subjective measure. Hence, these two episodes illustrate that the interpretation of yield movements around monetary policy announcements can

[16]These figures are not reported in any table. The corresponding time series are available upon request.

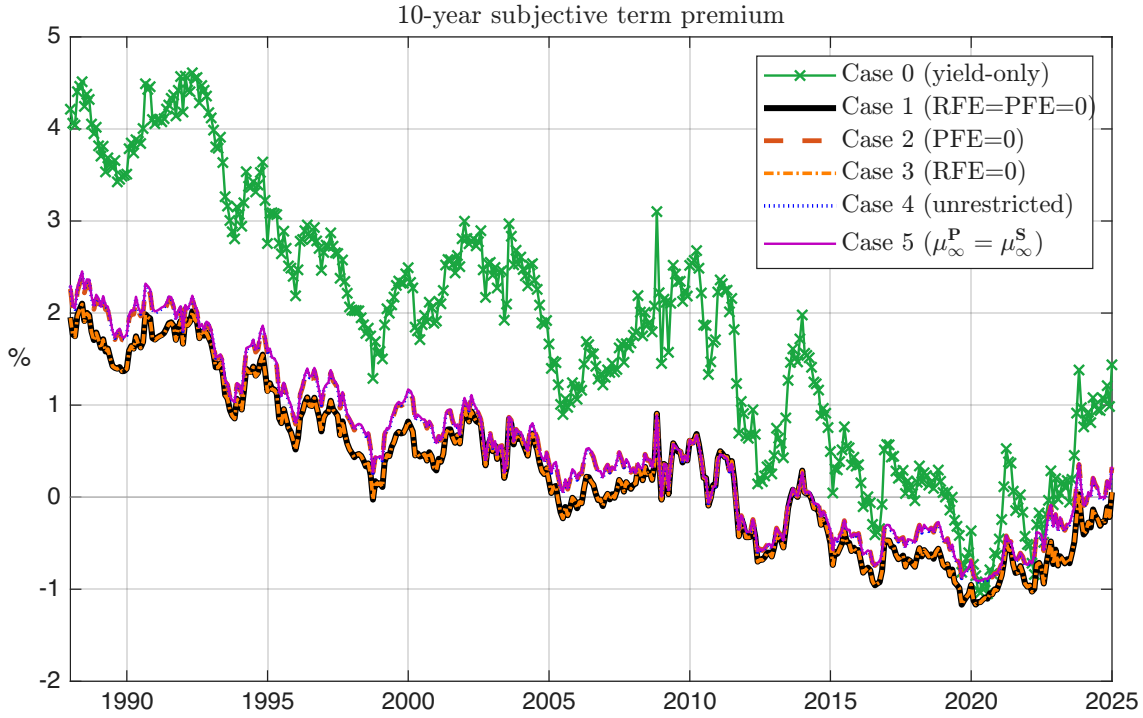


Figure 6. 10-year subjective term premium. The specifications for Case 1 to Case 5 are detailed in Table 1. For Case 0, we report the physical term premium.

differ according to whether it is assessed with beliefs of market (survey) participants or with historical interest rates dynamics.

Given the large discrepancy between the survey and the physical term premium, we now assess the implications of imposing restrictions on term structure models for interest rates and surveys. In Figure 6, we report estimates for the 10-year subjective term premium for the different model specification cases in Table 1 plus Case 0, which only uses interest rates. The figure indicates that the survey term premium is always well identified, as long as surveys are included in the model. However, the estimates for the physical term premium in Figure 7 reveal that model specification matters. Models that allow for probability forecast errors (Cases 3 and 4) recover the physical term premium of the yield-only model (Case 0).

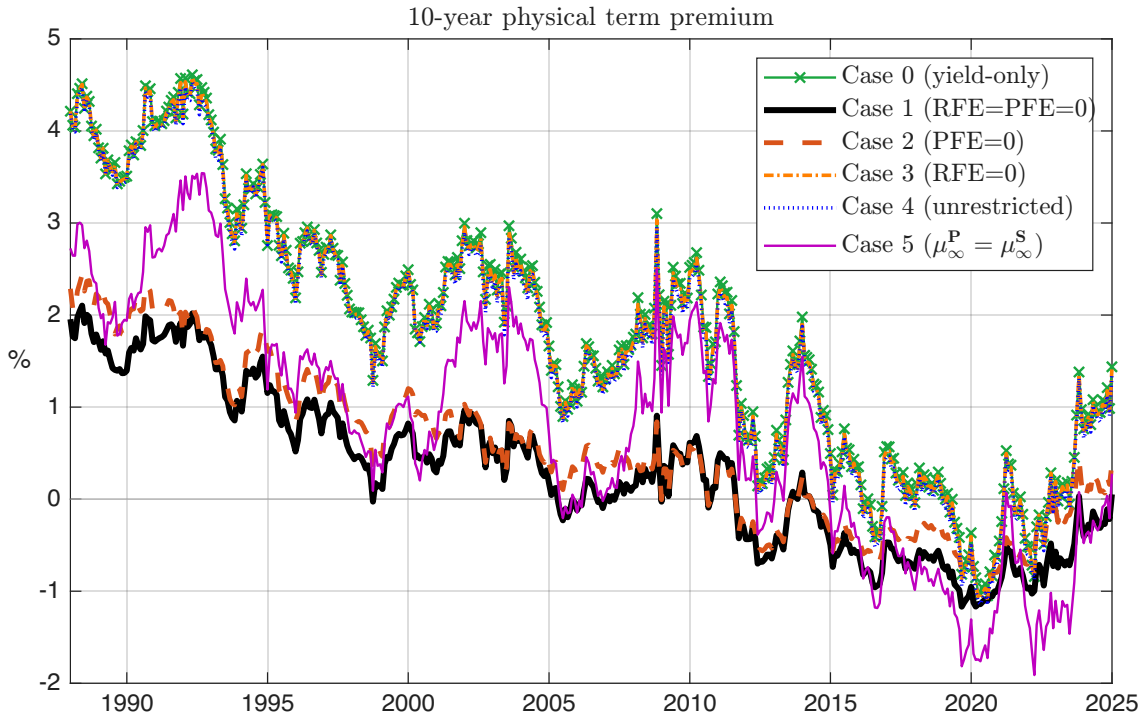


Figure 7. 10-year physical term premium. The specifications for Case 1 to Case 5 are detailed in Table 1.

Instead, the estimates for Cases 1 and 2, which impose that the probability forecast error is zero, are biased. Case 5, which allows for probability forecast errors but ties the long-run mean of the short rate under the physical measure to its survey counterpart, lies in between: its physical term premium averages 0.8% over the sample, compared with 1.8% for Case 0 and between 0.3% and 0.5% for Cases 1 and 2, and it falls below all other estimates after 2016, reaching −1.9% in 2022. This reflects the higher long-run mean imposed under the physical measure (see Table 7), which raises the expectation component of the 10-year yield, and hence lowers the term premium, most when interest rates are low. Comparing with Figure 6, we can see that excluding probability forecast errors effectively imposes equality between the estimated physical and survey term premium. This might be desirable if we

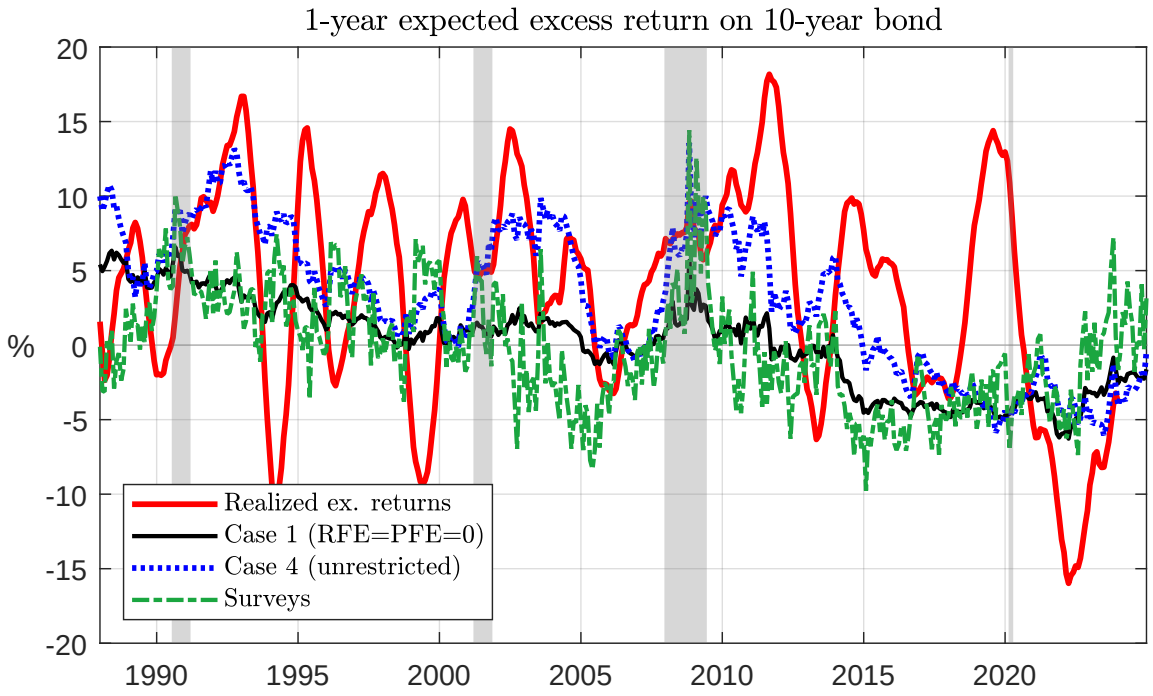


Figure 8. Expected excess returns comparison. The figure shows 1-year expected excess returns on the 10-year bond implied by surveys and the model (Case 1 and Case 4) under the physical probability measure. We also plot realized 1-year excess returns on a 10-year bond smoothed over 12 months, shifted forward to align with the timing of expectations. Shaded areas indicate NBER recessions.

want to estimate the market term premium, as the survey term premium represents the compensation required by investors based on their expectations of future interest rates, regardless of whether they make systematic forecast errors.

5.6 Forecasting realised returns and rates

We now turn to the ability of the different model specifications to predict realized returns and rates. Figure 8 plots the time series of expected excess returns on a long-maturity bond, as measured by surveys and implied by the model. Given that the expected yield on a

Model	2y	5y	10y
Case 0 (yield-only)	0.164	0.162	0.206
Case 1 (RFE=PFE=0)	0.132	0.118	0.130
Case 2 (PFE=0)	0.100	0.078	0.102
Case 3 (RFE=0)	0.163	0.161	0.205
Case 4 (unrestricted)	0.162	0.160	0.205
Case 5 ($\mu_\infty^{\mathbb{P}} = \mu_\infty^{\mathbb{S}}$)	0.134	0.148	0.201
Surveys	0.045	0.051	0.074

Table 9. Adjusted R^2 for predictive regressions of one year holding period excess returns on model expected excess returns. The specifications for Case 1 to Case 5 are detailed in Table 1.

10-year bond is directly available from survey data, expected excess returns over a one-year holding period can be computed as: $-10y^s_{10\times12,t,12} + 11y_{11\times12,t} - y_{12,t}$. With a slight abuse of terminology, we refer to these excess returns as excess returns on a 10-year bond, noting that the bond has a remaining maturity of 10 years at the end of the holding period. The figure also reports the corresponding expected returns under the physical probability measure, as well as a 12-month moving average of realized excess returns on the same bond, shifted one year forward to align with the timing of expectations. The expected excess returns generated by Case 1 closely track the survey-implied measures, despite the latter exhibiting substantial volatility, but both series show little correlation with realized returns. In contrast, the expected excess returns implied by Case 4 display markedly different dynamics — particularly in the early 1990s and 2000s — and align more closely with realized returns.

In Table 9, we report the adjusted R^2 for predictive regressions of one year holding period excess returns for bonds with maturities 2, 5 and 10 years on model expected excess returns for Case 0 (yield-only), 1 to 5 (from Table 1) and surveys. The table indicates that surveys have the lowest R^2 followed by expectations from term structure models for yields and surveys that impose zero probability forecast errors. The yield-only and the term structure models that allow for non-zero probability forecast errors (Cases 3 and 4) all have

the highest R^2, indicating that it is crucial to allow for survey biases to correctly capture the physical dynamics. Case 5 lies in between: its R^2 is close to those of Cases 0, 3 and 4 for the 10-year bond, but similar to those of the models without probability forecast errors for the 2-year bond.

To corroborate further these findings, in Table 10 we report the root mean-square forecasting error (RMSFE) for forecasts made with physical dynamics of different model specifications. For all periods from January 1988 to December 2024, we forecast 3 month, 6 month and 1, 2, 5 and 10-year yields for 1 through 4 quarters ahead. We highlight the smallest RMSFE for each yield and at each forecasting horizon in bold font. Results indicate that the yield-only model and the models with non-zero probability forecast errors (Cases 3 and 4) deliver the most accurate in-sample forecasts, while Case 5, which ties the long-run mean to surveys, lies in between, with a gap that widens with the forecast horizon. This evidence further corroborates that it is important to distinguish between survey and physical dynamics, by explicitly incorporating survey biases into joint term structure models for interest rates and surveys.

As final evidence, we perform an out-of-sample forecasting exercise. We use a rolling scheme, as follows. Using the initial sample from 1987:12 to 2009:01, we estimate the model and make forecasts (under the $\mathbb{P}$ measure) for 3, 6, 9 and 12 months ahead. Next, we remove the data for 1987:12 and add the data for 2009:02, re-estimate the model and make predictions for the four forecast horizons again. We repeat this procedure until 2023:12, so that we obtain 180 predictions for each forecast horizon.

Table 11 combines the scale of the forecasting errors with the relative comparison across specifications. Throughout this exercise, the benchmark is Case 0, the yield-only model, because it estimates the physical dynamics from yields alone and therefore provides the natural reference for assessing whether adding survey restrictions improves forecasts of realized

Model	Horizon\Yield	3m	6m	1y	2y	5y	10y	Av.
Case 0 (yield-only)	1q	**41**	**43**	**48**	**53**	**51**	**45**	**47**
	2q	**69**	**72**	**76**	**77**	**71**	**61**	**71**
	3q	**93**	**96**	**99**	**96**	**83**	**71**	**90**
	4q	**118**	**121**	**122**	**117**	**97**	**80**	**109**
Surveys	1q	46	51	57	58	58	52	53
	2q	75	81	86	84	78	70	79
	3q	106	111	116	109	95	83	103
	4q	137	142	145	135	113	96	128
Case 1 (RFE=PFE=0)	1q	45	47	51	55	53	47	50
	2q	77	79	83	84	76	65	77
	3q	106	108	110	107	92	78	100
	4q	134	136	136	131	109	90	123
Case 2 (PFE=0)	1q	45	47	52	56	53	47	50
	2q	78	80	84	84	76	65	78
	3q	107	109	111	108	92	78	101
	4q	135	137	137	131	109	90	123
Case 3 (RFE=0)	1q	**41**	44	**48**	**53**	**51**	**45**	**47**
	2q	**69**	**72**	**76**	**77**	**71**	**61**	**71**
	3q	94	**96**	**99**	**96**	**83**	**71**	**90**
	4q	**118**	**121**	123	**117**	**97**	**80**	**109**
Case 4 (unrestricted)	1q	**41**	44	**48**	**53**	**51**	**45**	**47**
	2q	**69**	**72**	**76**	**77**	**71**	**61**	**71**
	3q	94	97	**99**	97	84	**71**	**90**
	4q	119	**121**	123	**117**	**97**	**80**	**109**
Case 5 ($\mu_\infty^{\mathbb{P}} = \mu_\infty^{\mathbb{S}}$)	1q	42	44	49	**53**	**51**	**45**	48
	2q	72	74	78	79	72	62	73
	3q	99	101	103	100	85	73	94
	4q	126	128	129	123	100	82	115

Table 10. In-sample $RMSFE$s for the model-implied objective expectations for the period from 1987:12 to 2024:12. The smallest $RMSFE$s (after rounding to the nearest integer) for each horizon/maturity pair are denoted in bold font. The last column reports the average across maturities.

rates. Panel A reports the out-of-sample $RMSFE$s of Case 0 in basis points, while Panel B reports the ratio of the $RMSFE$ of each competing forecast to the $RMSFE$ of Case 0, the yield-only benchmark. A value larger than 1 means that the forecasts generated by the

Model	Horizon\Yield	3m	6m	1y	2y	5y	10y
Panel A: RMSFE (basis points)							
Case 0 (yield-only)	1q	29	31	35	37	45	48
	2q	52	55	58	60	65	67
	3q	74	78	81	81	82	82
	4q	98	102	105	103	99	94
Panel B: RMSFE relative to Case 0							
Surveys	1q	0.92	0.98	1.05	1.11**	1.08	1.08
	2q	0.95	0.98	1.04	1.10**	1.09	1.06
	3q	1.02	1.05	1.09*	1.12**	1.10	1.08
	4q	1.07	1.08	1.10**	1.13**	1.12	1.12
Case 1 (RFE=PFE=0)	1q	1.26	1.26	1.22	1.19**	1.03	1.04
	2q	1.28	1.27	1.26	1.25**	1.13**	1.08
	3q	1.25	1.24	1.23**	1.23**	1.15*	1.12
	4q	1.21	1.20*	1.19**	1.20**	1.16	1.17*
Case 2 (PFE=0)	1q	1.34	1.33	1.28	1.21*	1.04	1.04
	2q	1.34	1.33	1.30	1.27**	1.14**	1.09
	3q	1.28	1.27	1.25	1.23**	1.16*	1.13*
	4q	1.23	1.21	1.20*	1.20**	1.16	1.17*
Case 3 (RFE=0)	1q	1.04*	1.04	1.02	1.01	0.97	0.99
	2q	1.03*	1.03	1.01	1.01	1.00	0.99
	3q	1.03*	1.02	1.02	1.01	1.00	1.00
	4q	1.03	1.02	1.01	1.01	0.99	1.01
Case 4 (unrestricted)	1q	1.07*	1.05	1.02	1.03	0.99	1.00
	2q	1.05	1.04	1.03	1.04	1.03	1.01
	3q	1.05	1.04	1.04	1.05	1.05	1.04
	4q	1.04	1.04	1.04	1.05	1.04	1.06
Case 5 ($\mu_\infty^{\mathbb{P}} = \mu_\infty^{\mathbb{S}}$)	1q	1.12**	1.08	1.04	1.04	0.99	1.01
	2q	1.11	1.08	1.06	1.07	1.04	1.03
	3q	1.10	1.08	1.07	1.07	1.05	1.05
	4q	1.09	1.08	1.07	1.07	1.05	1.07

Table 11. Out-of-sample forecasting performance. Panel A reports Case 0 RMSFEs in basis points. Panel B reports RMSFE ratios relative to Case 0, so values above one indicate lower forecast accuracy than the yield-only benchmark. The forecast origins are from 2009:01 to 2023:12. ** and * denote two-sided significance at the 5% and 10% level, respectively, against Case 0 using the Diebold and Mariano (1995) test of equal predictive ability with fixed-b asymptotics as in Coroneo and Iacone (2020).

considered model are less accurate than those of Case 0, while a value smaller than 1 means that they are more accurate.

To assess whether the predictive performance of each competing specification differs statistically from Case 0, we apply the Diebold and Mariano (1995) test of equal predictive ability using fixed-b asymptotics as in Coroneo and Iacone (2020), which allows us to obtain correctly sized tests with our sample size (180 out-of-sample observations) and in the presence of autocorrelated forecast errors.[17] The null hypothesis is that the benchmark model (Case 0) and the competing specification exhibit equal forecast accuracy, while the alternative is that their predictive accuracies differ.[18] As shown in Table 11, the null hypothesis is frequently rejected for Cases 1 and 2, which impose zero probability forecast errors. In contrast, Cases 3, 4 and 5 remain close to Case 0, with only isolated rejections, mostly at the 10% level. The survey forecasts are also generally less accurate than Case 0 at longer maturities. These results provide further evidence that forecasts of realized yields and excess returns are primarily disciplined by the physical dynamics estimated from yields, while imposing survey-based restrictions on those dynamics worsens out-of-sample predictive performance.

6 Conclusion

A popular approach to deal with the poor identification of the parameters that determine the dynamics of interest rates, and thus the risk premium, is to augment term structure

[17] Although interest rates and forecast errors are highly persistent, the relevant object for the Diebold and Mariano (1995) test is the loss differential. Unreported sample autocorrelations of the loss differentials decline substantially beyond the first lag, suggesting that they do not display the degree of dependence under which the Diebold and Mariano (1995) test may suffer from severe power distortions; see Coroneo and Iacone (2025).

[18] As in Coroneo and Iacone (2020), we use a quadratic loss and a weighted covariance estimate for the long-run variance with Bartlett kernel and truncation $\lfloor T^{1/2} \rfloor$ (which equals 13 for our out-of-sample size of 180 observations).

models of interest rates with expectations obtained from interest rate surveys. It has been assumed offhand that the resulting decomposition can be taken as *both* the risk premium demanded by investors for holding long-maturity Treasury securities *and* the expectation of the short interest rate with respect to the real-world probability measure. Our results demonstrate that this interpretation holds only in the absence of survey biases, a condition that is rejected in the data most of the time.

Specifically, this paper develops a joint term structure model for zero-coupon yields and survey expectations that allows for subjective and objective dynamics and reaction functions to differ. By relaxing the common equal reaction and equal dynamics assumptions, we are able to test whether survey expectations align with the dynamics implied by bond prices. We find that survey participants make two types of systematic errors: misperceptions of the policy reaction function at short maturities and biased probability forecasts at long maturities. These errors imply that long-term interest rates admit two distinct decompositions: one under subjective beliefs, combining survey expectations with survey-implied term premia, and another under historical dynamics, combining objective expectations with physical term premia. Moreover, even the weaker restriction that the long-run mean of the short rate is the same under the two measures is not supported by the data: although it cannot be rejected by a likelihood ratio test, it implies counterfactually high long-run means of yields and worsens forecasts.

Our key result is that survey-based interest rate dynamics are markedly different from those inferred from bond prices, leading to a systematically lower subjective term premium. Given these findings and the broader evidence of biases in survey expectations, one might naturally question whether survey data should be used at all to augment model estimation and interest rate decompositions. In our view, the answer depends on the objective under consideration. If the objective is to forecast interest rates or expected returns, then surveys

are not essential, and a yield-only model is sufficient or, indeed, preferred since using surveys without accounting for biases risks distorting the estimated dynamics of interest rates. If the goal, however, is to measure the premium that investors (or survey participants) require for holding long-maturity bonds, then survey information becomes indispensable, as such a premium cannot be reliably inferred from yields alone. These two questions have often been treated as if they were identical, but they need not—and should not—be.

References

Adam, Klaus, Albert Marcet, and Johannes Beutel (2017) 'Stock price booms and expected capital gains.' *American Economic Review* 107(8), 2352–2408

Adam, Klaus, and Albert Marcet (2011) 'Internal rationality, imperfect market knowledge and asset prices.' *Journal of Economic Theory* 146(3), 1224–1252

Ang, Andrew, and Monika Piazzesi (2003) 'A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables.' *Journal of Monetary Economics* 50(4), 745–787

Bauer, Michael D., and Glenn D. Rudebusch (2020) 'Interest rates under falling stars.' *American Economic Review* 110(5), 1316–54

Bauer, Michael D., Carolin E. Pflueger, and Adi Sunderam (2024) 'Perceptions about monetary policy.' *The Quarterly Journal of Economics* 139(4), 2227–2278

Bauer, Michael D., Glenn D. Rudebusch, and Jing C. Wu (2012) 'Correcting estimation bias in dynamic term structure models.' *Journal of Business & Economic Statistics* 30(3), 454–467

Bianchi, Francesco, Sydney C. Ludvigson, and Sai Ma (2022) 'Belief distortions and macroeconomic fluctuations.' *American Economic Review* 112(7), 2269–2315

Black, Fischer (1995) 'Interest rates as options.' *The Journal of Finance* 50(5), 1371–1376

Buraschi, Andrea, Ilaria Piatti, and Paul Whelan (2022) 'Subjective bond returns and belief aggregation.' *The Review of Financial Studies* 35(8), 3710–3741

Chernov, Mikhail, and Philippe Mueller (2012) 'The term structure of inflation expectations.' *Journal of Financial Economics* 106(2), 367–394

Chun, Albert L. (2011) 'Expectations, bond yields, and monetary policy.' *The Review of Financial Studies* 24(1), 208–247

Cieslak, Anna (2018) 'Short-rate expectations and unexpected returns in Treasury bonds.' *The Review of Financial Studies* 31(9), 3265–3306

Coibion, Olivier, and Yuriy Gorodnichenko (2015) 'Information rigidity and the expectations formation process: A simple framework and new facts.' *American Economic Review* 105(8), 2644–78

Connolly, Michael F, and Ethan Struby (2026) 'Shadow rate models and monetary policy.' *Journal of Financial Econometrics* 24(4), nbag015

Coroneo, Laura, and Fabrizio Iacone (2020) 'Comparing predictive accuracy in small samples using fixed-smoothing asymptotics.' *Journal of Applied Econometrics* 35(4), 391–409

—— (2025) 'Testing for equal predictive accuracy with strong dependence.' *International Journal of Forecasting* 41(3), 1073–1092

Crump, Richard K., Stefano Eusepi, and Emanuel Moench (2018) 'The term structure of expectations and bond yields.' Staff Reports 775, Federal Reserve Bank of New York

De la O, Ricardo, and Sean Myers (2021) 'Subjective cash flow and discount rate expectations.' *The Journal of Finance* 76(3), 1339–1387

Diebold, Francis X., and Roberto S. Mariano (1995) 'Comparing predictive accuracy.' *Journal of Business & Economic Statistics* 13(3), 253–263

Duffee, Gregory R. (2011) 'Information in (and not in) the term structure.' *The Review of Financial Studies* 24(9), 2895–2934

Duffee, Gregory R, and Richard H Stanton (2012) 'Estimation of dynamic term structure models.' *The Quarterly Journal of Finance* 2(02), 1250008

d'Amico, Stefania, Don H. Kim, and Min Wei (2018) 'Tips from TIPS: the informational content of Treasury Inflation-Protected Security prices.' *Journal of Financial and Quantitative Analysis* 53(1), 395–436

Farmer, Leland E, Emi Nakamura, and Jón Steinsson (2024) 'Learning about the long run.' *Journal of Political Economy* 132(10), 3334–3377

Golinski, Adam, and Peter Spencer (2024) 'Unconventional monetary policies and the yield curve: Estimating non-affine term structure models with unspanned macro risk by factor extraction.' *The Review of Asset Pricing Studies* 14(1), 119–152

Greene, William H. (2011) *Econometric Analysis* (7th Edition, Pearson Education)

Greenwood, Robin, and Andrei Shleifer (2014) 'Expectations of returns and expected returns.' *The Review of Financial Studies* 27(3), 714–746

Hamilton, James D., and Jing C. Wu (2012) 'Identification and estimation of Gaussian affine term structure models.' *Journal of Econometrics* 168(2), 315–331

Joslin, Scott, Kenneth J. Singleton, and Haoxiang Zhu (2011) 'A new perspective on Gaussian dynamic term structure models.' *The Review of Financial Studies* 24(3), 926–970

Kim, Don H., and Athanasios Orphanides (2012) 'Term structure estimation with survey data on interest rate forecasts.' *Journal of Financial and Quantitative Analysis* 47(1), 241–272

Kim, Don H., and Jonathan H. Wright (2005) 'An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates.' Finance and Economics Discussion Series 2005-33, Board of Governors of the Federal Reserve System (U.S.)

Kim, Don H., and Marcel A. Priebsch (2020) 'Are shadow rate models of the Treasury yield curve structurally stable?' Finance and Economics Discussion Series 2020-061, Board of Governors of the Federal Reserve System

Liu, Yan, and Jing C. Wu (2021) 'Reconstructing the yield curve.' *Journal of Financial Economics* 142(3), 1395–1425

Mankiw, N. Gregory, Ricardo Reis, and Justin Wolfers (2003) 'Disagreement about inflation expectations.' *NBER Macroeconomics Annual* 18, 209–248

Nagel, Stefan, and Zhengyang Xu (2023) 'Dynamics of subjective risk premia.' *Journal of Financial Economics* 150(2), 103713

Nelson, Charles R., and Andrew F. Siegel (1987) 'Parsimonious modeling of yield curves.' *The Journal of Business* 60(4), 473–489

Piazzesi, Monika, Juliana Salomao, and Martin Schneider (2015) 'Trend and cycle in bond premia.' Working paper

Politis, Dimitris N., and Joseph P. Romano (1994) 'The stationary bootstrap.' *Journal of the American Statistical Association* 89(428), 1303–1313

Priebsch, Marcel A (2023) 'Computing arbitrage-free yields in multi-factor Gaussian shadow rate term structure models.' *The Quarterly Journal of Finance* 13(04), 2350013

Sastry, Karthik A. (2026) 'Disagreement about monetary policy.' *American Economic Journal: Macroeconomics* 18(2), 486–522

Schmeling, Maik, Andreas Schrimpf, and Sigurd A.M. Steffensen (2022) 'Monetary policy expectation errors.' *Journal of Financial Economics* 146(3), 841–858

Singleton, Kenneth J. (2021) 'Presidential address: How much "rationality" is there in bond-market risk premiums?' *The Journal of Finance* 76(4), 1611–1654

Svensson, Lars E. O. (1994) 'Estimating and interpreting forward interest rates: Sweden 1992-1994.' IMF Working Papers 1994/114, International Monetary Fund

Veronesi, Pietro (2010) *Fixed income securities: Valuation, risk, and risk management* (John Wiley & Sons)

Wu, Jing Cynthia, and Fan Dora Xia (2016) 'Measuring the macroeconomic impact of monetary policy at the zero lower bound.' *Journal of Money, Credit and Banking* 48(2-3), 253–291

Zellner, Arnold (1962) 'An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias.' *Journal of the American Statistical Association* 57(298), 348–368

Zeng, Ming, and Guihai Zhao (2026) 'Expectation-driven term structure of equity and bond yields.' *Journal of Monetary Economics* 157, 103881

Appendix A Survey factors

Unspanned risk factors that are not accounted for in the physical dynamics of the state vector could drive a wedge between the survey dynamics and the physical dynamics (see Duffee 2011). Any such unspanned risk factors should then be identifiable from both the physical law of motion of the state vector and bond excess returns. To this end, we estimate a $VAR(1)$ with the first three principal components extracted from yields (YPC) and the survey-specific factor ($SPC_{\perp}$). The survey-specific factor is computed from survey observations orthogonalized with respect to the first three principal components of yields.

$VAR(1)$

	$YPC_{1,t-1}$	$YPC_{2,t-1}$	$YPC_{3,t-1}$	$SPC_{\perp,t-1}$
$YPC_{1,t}$	0.993^{***}	-0.013	-0.425^{***}	-0.009
$YPC_{2,t}$	-0.002	0.974^{***}	0.261^{***}	-0.012
$YPC_{3,t}$	0.002^{**}	-0.001	0.881^{***}	0.001
$SPC_{\perp,t}$	0.007	0.099^{***}	-0.843^{***}	0.721^{***}

Excess returns

	$YPC_{1,t-1}$	$YPC_{2,t-1}$	$YPC_{3,t-1}$	$SPC_{\perp,t-1}$
$xret_{1,t}$	0.004^{***}	0.019^{***}	0.090^{***}	-0.001
$xret_{2,t}$	0.008^{**}	0.060^{***}	0.116	0.003
$xret_{5,t}$	0.013	0.155^{***}	0.193	0.048
$xret_{7,t}$	0.017	0.218^{***}	0.306	0.069
$xret_{10,t}$	0.019	0.304^{***}	0.500	0.075

Table A-1. Estimates of VAR(1) coefficients (top panel) and coefficients of predictive regression of annualised one-month excess returns on yield curve factors and the survey factor (bottom panel). ***, ** and * denote significance at the 1%, 5% and 10% level.

In the top panel of Table A-1, we report the coefficient estimates, which indicate that the survey factor does not Granger-cause the principal components extracted from yields at any conventional significance level. In the bottom panel of Table A-1, we report predictive regressions of monthly excess returns for 1, 2, 5, 7 and 10-year zero coupon bonds on the

yield curve factors (YPC) and the survey-specific factor ($SPC_{\perp}$). In line with the literature, we find that bond excess returns are predictable by the second YPC, usually referred to as the slope of the yield curve. We also find some evidence of predictability for the first yield principal component for the 1 and 2-year bond excess returns, and for the third principal component for the 1-year excess returns. [19] Importantly, however, the results indicate that the survey specific factor does not have any forecasting power in the excess return regressions.

Thus, to summarize, these results do not support the hypothesis that an unaccounted unspanned risk factor drives the wedge between the survey and the physical dynamics of the state vector.

[19] We also checked that the results in Table A-1 are virtually unchanged if we include the fourth and the fifth YPCs in the regressions, or if we use as survey-specific factor the first principal component from surveys orthogonalized with respect to the first five YPCs.

Appendix B Shadow rate survey expectations

From Eq. (10), the perceived one-period forward rate at horizon h is

$$f^{\mathbb{S}}(\mathbf{x}_{t+h}, n; \mathbf{\Psi}) \approx \underline{r} + \sigma_n^{\mathbb{S}} g\left(\frac{a_n^{\mathbb{S}} + \mathbf{b}_n^{\mathbb{S}\prime}\mathbf{x}_{t+h} - \underline{r}}{\sigma_n^{\mathbb{S}}}\right), \tag{B-1}$$

where $g(z) = z\mathtt{N}(z) + \mathtt{n}(z)$ with $\mathtt{N}(z)$ and $\mathtt{n}(z)$ denoting the standard normal CDF and the PDF of z.

Since $f^{\mathbb{S}}(\mathbf{x}_{t+h}, n; \mathbf{\Psi})$ is a nonlinear function of $\mathbf{x}_{t+h}$, its expectation is non-trivial. However, the function depends on $\mathbf{x}_{t+h}$ only through the scalar index

$$z_{n,t+h} = \frac{a_n^{\mathbb{S}} + \mathbf{b}_n^{\mathbb{S}\prime}\mathbf{x}_{t+h} - \underline{r}}{\sigma_n^{\mathbb{S}}}, \tag{B-2}$$

which is conditionally Gaussian under $\mathbb{S}$:

$$z_{n,t+h} \mid \mathcal{F}_t \overset{\mathbb{S}}{\sim} \mathcal{N}(\bar{z}_{n,t,h}, s_{n,h}^2). \tag{B-3}$$

where

$$\begin{aligned}
\bar{z}_{n,t,h} &= \frac{a_n^{\mathbb{S}} + \mathbf{b}_n^{\mathbb{S}\prime} E_t^{\mathbb{S}}[\mathbf{x}_{t+h}] - \underline{r}}{\sigma_n^{\mathbb{S}}}, \\
s_{n,h}^2 &= \frac{\mathbf{b}_n^{\mathbb{S}\prime}\mathbf{\Sigma}_h^{\mathbb{S}}\mathbf{b}_n^{\mathbb{S}}}{(\sigma_n^{\mathbb{S}})^2}, \\
\mathbf{\Sigma}_h^{\mathbb{S}} &= \sum_{j=0}^{h-1} (\mathbf{\Phi}^{\mathbb{S}})^j \mathbf{\Sigma} (\mathbf{\Phi}^{\mathbb{S}\prime})^j.
\end{aligned}$$

Using the fact that

$$g(x) = x\mathtt{N}(x) + \mathtt{n}(x) = E[\max\{x + u, 0\}], \qquad u \sim \mathcal{N}(0, 1), \tag{B-4}$$

we can evaluate the expectation of $g(z)$ when $z \sim \mathcal{N}(\bar{z}, s^2)$. Writing $z = \bar{z} + s\varepsilon$, with $\varepsilon \sim \mathcal{N}(0,1)$, gives

$$E[g(z)] \;=\; E[\max\{\bar{z} + s\varepsilon + u, 0\}]\,. \tag{B-5}$$

Since $s\varepsilon + u \sim \mathcal{N}(0, 1+s^2)$, the random variable

$$\bar{z} + s\varepsilon + u$$

is Gaussian with mean $\bar{z}$ and variance $1+s^2$. Therefore,

$$E[\max\{\bar{z} + s\varepsilon + u, 0\}] = \sqrt{1+s^2}\, g\left(\frac{\bar{z}}{\sqrt{1+s^2}}\right),$$

where we use the identity $E[\max\{Y,0\}] = v\, g(m/v)$ for $Y \sim \mathcal{N}(m, v^2)$. Hence,

$$E[g(z)] = \sqrt{1+s^2}\, g\left(\frac{\bar{z}}{\sqrt{1+s^2}}\right). \tag{B-6}$$

Therefore

$$E_t^{\mathbb{S}}[f^{\mathbb{S}}(\mathbf{x}_{t+h}, n; \mathbf{\Psi})] \approx \underline{r} + \sigma_n^{\mathbb{S}} \sqrt{1+s_{n,h}^2}\, g\left(\frac{\bar{z}_{n,t,h}}{\sqrt{1+s_{n,h}^2}}\right). \tag{B-7}$$

Equivalently,

$$E_t^{\mathbb{S}}[f^{\mathbb{S}}(\mathbf{x}_{t+h}, n; \mathbf{\Psi})] \approx \underline{r} + \sigma_n^{\mathbb{S}} \sqrt{1+s_{n,h}^2}\, g\left(\frac{a_n^{\mathbb{S}} + \mathbf{b}_n^{\mathbb{S}\prime} E_t^{\mathbb{S}}[\mathbf{x}_{t+h}] - \underline{r}}{\sigma_n^{\mathbb{S}} \sqrt{1+s_{n,h}^2}}\right). \tag{B-8}$$

Thus, the survey expectation has the same functional form as the original Wu and Xia (2016) approximation, but with adjusted arguments.

The special case is the expectation of the short rate

$$E_t^{\mathbb{S}}[r_{t+h}] = E_t^{\mathbb{S}}[f^{\mathbb{S}}(\mathbf{x}_{t+h}, 0; \mathbf{\Psi})]. \tag{B-9}$$

Under the perceived policy rule, the shadow short rate is $ssr_{t+h} = \delta_0^{\mathbb{S}} + \boldsymbol{\delta}_1^{\mathbb{S}\prime}\mathbf{x}_{t+h}$ which is conditionally Gaussian under $\mathbb{S}$. Therefore, the expectation of the short rate can be computed as:

$$\begin{aligned} E_t^{\mathbb{S}}[r_{t+h}] &= E_t^{\mathbb{S}}\left[\max\{\underline{r}, \delta_0^{\mathbb{S}} + \boldsymbol{\delta}_1^{\mathbb{S}\prime}\mathbf{x}_{t+h}\}\right] \\ &= \underline{r} + (\delta_0^{\mathbb{S}} + \boldsymbol{\delta}_1^{\mathbb{S}\prime}E_t^{\mathbb{S}}[\mathbf{x}_{t+h}] - \underline{r})\mathtt{N}\left(\frac{\delta_0^{\mathbb{S}} + \boldsymbol{\delta}_1^{\mathbb{S}\prime}E_t^{\mathbb{S}}[\mathbf{x}_{t+h}] - \underline{r}}{\sigma_h^{\mathbb{S}}}\right) \\ &\quad + \sigma_h^{\mathbb{S}}\mathtt{n}\left(\frac{\delta_0^{\mathbb{S}} + \boldsymbol{\delta}_1^{\mathbb{S}\prime}E_t^{\mathbb{S}}[\mathbf{x}_{t+h}] - \underline{r}}{\sigma_h^{\mathbb{S}}}\right), \end{aligned} \tag{B-10}$$

where

$$\left(\sigma_h^{\mathbb{S}}\right)^2 = \boldsymbol{\delta}_1^{\mathbb{S}\prime}\mathbf{\Sigma}_h^{\mathbb{S}}\boldsymbol{\delta}_1^{\mathbb{S}}. \tag{B-11}$$

Appendix C BCFF data

We use BCFF data for all available rates, which are generally different at each point in time.

Although the BCFF forecasts are published at the beginning of a month, they are collected in the last few days of the preceding month. Thus, to align the timing of the conditioning information set of the survey forecasts with the one in spot interest rates, we assume that surveys are observed as soon as they are collected (i.e. at the end of the month), so we treat the BCFF surveys published in the first few days of each month as observed at the end of the preceding month.

Since the surveys refer to T-bill discount rates and par yields, we extract zero-coupon rates using the Nelson and Siegel (1987) and Svensson (1994) model.[20] In particular, we parameterize the zero-coupon yield curve to compute the implied T-bill discount rates and par yields, which we then fit to the survey data. We fit the Svensson (1994) model that includes four shape parameters and two decay parameters. Throughout this paper, we use 3-month, 6-month, 1, 2, 5 and 10-year zero-coupon survey yields based on this procedure.

Finally, we note that the BCFF forecasts are for average rates over calendar quarters. To align the forecasts with spot interest rates, following the standard approach in the literature (e.g. Kim and Orphanides 2012, Buraschi, Piatti and Whelan 2022, Nagel and Xu 2023), we assume that the survey forecasts are for mid-quarter rates, and, at each point in time, we interpolate linearly between the quarterly horizons. We use forecasts for up to 5 quarters ahead, which, after interpolation, gives us forecasts for 3, 6, 9 and 12 months ahead.

[20] Although T-bills are quoted on the actual-to-360 days basis, for simplicity, we assume that each calendar quarter has 90 days. For a textbook treatment of quoting conventions and the Nelson and Siegel (1987) model, see Veronesi (2010).

Appendix D Estimation procedure

Factor extraction

The following factor extraction method is described in Golinski and Spencer (2024), which provides further details.

First, for convenience, we restate the notation. We denote the model-implied zero coupon yields at time t by $\mathbf{y}_t$ and the associated vector of pricing errors by $\mathbf{v}_t \sim N(\mathbf{0}, \mathbf{\Sigma}_y)$, such that the observed yields are

$$\widetilde{\mathbf{y}}_t = \mathbf{y}_t + \mathbf{v}_t. \tag{D-1}$$

We assume that there are $K < N - 1$ yields or fixed combinations (or 'portfolios') of yields, given by a fixed $N \times K$ weighting matrix $\mathbf{W}$, that are nevertheless fitted without error:

$$\mathbf{q}_{y,t} \equiv \mathbf{W}'\widetilde{\mathbf{y}}_t = \mathbf{W}'\mathbf{y}_t \tag{D-2}$$

for all t, which is equivalent to assuming $\mathbf{W}'\mathbf{v}_t = \mathbf{0}$. Following Golinski and Spencer (2024), we refer to Eq.(D-2) as the *observability restriction*. As in Joslin et al. (2011), we assume that the first three principal components of yields are observed without error, so $\mathbf{W}$ is an $N \times 3$ matrix that consists of eigenvectors corresponding to the three largest eigenvalues of $\text{Cov}(\mathbf{y}_t)$.

Denote by $\mathbf{q}_y(\mathbf{x}_t; \mathbf{\Psi})$ the vector-valued function in $\mathbb{R}^K$ that maps the latent state vector $\mathbf{x}_t$ to the observable principal components $\mathbf{q}_{y,t}$, where $\mathbf{\Psi}$ denotes a generic vector of relevant parameters, possibly different in every equation. Also, denote the inverse of $\mathbf{q}_y(\mathbf{x}_t; \mathbf{\Psi})$ by $\mathbf{q}_y^{-1}(\mathbf{q}_{y,t}; \mathbf{\Psi})$, so that:

$$\mathbf{q}_{y,t} = \mathbf{q}_y(\mathbf{x}_t; \mathbf{\Psi}) = \mathbf{W}'\mathbf{y}(\mathbf{x}_t; \mathbf{\Psi}) \iff \mathbf{x}_t = \mathbf{q}_y^{-1}(\mathbf{q}_{y,t}; \mathbf{\Psi}) = \mathbf{x}(\widetilde{\mathbf{y}}_t; \mathbf{\Psi}). \tag{D-3}$$

Substituting $\mathbf{x}_t$ back into (D-1) gives a non-linear econometric model of the cross-section of N observed yields:

$$\widetilde{\mathbf{y}}_t = \mathbf{y}(\mathbf{x}(\widetilde{\mathbf{y}}_t; \boldsymbol{\Psi}); \boldsymbol{\Psi}) + \mathbf{v}_t. \tag{D-4}$$

Thus, given a set of parameters $\boldsymbol{\Psi}$, we can find $\mathbf{x}_t$ that solves the equation

$$\mathbf{q}_{y,t} = \mathbf{W}'\mathbf{y}(\mathbf{x}(\widetilde{\mathbf{y}}_t; \boldsymbol{\Psi}); \boldsymbol{\Psi}) \tag{D-5}$$

for any t. Since there is a direct mapping between $\mathbf{q}_{y,t}$ and $\mathbf{x}_t$, the latter can be treated as observable, conditional on $\boldsymbol{\Psi}$. The parameters that are necessary for solving Eq.(D-5) are those necessary for spanning the cross-section of yields, i.e. the risk-neutral parameters: $\boldsymbol{\mu}^{\mathbb{Q}}$, $\boldsymbol{\Phi}^{\mathbb{Q}}$ and $\boldsymbol{\Sigma}$.

The likelihood function

Denote $\boldsymbol{\Theta} \equiv \{\mu_1^{\mathbb{Q}}, \boldsymbol{\lambda}^{\mathbb{Q}}, \boldsymbol{\Sigma}_{\mathbf{y}}, \delta_0^{\mathbb{S}}, \boldsymbol{\delta}_{\mathbf{1}}^{\mathbb{S}}, \boldsymbol{\mu}^{\mathbb{S}}, \boldsymbol{\Phi}^{\mathbb{S}}, \boldsymbol{\Sigma}_{\mathbf{s}}, \boldsymbol{\mu}^{\mathbb{P}}, \boldsymbol{\Phi}^{\mathbb{P}}, \boldsymbol{\Sigma}\}$ and the conditional likelihood function as

$$\log \mathcal{L}(\boldsymbol{\Theta}) = \sum_{t=2}^{T} \log \ell(\widetilde{\mathbf{y}}_t, \widetilde{\mathbf{y}}_t^s | \widetilde{\mathbf{y}}_{t-1}, \widetilde{\mathbf{y}}_{t-1}^s; \boldsymbol{\Theta}). \tag{D-6}$$

Then, the time-t conditional likelihood can be decomposed as

$$\begin{aligned} \ell(\widetilde{\mathbf{y}}_t, \widetilde{\mathbf{y}}_t^s | \widetilde{\mathbf{y}}_{t-1}, \widetilde{\mathbf{y}}_{t-1}^s; \boldsymbol{\Theta}) \quad = \quad & \ell^{\mathbb{Q}}(\widetilde{\mathbf{y}}_t | \mathbf{q}_{y,t}; \mu_1^{\mathbb{Q}}, \boldsymbol{\lambda}^{\mathbb{Q}}, \boldsymbol{\Sigma}_{\mathbf{y}}, \boldsymbol{\Sigma}) \\ & \times \ell^{\mathbb{S}}(\widetilde{\mathbf{y}}_t^s | \mathbf{q}_{y,t}; \mu_1^{\mathbb{Q}}, \boldsymbol{\lambda}^{\mathbb{Q}}, \delta_0^{\mathbb{S}}, \boldsymbol{\delta}_1^{\mathbb{S}}, \boldsymbol{\mu}^{\mathbb{S}}, \boldsymbol{\Phi}^{\mathbb{S}}, \boldsymbol{\Sigma}_{\mathbf{s}}, \boldsymbol{\Sigma}) \\ & \times \ell^{\mathbb{P}}(\mathbf{q}_{y,t} | \mathbf{q}_{y,t-1}; \boldsymbol{\mu}^{\mathbb{P}}, \boldsymbol{\Phi}^{\mathbb{P}}, \boldsymbol{\Sigma}). \end{aligned} \tag{D-7}$$

Note that the last part of the likelihood, $\ell^{\mathbb{P}}$, involves only the $\mathbb{P}$ parameters. To take advantage of the Zellner (1962) result, we need to rotate the nonlinear principal components

$\mathbf{q}_{y,t}$ to Gaussian factors $\mathbf{x}_t$. Due to the observability restriction, the parts of the likelihood function that correspond to cross sections of yields and surveys are identical regardless of whether we condition on the observed principal components or on the underlying factors, i.e.

$$\begin{aligned}
\ell^{\mathbb{Q}}(\widetilde{\mathbf{y}}_t|\mathbf{q}_{y,t};\mu_1^{\mathbb{Q}},\boldsymbol{\lambda}^{\mathbb{Q}},\boldsymbol{\Sigma}_{\mathbf{y}},\boldsymbol{\Sigma}) &= \ell^{\mathbb{Q}}(\widetilde{\mathbf{y}}_t|\mathbf{x}_t;\mu_1^{\mathbb{Q}},\boldsymbol{\lambda}^{\mathbb{Q}},\boldsymbol{\Sigma}_{\mathbf{y}},\boldsymbol{\Sigma}) \\
\ell^{\mathbb{S}}(\widetilde{\mathbf{y}}_t^s|\mathbf{q}_{y,t};\mu_1^{\mathbb{Q}},\boldsymbol{\lambda}^{\mathbb{Q}},\delta_0^{\mathbb{S}},\boldsymbol{\delta}_1^{\mathbb{S}},\boldsymbol{\mu}^{\mathbb{S}},\boldsymbol{\Phi}^{\mathbb{S}},\boldsymbol{\Sigma}_{\mathbf{s}},\boldsymbol{\Sigma}) &= \ell^{\mathbb{S}}(\widetilde{\mathbf{y}}_t^s|\mathbf{x}_t;\mu_1^{\mathbb{Q}},\boldsymbol{\lambda}^{\mathbb{Q}},\delta_0^{\mathbb{S}},\boldsymbol{\delta}_1^{\mathbb{S}},\boldsymbol{\mu}^{\mathbb{S}},\boldsymbol{\Phi}^{\mathbb{S}},\boldsymbol{\Sigma}_{\mathbf{s}},\boldsymbol{\Sigma})
\end{aligned}$$

The rotation of the factors, however, requires an adjustment to the $\mathbb{P}$ part of the likelihood. Applying the change-of-variable technique gives:[21]

$$\ell^{\mathbb{P}}(\mathbf{q}_{y,t}|\mathbf{q}_{y,t-1};\boldsymbol{\mu}^{\mathbb{P}},\boldsymbol{\Phi}^{\mathbb{P}},\boldsymbol{\Sigma}) = \ell^{\mathbb{P}}(\mathbf{x}_t|\mathbf{x}_{t-1}) \times \left|\det\left(\mathbf{J}_{x,t}\right)\right|^{-1}, \tag{D-8}$$

where $\mathbf{J}_{x,t}$ is the Jacobian term resulting from the change of variables:

$$\mathbf{J}_{x,t} \equiv \left[\frac{\partial \mathbf{q}_{y,t}}{\partial x_{1,t}}, \cdots, \frac{\partial \mathbf{q}_{y,t}}{\partial x_{K,t}}\right].$$

In the shadow rate model the first derivative for the jth yield with the remaining maturity m_j is

$$\frac{\partial y_{j,t}}{\partial \mathbf{x}_t} = \frac{1}{m_j}\frac{\partial}{\partial \mathbf{x}_t}\sum_{m=0}^{m_j-1} f(\mathbf{x}_t, m; \Psi) \approx \frac{1}{m_j}\sum_{m=0}^{m_j-1} \mathtt{N}\left(\frac{a_{f,m} + \mathbf{b}_{f,m}'\mathbf{x}_t - \underline{r}}{\sigma_m^{\mathbb{Q}}}\right)\mathbf{b}_{f,m}. \tag{D-9}$$

Denote by $\mathbf{H}_{x,t}$ stacked vectors $\partial y_{j,t}/\partial \mathbf{x}_t$:

$$\mathbf{H}_{x,t} = \frac{\partial \mathbf{y}_t}{\partial \mathbf{x}_t'} = \left[\frac{\partial y_{1,t}}{\partial \mathbf{x}_t'}, \cdots, \frac{\partial y_{J,t}}{\partial \mathbf{x}_t'}\right]'. \tag{D-10}$$

[21] See Greene (2011), Appendix B.

Then, the Jacobian $\mathbf{J}_{x,t}$ is given by

$$\mathbf{J}_{x,t} = \mathbf{W}'\mathbf{H}_{x,t}. \tag{D-11}$$

The rotation of factors from $\mathbf{q}_{y,t}$ to $\mathbf{x}_t$ allows us to find the conditional mean parameters under the $\mathbb{P}$ measure in Eq.(8), $\boldsymbol{\mu}^{\mathbb{P}}$ and $\boldsymbol{\Phi}^{\mathbb{P}}$, by *OLS*, conditional on the risk-neutral and subjective parameters that are necessary for factor rotation, which greatly facilitates the estimation since it considerably reduces the number of parameters in numerical optimization.

Appendix E Bootstrap

We use a bootstrap procedure for two purposes: to obtain finite-sample critical values for the likelihood ratio tests in Table 4, and to quantify the small-sample estimation bias in the estimated reaction function and probability forecast errors in Table 8.

Likelihood ratio tests

For the likelihood ratio tests, the null and alternative models are those defined in Section 5. For each restricted null case model $\mathcal{C}_k$, with $k \in \{1, 2, 3, 5\}$, the alternative model is the unrestricted specification $\mathcal{C}_4$. The observed likelihood ratio statistic is

$$LR_k = 2\left(\widehat{L}_4 - \widehat{L}_k\right), \tag{E-1}$$

where $\widehat{L}_k$ and $\widehat{L}_4$ denote the maximized log-likelihood values under the restricted null model and the unrestricted model, respectively. The bootstrap is implemented separately for each null hypothesis as follows:

1. For each $k \in \{1, 2, 3, 5\}$, we estimate the joint model for interest rates and surveys under the restricted null model $\mathcal{C}_k$.

2. Using the parameter estimates from $\mathcal{C}_k$, we simulate 1,000 artificial data sets with the same time-series length and cross-sectional dimension as the original sample in two steps. In the first step, we simulate the state variables $\mathbf{x}_t$ using the estimated $\mathbb{P}$-dynamics. In the second step, based on these simulated factors, we generate a cross-section of yields using the forward-rate formula of Wu and Xia (2016) in Eq. (6), and a cross section of surveys for each horizon using the survey expectation formula in Eq. (11), to which we add measurement errors. To preserve the strong persistence in

the residuals, we apply the circular stationary bootstrap of Politis and Romano (1994). Concretely, for the initial observation, we draw a joint residual pair $\{\mathbf{v}_t, \mathbf{v}_t^s\}$ at random from the residuals of the null $\mathcal{C}_k$ and set the probability of observing the consecutive set of residuals to 0.95.

3. For each bootstrapped sample generated under $\mathcal{C}_k$, we estimate both the restricted null model $\mathcal{C}_k$ and the unrestricted model $\mathcal{C}_4$, and compute the bootstrap likelihood ratio statistic $LR_{k,b} = 2(\widehat{L}_{4,b} - \widehat{L}_{k,b})$. The bootstrap p-value is the fraction of bootstrap replications for which $LR_{k,b} \geq LR_k$.

Small-sample estimation bias

To assess whether the estimated RFE and PFE can be explained by finite-sample estimation bias, we use a separate bootstrap exercise under Case 1 only. This null imposes both $RFE = 0$ and $PFE = 0$, so any RFE or PFE obtained after re-estimating the unrestricted model on the simulated samples is spurious by construction and can be interpreted as small-sample estimation bias. Specifically, we simulate 1,000 artificial data sets under Case 1, estimate the unrestricted model $\mathcal{C}_4$ on each simulated sample, and compute the RFE and PFE as in Eqs. (15)–(16). We repeat this exercise for samples of length $T = 445$, matching the original sample, and for shorter samples of length $T = 200$. These simulated RFE and PFE distributions are used to construct the bias summary statistics in Table 8 and the bias bands in Figure 3.

Appendix F Additional tables

Model	$\mu_1^{\mathbb{Q}} \times 100$	$Eig(\Phi^{\mathbb{Q}})$	
Case 0 (yield-only model)	0.0227	0.9990	$0.9499 \pm 0.0025\,i$
Case 1 (RFE=PFE=0)	0.0229	0.9992	$0.9465 \pm 0.0024\,i$
Case 2 (PFE=0)	0.0241	0.9993	$0.9473 \pm 0.0042\,i$
Case 3 (RFE=0)	0.0225	0.9993	$0.9467 \pm 0.0041\,i$
Case 4 (unrestricted)	0.0242	0.9995	$0.9472 \pm 0.0045\,i$
Case 5 ($\mu_\infty^{\mathbb{P}} = \mu_\infty^{\mathbb{S}}$)	0.0243	0.9995	$0.9473 \pm 0.0045\,i$

Table F-1. Estimates of risk-neutral parameters. The table reports the estimates of the first element of the intercept vector $\boldsymbol{\mu}^{\mathbb{Q}}$ and the eigenvalues of the autoregressive matrix $\boldsymbol{\Phi}^{\mathbb{Q}}$.

Appendix G Additional figures

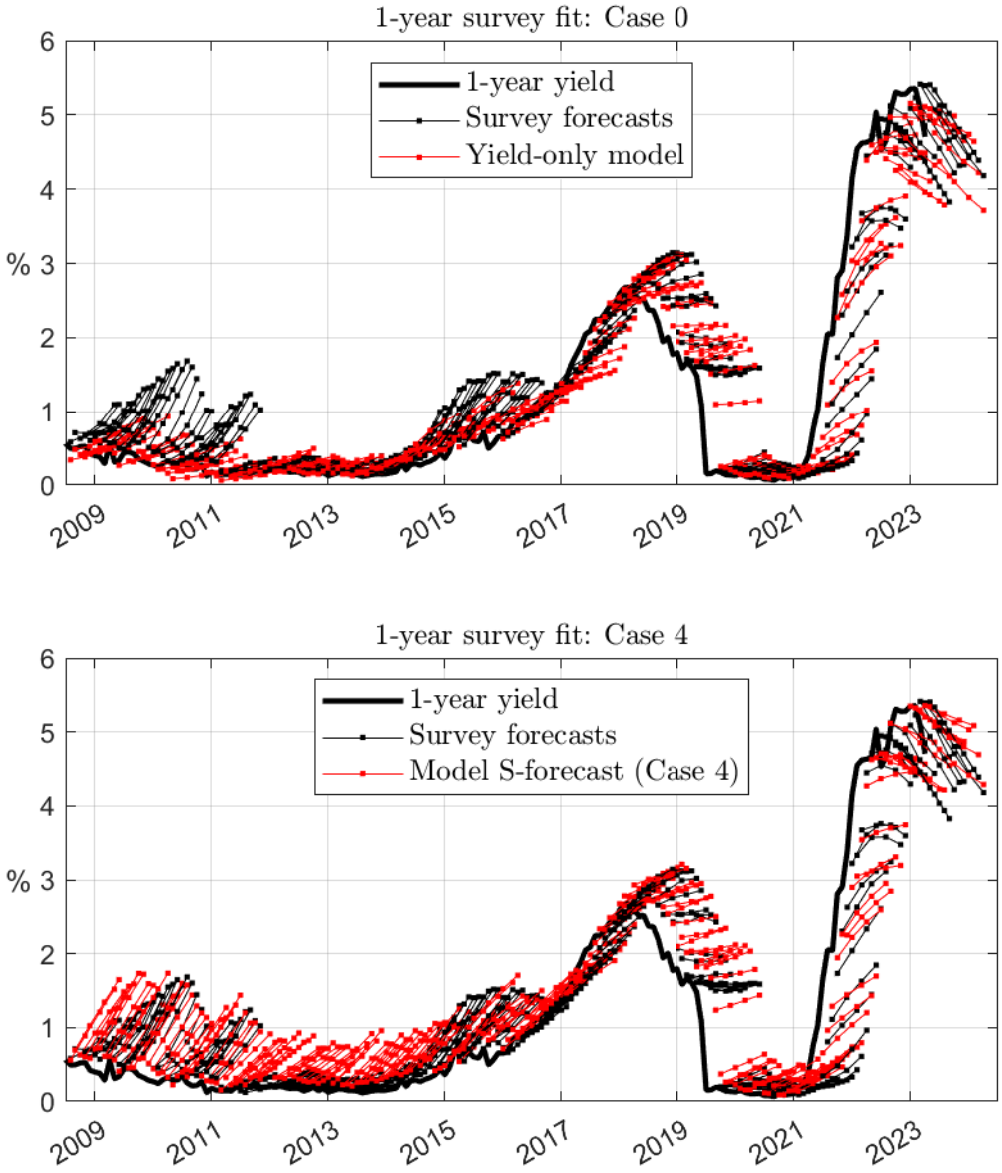


Figure G-1. Real-time survey expectations and model-implied forecasts. The black dotted lines indicate survey expectations of the 1-year rate (from BCFF) for horizons of 3, 6, 9 and 12 months made from 2009:01 to 2023:12. The red dotted lines refer to the corresponding real-time out-of-sample forecasts of the 1-year rate implied by the yield-only (Case 0) term structure model (top panel) and the unrestricted (Case 4) term structure model (bottom panel).