A GMM-Based Decomposition of the Brazilian Term Structure of Yields

Livio C. Maya*

Abstract

I estimate a no-arbitrage affine model of the term structure to decompose observed Brazilian yields and their response to identified monetary policy shocks. I introduce a GMM procedure that, through the weighting of overidentifying moment conditions, allows to control the amount of risk-premia information incorporated in the estimation of factors' dynamics. Incorporating no risk-premia information is equivalent to estimating it by OLS. Incorporating all information is equivalent to reproducing return-forecasting regressions. Does that choice matter? In the case of Brazil, I present previously undocumented evidence that the level factor accounts for the bulk (up to 92%) of long-term risk premia spanned by the first three principal components of the term structure. Compared to a model that reproduces this empirical pattern, OLS dynamics tends to understate term premia in times of elevated term structure level, and overstate the persistence of monetary policy shocks.

*Banco Safra. Email: liviomaya@gmail.com. I thank John Cochrane and Joaquim Levy for numerous useful comments.

1. Introduction

The objective of this paper is to decompose the term structure of forward rates in the Brazilian economy between expected interest and term premium components. I estimate a quarterly multifactor version of the Gaussian affine model (Dai and Singleton (2000)) by a new, two-step generalized method of moments (GMM, Hansen (1982)). The GMM procedure includes orthogonality conditions associated to return-forecasting regressions, and can thus incorporate information from long-term risk premia. The estimated model characterizes the conditional distribution of interest rates and bond risk premia. I apply it to decompose: 1. the observed time series of yields, and 2. the response of the term structure to monetary policy shocks identified using high-frequency movements in market prices (Cochrane and Piazzesi (2002)).

The estimation of the model is a central contribution of the paper. The solution of the affine model embeds a set of no-arbitrage restrictions on the dynamics of factors, bond prices and expected returns (*i.e.*, risk premia). These restrictions are most clearly represented algebraically by a linear relationship involving: the *physical* (*i.e.*, data-generating) distribution of X, its *risk-neutral* distribution (which prices the term structure), and the distribution of *market prices of risk* governing expected returns. We can freely set two of these distributions; the third is endogenous.

Once you pick risk-neutral distribution parameters to fit observed yields, the overidentifying restrictions of the affine model offer two paths to complete the estimation. They differ in the information set used to select the remaining parameters. One option is to use the time series of X and estimate its physical dynamics using, for example, OLS (as Adrian et al. (2013)); the model then gives price of risk parameters. A different route is to focus on *risk-premium* evidence to discipline the market price of risk, and then infer the physical distribution of X from the restrictions (as Cochrane and Piazzesi (2009)). Bond return predictability gets clearer as we increase investment horizons (Baker et al. (2003), Boudoukh et al. (2008), I present evidence this is the case in Brazil). Hence, the key advantage of looking at expected return evidence is that it conveys information about the long-term dynamics of bond yields, and hence of X. Such information is critical for decompositions of the term structure, as expected interest is tightly connected to the *persistence* of interest rates.

I propose a new, two-step GMM estimation procedure that unifies these seemingly opposing approaches by converting them into separate types of moment conditions. There are three of them: *pricing moments*, *OLS moments* and *risk-premia moments*. In the first step, I solve a GMM problem that targets pricing moments. It chooses risk-neutral parameters to minimize the distance between observed and model-implied yields.

The second step is the main novelty. Fixing the risk-neutral dynamics from the first step, it solves another GMM problem, which now selects the physical distribution of X (and, by extension, price of risk parameters). The GMM targets a set of overidentifying *orthogonality conditions* on: 1. the residuals of the vector autoregression (VAR) governing X (OLS moments); and 2. the residuals of the model-implied return-forecasting regressions (risk-premium moments). Hence, it uses information from two sources: the time series of X, and realized returns. The weighting matrix of the GMM problem captures the importance we attribute to each set of orthogonality conditions. When giving all the weight to OLS moments, we recover the OLS estimates of the physical law of motion of X. When giving all the weight to risk premium moments, we replicate long-term return-forecasting regressions. I describe all results in the paper using three weighting matrix choices: the one leading to OLS, the one replicating return-forecasting regressions and an "intermediary" one that gives equal weight to each targeted moment.

Besides providing a natural framework to work with overidentifying restrictions, the GMM approach offers other advantages. It allows the comparison between competing estimation strategies in a single setup. It provides a ready-to-use distribution theory for point estimates. It does not require the introduction of measurement errors to deal with the stochastic singularity problem of the affine model.

A second contribution of the paper is the application to the Brazilian bond market. Studies of predictability in emerging economies are relatively scarce (Akgiray et al. (2016) and Devpura et al. (2021) are recent examples), as many markets lack the required data or liquidity. In the case of Brazil, we now have over ten years of market price data on long-term bond prices, provided by the Brazilian Financial and Capital Markets Association (ANBIMA), to study the connection between yields and risk premia.[1]

To discipline the choice of which risk-premium moments to include in the GMM estimation, I run a series of return-forecasting regressions with the Brazilian data. Two patterns emerge. First: bond excess returns are predictable. In particular, the commonly used *level*, *slope* and *curve* factors of the term structure capture the bulk of predictability spanned by yields. Predictability increases over investment horizons, and reaches 54% in the three-year case. This pattern motivates me to use these three factors as state vector X in the affine model.

Second: as we increase investment horizon, movements of the *level* of the term structure capture an increasing share - up to 92% - of the variance of risk premia spanned by X. To the best of my knowledge, this is a previously undocumented pattern of Brazilian returns. Long-term bond returns tend to be higher when the level of the term structure is higher. One might find this result expected, and attributable to interest mean-reversion. But the key information we take from the regressions refers to *timing*. *It usually takes two years* for the level factor to deliver high returns. Over investment horizons of one year or less, when returns are less predictable, the pattern vanishes: the level of the term structure explains returns about as much as its slope and curvature.

Considering the empirical findings, I include in the GMM estimation the orthogonality condition that defines the return-forecasting regressions for two-year investment horizons. We want the affine model to replicate the return-predictive power of the level factor.

However, the GMM estimator that targets only OLS moments (*i.e.*, the OLS estimate of

[1] Papers studying the Brazilian term structure typically use swap data on the overnight lending rate (the DI rate). Datasets beginning prior to 2009 usually contain data on bonds with maturity up to two years. See, for instance, Tabak and Andrade (2003), Osmani and Tabak (2008), Almeida and Faria (2014), Almeida et al. (2015) and Faria and Almeida (2018). As I show in the text, availability of long-term bond yield data is critical to identify the persistence of the risk-neutral distribution.

the physical distribution of X) soundly fails to match risk-premium moments. The share of two-year risk premia variance accounted for by the level of yields is lower - by as much as half - than what is implied by return-forecasting regressions. Hence the OLS model leads to risk premia dynamics at odds with the evidence. As we change the weighting matrix of the GMM problem to privilege risk-premia moments, we get increasingly closer to the return-forecasting regressions. On the other hand, model shocks become more correlated with the state X.

How do these changes in the set of moments we focus affect the decompositions of the Brazilian term structure? As we attribute more weight to risk-premium moments and thus deviate from OLS, interest dynamics presents *higher frequency* and *amplified* cycles, particularly in response to a level shock. Its impulse-response functions gain an "S" shape, as opposed to an uniformly-decaying, AR(1) look under OLS. The faster and larger decline of interest after a positive level shock leads to the higher expected returns the return-forecasting regressions call for. Consequently, periods of elevated term structure level are more attributed to *term premia*, since agents expect interest to decline faster and by a larger amount.

Finally, I also apply the model to decompose the effect of an identified monetary policy shock on the term structure. Following the tradition started by Cochrane and Piazzesi (2002) (also Gertler and Karadi (2015) and Nakamura and Steinsson (2018), for example), I identify monetary surprises through changes in the interest implied by futures contracts around policy rate announcements. Gomes et al. (2023) and Da Costa Filho (2021) follow a similar approach using Brazilian data. Since model factors X also derive from market prices, I can observe them in high frequency and measure the average effect of monetary surprises. They mostly associate with movements in the *slope* of the term structure. A 0.1% surprise in the policy rate causes the slope factor to increase by 7.2% of one standard deviation. (Short-term forward rates increase relative to long-term ones.)

The signs of the shock on each of the three factors point to a *decline in risk premia, regardless of weighting matrix specification*. Term premia falls in the quarter of the policy announcement.

Therefore, the negative inclination of the term structure overstates agents' expectations of how fast the effect of the monetary shock on interest will fade. And this is where moment weighting in the GMM estimation matters. The model estimated by OLS says that the effect of the monetary shock on interest rates fades slowly over time, akin to a persistent AR(1). On the other hand, the specification that puts all weight on risk-premia moments and the "intermediary" one agree that the positive effect of the monetary shock on interest lasts ten quarters.

Taken together, the results I present for the Brazilian bond market reinforce the importance of carefully considering the choice/weighting of different empirical information to estimate term structure models.

2. The Affine Model of the Term Structure

2.1. Notation

One period corresponds to a quarter. The notation is the following: p_t^n is the log-price of a zero-coupon bond that pays one unit of currency after n quarters; $f_t^n = p_t^{n-1} - p_t^n$ is the log forward rate; $i_t = f_t^1$ is the one-period interest rate; and

$$\begin{aligned} rx_{t,t+m}^n &= p_{t+m}^{n-m} - p_t^n - \left[i_t + f_t^2 + \cdots + f_t^m \right] \\ &= p_{t+m}^{n-m} - p_t^n + p_t^m \end{aligned}$$

is the return on the holding of an n-maturity bond for m quarters, in excess of the interest charged on an m-period risk-free borrowing $-p_t^m = i_t + f_t^2 + \cdots + f_t^m$. When $m = 1$ we get the one-period excess return $rx_{t,t+1}^n = p_{t+1}^{n-1} - p_t^n - i_t$. There is one risk premium $E_t rx_{t,t+m}^n$ to each pair of bond maturity n and investment horizon m. I use the expressions "risk premium" and "expected excess return" interchangeably.

2.2. Forward Rate Decomposition

By no arbitrage, the n-maturity forward rate f_t^n is the interest rate at which one can lock, in period t, a borrowing that starts in $t+n-1$, with re-payment due in $t+n$. As such, there should be a connection between f_t^n and the interest rate expected to prevail in $t+n-1$, $E_t i_{t+n-1}$. It is possible to show that

$$f_t^n = \underbrace{\sum_{j=0}^{n-2}\left[E_t rx_{t+j,t+j+1}^{n-j} - E_t rx_{t+j,t+j+1}^{n-j-1}\right]}_{\text{Term Premium } tp_t^n} + E_t i_{t+n-1}. \tag{1}$$

Equation (1) decomposes the forward rate between an expected interest component and a term premium component that depends on a combination of expected excess returns.[1]

The objective of this paper is to measure decomposition (1) using information from the cross-section of yields, interest rate dynamics and risk premia. (In theory, either a model of expected interest or a model of risk premia suffice to measure the decomposition, as they provide one of the terms on the right-hand side of (1) - we can infer the other.)

2.3. The Affine Model: Main Equations

Following Ang and Piazzesi (2003) and Ang et al. (2007), I use a discrete-time version of the Gaussian exponential-affine model presented by Duffie and Kan (1996). This section presents the main equations necessary to understand the mechanics of the affine model. I defer to the appendix a more detailed description.

The state is a P-sized vector process X_t that evolves according to the law of motion

$$X_t = \mu + \Phi X_{t-1} + e_{t+1} \qquad e_{t+1} \sim N(0, \Sigma). \tag{2}$$

[1] The decomposition holds in expected value and for any path of bond prices (i.e., without expected values). The identity follows from the fact that the forward rate is the difference between the price of bonds that eventually pay the same one unit of currency. Cochrane and Piazzesi (2009) provide pretty art illustrating it.

The conditional distribution of X_t in (2) is the "physical" distribution because it generates actual data. I call $X_{t,p}$ the p-th element of X_t, and the same for $e_{t,p}$ and other vectors in the model.

The Gaussian model imposes the absence of arbitrage and the log stochastic discount rate

$$m_{t+1} = -\left[\delta_0 + \delta_1' X_t\right] - \frac{1}{2}\lambda_t' \Sigma \lambda_t - \lambda_t' e_{t+1}. \tag{3}$$

Vector $\lambda_t = \lambda_0 + \lambda_1 X_t$ is linear on the factors and captures the sensitivity of m_{t+1} to each of the P reduced-form shocks e_{t+1}. Discount rate (3) implies that the interest rate is

$$i_t = \delta_0 + \delta_1' X_t. \tag{4}$$

Bond prices are also an affine function of the state:

$$p_t^n = A_n + B_n' X_t. \tag{5}$$

Coefficient A_n and P-sized vector B_n solve well-known Riccati equations outlined in the appendix. These Riccati equations constitute the main restrictions imposed by no-arbitrage on the cross-section of yields. In their absence, we could estimate A_n and B_n by standard OLS. To derive (5), we use the main pricing condition of the model:

$$E_t r x_{t,t+1}^n = v_n + \text{cov}_t(r x_{t,t+1}^n, e_{t+1}')\ \lambda_t = v_n + B_{n-1}' \Sigma\ \lambda_t \tag{6}$$

where v_n is a small Jensen inequality constant.[1] You can read (6) as a beta model. The risk premium on a one-quarter holding of a given bond equals the product of the conditional covariances between its return and the shocks (the *quantity of risk*) and the discount rate sensitivity to these shocks λ_t (the *price of risk*). If the conditional covariance of the excess return with shock $e_{t,p}$ is 1% larger, investors demand $\lambda_{t,p}$% higher risk premium to hold the bond.

[1] I define $v_n = -\frac{1}{2} B_{n-1}' \Sigma B_{n-1}$.

Given the format of solution (5), coefficient B_{n-1} determines this conditional covariance, which is constant over time. Thus, all variation in risk premia follows from variation in the price of risk λ_t.

Critically for our estimation strategy, (6) also implies that we can price bonds using a "risk-neutral" distribution of X_t, different from the "physical" distribution generating the data. Under this risk-neutral distribution, X_t follows

$$\begin{aligned} X_t &= \mu^* + \Phi^* X_{t-1} + e_{t+1} \qquad e_{t+1} \sim N(0, \Sigma), \\ \mu^* &= \mu - \Sigma\,\lambda_0, \\ \Phi^* &= \Phi - \Sigma\,\lambda_1. \end{aligned} \tag{7}$$

The pair (μ^*, Φ^*) determines the drift of X_t under this new measure. For a fixed (μ^*, Φ^*), coefficients A_n and B_n do not depend on (μ, Φ) or (λ_0, λ_1). Thus, to price bonds all we need is the risk-neutral distribution. We do not need to know the physical distribution (2) or the market price of risk λ_t governing risk premia (6).

Finally, the appendix shows that

$$f_t^n = v_n + E_t^* i_{t+n-1} = \text{constant} + \delta_1' \Phi^{*n-1} X_t, \tag{8}$$

for all n, where expectation E_t^* integrates using (7). In words, if agents expect X_t to evolve according to (μ^*, Φ^*) instead of (μ, Φ), there is no term premium (except for the small v_n), and forward rates reflect expected short-term interest. Hence the "risk neutral" designation.[1]

In all, we can summarize the affine model as follows:

- (μ, Φ) govern the dynamics of X_t through (2);
- (μ^*, Φ^*) govern the cross-section of yields through (8);

[1] We build the risk-neutral measure as a measure under which prices would be observationally equivalent if risk was not priced ($\lambda_t = 0$). By the pricing condition (6), that requires $E_t^* rx_{t,t+1}^n = v_n$, an equation that yields the terms μ^* and Φ^* defined in (7) as well as equation (8).

- (λ_0, λ_1) govern risk premia through (6).

By (7), the affine model allows us to fix two of these pairs of parameters, but not the third. This restriction leads to a trade-off during estimation. We need to carefully consider which aspects of the data we want to reproduce more closely. The GMM framework brings this trade-off to light through the imposition of overidentifying moment conditions which we can weight differently during estimation.

3. Yields and Risk Premia in Brazil

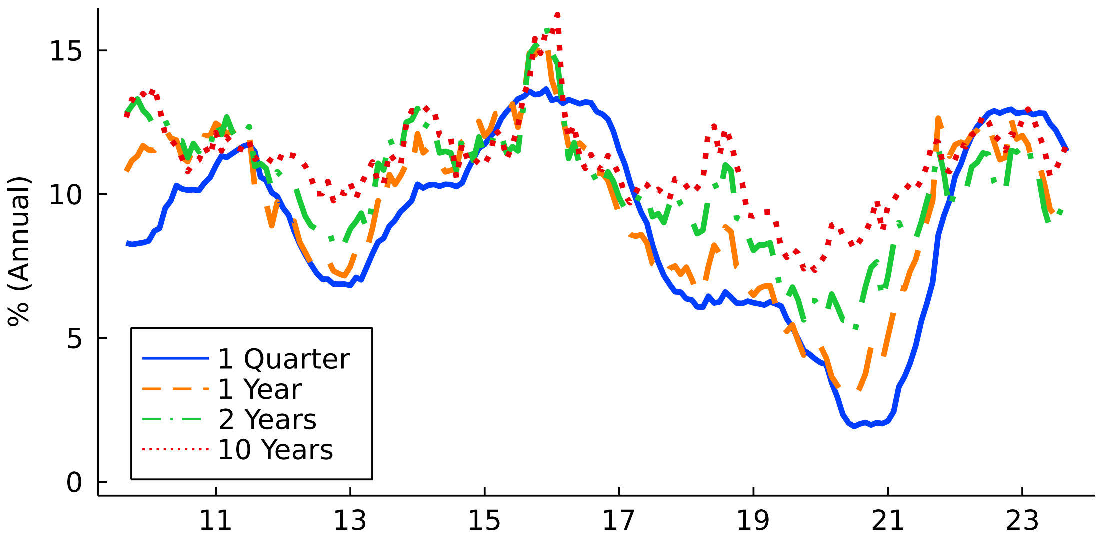
3.1. Data

The Brazilian Financial and Capital Markets Association (ANBIMA) collects market price data of nominal bonds issued by the Brazilian federal government. At any point in time, available bonds have different maturities that do not coincide with a complete array of uniformly spaced quarters, hence some interpolation is necessary. An interpolation similar to that of the most commonly used Fama and Bliss (1987) dataset is not available. Instead, ANBIMA provides an estimate of the term structure of interest rates through a daily series of estimated load and decay parameters of a fitted Svenson (1994) model. It estimates these parameters to approximate the discount applied to the most liquid trading bonds.[1] The closest analogue to this interpolation approach for the US is studied by Gürkaynak et al. (2007).

The Svenson (1994) model contains a total of six parameters, and provides a lot of flexibility. But how far is it from actual price data? Not much. I measure the distance between ANBIMA yields and actual market price yields of available zero-coupon bonds, and find a root mean squared error (RMSE) of 0.044% in the annualized yield to maturities.[2] In comparison, linear regressions of forward rates on their first three principal components, which explain over

[1] ANBIMA provides market prices of different traded bonds, but not liquidity data, therefore replication of the models' parameters is not feasible.

[2] The RMSE of an error series $\{\epsilon_t\}$ is defined as $\left[T^{-1}\sum_{t=1}^{T}\epsilon_t^2\right]^{1/2}$. "Annualized" means I multiply RMSE by four (since yields are in quarterly units).



Notes. I multiply forward rates by four to annualize them.

Figure 1: Quarterly Forward Rates f_t^n (Data)

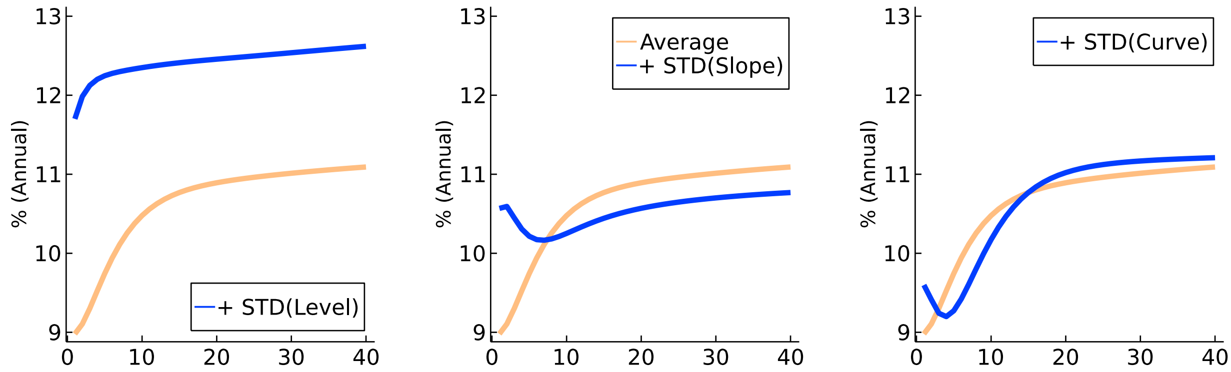
99% of the quadratic variation of yields, result in an average RMSE of 0.19%. The appendix provides additional details.

I use ANBIMA data to build monthly observations of the prices of bonds with maturities $n = 1, 2, \ldots, 40$ quarters. Price data then leads to a full term structure of forward rates (figure 1) and excess returns (figure 3). It covers the period 2009-M1 to 2023-M9. Each data point corresponds to the last trading day of the corresponding month.

3.2. Factors

The vector of factors X_t contains the first three principal components of the term structure of forward rates.[1] Like in the US case, these three factors capture the bulk (> 99%) of cross-section variation of Brazilian forward rates (Litterman and Scheinkman (1991), Knez et al. (1994), Dai and Singleton (2000)) and therefore stands as a natural candidate for model factors.

[1] The principal component decomposition amounts to the spectral decomposition of the sample covariance matrix of f_t^n. Letting λ_i be the i-th eigenvalue, the share of variance attributable to the first principal component is $\lambda_1 / \sum_i \lambda_i$.



(a) Level **(b)** Slope **(c)** Curve

Notes. Each plot shows the average term structure of forward rates in the data (orange) and the effect of a one standard deviation change in the level/slope/curve principal component (blue).

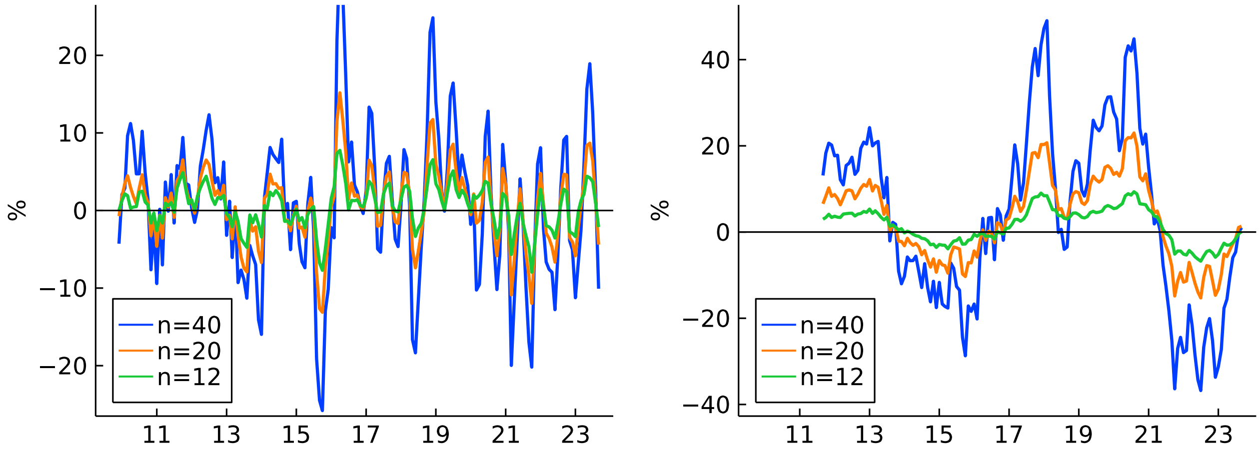
Figure 2: Term Structure of Forward Rates: Effect of Change in Principal Components

As usual, they describe the *level*, *slope* and *curvature* of the term structure of forward rates.

$$X_t = \begin{bmatrix} \text{Level}_t \\ \text{Slope}_t \\ \text{Curve}_t \end{bmatrix} \qquad \text{(normalized to mean = 0, standard deviation = 1\%).}$$

I choose signs so that the interest rate is increasing in all three factors. Figure 2 depicts the effects of a one standard deviation change in each factor on the average term structure. When Level_t is large, all yields are high. When Slope_t is large, short-term yields are high relative to long-term yields. When Curve_t is large, short and long-term yields are high relative to mid-term yields.

By construction, principal components have zero sample covariance with each other. I also normalize each factor to have a sample mean of zero and a standard deviation of 1% (for comparison, the quarterly interest rate has a standard deviation of 0.8%). The appendix plots X_t and forward rates' loadings on it.



(a) 1-quarter holdings ($m = 1$) **(b)** 2-year holdings ($m = 8$)

Notes. Each curve corresponds to the excess return on a bond of a given maturity $n \in (12, 20, 40)$.

Figure 3: Excess Returns $rx^n_{t,t+m}$ (Data)

3.3. Return Predictability

One central tenet of modern empirical finance is that returns are, to some extent, predictable (Fama and French (1988), Cochrane (2008), Golez and Koudijs (2018), and many others). Fixed-income markets are no exception (Fama (1984), Cochrane and Piazzesi (2005), Andreasen et al. (2021)), and, as I show below, neither is Brazil. In the affine model, the market price of risk λ_t governs risk premia, which means we can use the available evidence on return predictability to estimate it. In turn λ_t ties together the physical and risk-neutral distributions of model factors X_t.

In this section, I study predictability in the Brazilian bond market by running return-forecasting regressions of the format

$$rx^n_{t,t+m} = a^{(n)}_m + b^{(n)}_m \cdot X_t + \varepsilon^n_{t,t+m}. \tag{9}$$

The empirical patterns that emerge guide the selection of moment conditions we target in the GMM estimation of the affine model, and deeply affect results and their interpretation. I analyze these patterns separately.

Empirical Pattern 1: The level, slope and curve factors forecast returns. Predictability starts at $R^2 = 11\%$, and grows with investment horizon m up to 54%.

The top panel of table 1 reports OLS estimates of (9). The terms in parentheses denote p-values for the test $b_{m,x}^{(n)} = 0$. ($b_{m,x}^{(n)}$ is the element of $b_m^{(n)}$ corresponding to factor x.) I compute standard errors using the Newey-West estimator with 24 lags (two years). The table reports results for four investment horizons: one quarter, and one through three years ($m = 1, 4, 8, 12$); and three bond maturities: three, five and seven years ($n = 12, 20, 28$). Results for other maturities are qualitatively very similar. Panels 3a and 3b plot excess returns in the cases of one quarter and two year investment horizons.

The regressions lead to R^2s that grow with horizon m, starting at 11% for the quarterly return and reaching 54% in the three-year horizon. Increasing predictability in investment horizons is a well-known pattern in the empirical finance literature (see Baker et al. (2003)), often associated to high-frequency noise that averages out over time (which is visible if we compare the two panels in figure 3).[1] For comparison, Cochrane and Piazzesi (2005) find R^2's in the range of 30 to 40% in the case of US annual returns. A Wald test of joint significance $b_m^{(n)} = 0$ (unreported) rejects the null of a spurious model at 1% significance when $m > 1$ and 4% significance when $m = 1$. Looking at point estimates, returns are higher when the term structure displays a high level, steepness or curvature.[2] The estimated loadings are economically significant in most cases. For example: a one standard deviation increase in the level factor raises returns on the five-year bond ($n = 20$) by 1.5 to 9%, depending on the investment horizon. For the slope and curve factors, the impact reaches 3 and 4%, respectively. These results reflect the elevated degree of predictability spanned by the three principal component factors X_t.

[1] Many studies emphasize that growing R^2 is not due to growing predictability *per se* (Kirby (1997), Boudoukh et al. (2008), Farmer et al. (2023)). Instead, it is the expected outcome of having a persistent predictive variable such as our factors X_t or forward rates themselves.

[2] Recall that the Slope_t factor is increasing in short-term interest and decreasing in steepness. The estimated negative loadings on Slope_t are therefore consistent with the classical Fama and Bliss (1987) finding that, in the US, a positive inclination of the term structure forecasts higher returns.

Model:	$rx^n_{t,t+m} = a^{(n)}_m + b^{(n)}_m \cdot X_t + \varepsilon^n_{m,t}$					
	Level$_t$	Slope$_t$	Curve$_t$	R^2(%)	χ^2_{Level}	R^2_f(%)
Investment Horizon: $m = 1$ (one quarter)						
$n = 12$	0.44 (0.21)			3		22
	0.44 (0.23)	-0.72 (0.01)	0.31 (0.08)	11	(0.01)	
$n = 20$	0.96 (0.08)			4		27
	0.97 (0.10)	-1.52 (0.00)	0.75 (0.02)	17	(0.00)	
$n = 28$	1.49 (0.04)			5		28
	1.49 (0.06)	-2.26 (0.00)	1.28 (0.01)	20	(0.00)	
Investment Horizon: $m = 4$ (one year)						
$n = 12$	1.67 (0.06)			12		33
	1.74 (0.05)	-0.37 (0.65)	1.35 (0.02)	19	(0.05)	
$n = 20$	3.55 (0.01)			17		44
	3.68 (0.01)	-1.76 (0.19)	2.69 (0.00)	30	(0.01)	
$n = 28$	5.51 (0.00)			20		49
	5.70 (0.00)	-3.14 (0.08)	4.08 (0.00)	37	(0.00)	
Investment Horizon: $m = 8$ (two years)						
$n = 12$	2.43 (0.00)			40		52
	2.53 (0.00)	-0.08 (0.89)	0.94 (0.10)	43	(0.16)	
$n = 20$	5.59 (0.00)			38		52
	5.75 (0.00)	-0.78 (0.57)	1.92 (0.16)	42	(0.31)	
$n = 28$	8.33 (0.00)			39		53
	8.54 (0.00)	-1.43 (0.48)	2.68 (0.17)	43	(0.31)	
Investment Horizon: $m = 12$ (three years)						
$n = 20$	5.01 (0.00)			47		55
	5.28 (0.00)	-0.05 (0.96)	1.58 (0.05)	50	(0.01)	
$n = 28$	8.75 (0.00)			50		59
	9.21 (0.00)	-0.56 (0.73)	3.05 (0.04)	54	(0.00)	

Notes. Monthly data. The dependent variable is the excess return $rx^n_{t,t+m}$, where t is in quarter units, n is the bond maturity, and m is the investment horizon. p-values in parentheses. Standard errors computed using Newey and West (1987) estimator with 24 lags. The χ^2_{Level} column reports the p-value of the Wald test of $b^{(n)}_{m,\text{Slope}} = b^{(n)}_{m,\text{Curve}} = 0$, that is, a specification that involves only the level factor on the right-hand side. The R^2_f column reports the R^2 from a regression with the forward rates of maturities $n = 4, 8, 12$ quarters on the right-hand side (instead of X_t).

Table 1: Return-Forecasting Regressions

Can we improve our forecasts by using additional information from the term structure? Column R_f^2 of table 1 reports the R^2 of an alternative regression model

$$rx^n_{t,t+m} = \alpha + \beta \cdot F_t + \varepsilon_{t,t+m}, \tag{10}$$

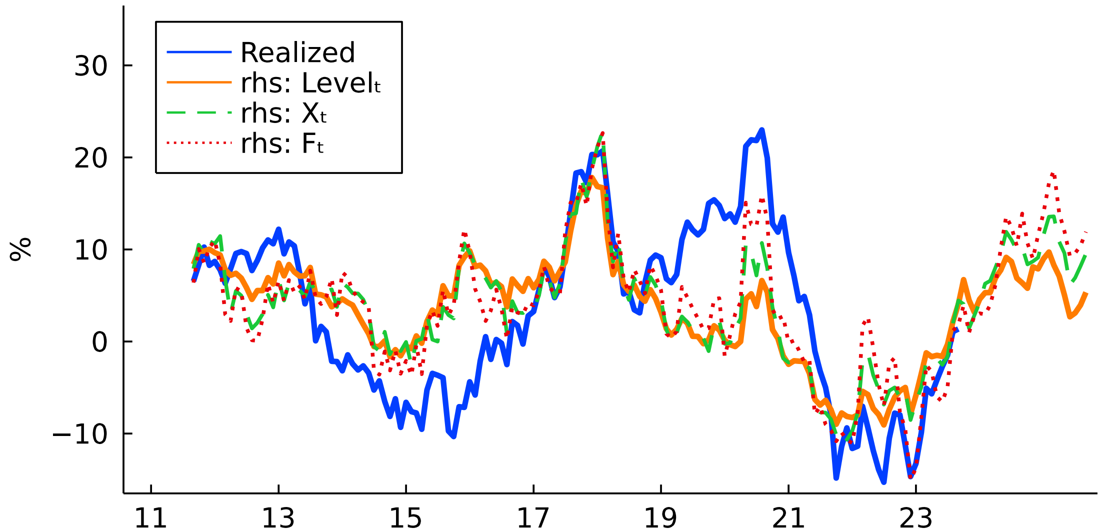
where vector F_t groups forward rates of maturities one, two and three years ($n = 4, 8, 12$; I try other combinations, and this one often leads to the best fit).[1] The table shows that we do gain some forecasting power, although these gains more or less decline as the investment horizon increases. In the three-year horizon regressions, the R^2 does not improve by more than 5%. In the two-year horizon, it increases by about 10%, but the additional predictability derives mostly from the high returns realized at the onset of the COVID pandemic, when the Brazilian central bank begins to cut interest rates. Figure 4, which plots fitted values from each model, illustrates that point. These interest cuts were hardly forecastable two years prior to COVID, hinting at some degree of model overfit in the case of regression (10). Overall, I do not find the gains in predictability sufficiently strong to justify adding more variables to the state vector X_t, particularly if we focus on investment horizons of two years or longer.

Empirical Pattern 2: As the investment horizon m grows, the Level_t factor accounts for an increasing share (up to 92%) of the variance of risk premia spanned by X_t.

Empirical pattern 2 is the main evidence on Brazilian bond risk premia I present in this paper, and the key risk premium condition I explore to estimate the affine model.

Results from the OLS regression (9) underscore the increasing return-predicting power of the level factor as investment horizon m grows. First, coefficients $b^{(n)}_{m,\text{Level}}$ become more economically significant. With quarterly returns, $b^{(n)}_{m,\text{Level}}$ is about as large as the other two loadings; in the two-year horizon case, it is more than twice as large. Second, estimates of $b^{(n)}_{m,\text{Level}}$ become more statistically significant. With $n = 20$, for example, the p-value declines

[1] Including more forward rates renders estimates too volatile due to multicollinearity, so I stop at three regressors. In any case, including more variables on the right-hand side does not increase R^2 too much.



Notes. Blue: realized excess return data ($rx^{20}_{t,t+8}$). Orange: model (12), fit projected by Level_t. Green, dash: model (9), fit projected by X_t. Red, dot: model (10), fit projected by F_t.

Figure 4: Return predictability ($rx^{20}_{t,t+8}$) under different models

from 0.10 using quarterly returns, to zero in the regressions with two and three-year returns. It is also zero for other maturities n. On the other hand, point estimates of the loadings on Slope_t and Curve_t become less precisely estimated, and more indistinguishable from zero, particularly in the two-year horizon regressions, $m = 8$. In this latter case, a Wald test of the joint hypothesis $b^{(n)}_{m,\text{Slope}} = b^{(n)}_{m,\text{Curve}} = 0$ fails to reject the null (p-values in table 1's χ^2_{Level} column).

In addition, the level factor accounts for increasing shares of the variance of risk premia $E_t rx^n_{t,t+m}$ spanned by X_t. This conclusion follows from the increasing loadings of excess returns on Level_t compared to the other two factors. Since the three regressors are orthogonal by construction and have the same $(1\%)^2$ variance, coefficients $b^{(n)}_m$ pin down the variance decomposition:

$$\text{var}\left(E_t rx^n_{t,t+m}\right) = \left[b^{(n)2}_{m,\text{Level}} + b^{(n)2}_{m,\text{Slope}} + b^{(n)2}_{m,\text{Curve}}\right] \times (1\%)^2. \tag{11}$$

In the quarterly return regression $m = 1$, the share of variance accounted for by $b^{(n)2}_{m,\text{Level}}$ is

24-25%: most predictability comes from Slope_t and Curve_t. But it grows to 88-89% in the regression with two-year returns and 90-92% in the three-year case. Almost all predictability of two and three-year bond returns stems from variation in the level factor.

We can make the point visually. Figure 4 plots, in solid red, the two-year risk premium on five-year bonds $E_t rx^{20}_{t,t+8}$ implied by the restricted model

$$rx^{n}_{t,t+m} = a^{(n)}_{m} + b^{(n)}_{m,\text{Level}}\ \text{Level}_t + \varepsilon^{n}_{t,t+m}. \tag{12}$$

We can hardly visualize any difference between the fitted risk premium spanned by Level_t (model (12)), and that spanned by X_t (model (9)). Again the most striking difference occurs at the start of the unpredictable COVID pandemic. Table 1 reports OLS estimates of (12). As the investment horizon m grows, the R^2 produced by the level factor becomes increasingly close to that produced by the entire state vector X_t. (Since we estimate all equations by OLS, the ratio of R^2 in model (12) to that in the model (9) also represents the share of risk premium variance attributable to Level_t.)

In conclusion: among X_t, the level factor is the key driver of long-term risk premia. We want the affine model to replicate this pattern. How it affects the other aspects of the model and hence decomposition (1) is the next topic.

4. GMM Estimation

Our job is to estimate the parameters of the affine model. They relate to: physical dynamics (μ, Φ), risk-neutral dynamics (μ^*, Φ^*), price of risk (λ_0, λ_1), interest rate (δ_0, δ_1) and shock covariance (Σ). I use the hat notation to differentiate parameter estimates from their true values.

I start by forming an estimate $\hat{\Sigma}$ of Σ by running OLS on (2) and using the resulting fitted errors $\hat{e}_t^{\text{OLS}}$:

$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T} \hat{e}_t^{\text{OLS}} \hat{e}_t^{\text{OLS}\prime}.$$

Using a first-stage estimate of Σ is common in the affine model literature. Re-computing it using the final estimates $(\hat{\mu}, \hat{\Phi})$ makes little difference.

I also estimate interest parameters $(\hat{\delta}_0, \hat{\delta}_1)$ by running OLS on (4). Table 2 reports results. Another possibility would be to estimate (δ_0, δ_1) jointly with the remaining parameters, targeting the entire term structure $\{f_t^n\}$ instead of $i_t = f_t^1$ only. However, mean squared errors of the interest rate equation (4) are the largest across maturities n, even if we estimate (δ_0, δ_1) to minimize it (the appendix shows a plot of average errors by maturity).[1]

Fixing $(\hat{\Sigma}, \hat{\delta}_0, \hat{\delta}_1)$, we focus on the remaining parameters (μ, Φ), (μ^*, Φ^*), and (λ_0, λ_1), which respectively govern the dynamics of X_t (hence expected interest $E_t i_{t+j}$), model pricing and risk premia. The affine model constrains these parameters through (7), so we are not free to optimize all aspects of the model. I apply a two-step GMM procedure that highlights which of these aspects we prioritize.

Both steps of the estimation procedure solve standard GMM problems. For a generic moment condition $E[h_t(\theta)] = 0$, where h_t is a function of parameter vector θ and the data, they involve minimizing the objective function

$$\text{Min}_\theta \quad g_\theta' \, W \, g_\theta \tag{13}$$

where $g_\theta = T^{-1}\sum_{t=1}^{T} h_t(\theta)$ is the sample counterpart of the moment condition and W is a positive definite weighting matrix.[2] In both steps of the estimation, I use numerical search algorithms to solve (13).

[1] The R^2 for the interest rate regression is 97.7%.

[2] See Hamilton (1994) for a textbook treatment of GMM estimators.

4.1. First Step: Term Structure Pricing

We want the Gaussian model to price the term structure of forward rates as well as possible, for two reasons. First, we are ultimately interested in the forward rate decomposition (1); and since the decomposition only holds using model-implied forward rates $\hat{f}_t^n$, the exercise becomes pointless if they are too different from observed rates. Second, as Cochrane and Piazzesi (2009) point out, in any period t the cross-section of forward rates $\{f_t^n\}_{n=1}^N$ allows us to very precisely estimate risk-neutral dynamics (μ^*, Φ^*) governing factors' drift $\{E_t^* X_{t+n}\}_{n=1}^N$, through equation (8): $f_t^n = v_n + \delta_0 + \delta_1' E_t^* X_{t+n-1}$. Estimation of the physical dynamics (μ, Φ), on the other hand, is subject to substantial model and statistical uncertainty, as I discuss below.

The solution to bond prices (5) implies an also affine solution to forward rates $f_t^n = A_n^f + B_n^{f\prime} X_t$.[1] By equation (8) (see last paragraph), given $(\hat{\Sigma}, \hat{\delta}_0, \hat{\delta}_1)$, A_n^f and B_n^f depend only on (μ^*, Φ^*). Hence, the first step of the estimation chooses (μ^*, Φ^*) to minimize pricing errors. In the GMM formulation (13), it matches the moment condition below.

Moment Condition 1 (Pricing Errors)**:**

$$E\left[\epsilon_{n,t}^2\right] = 0 \qquad \text{where} \quad \epsilon_{n,t} = f_t^n - \left(A_n^f + B_n^{f\prime} X_t\right). \tag{M1}$$

In the affine model, pricing errors are identically zero $\epsilon_{n,t}^2 = 0$, so moment M1 holds trivially. By targeting M1, the GMM algorithm minimizes squared pricing errors, like OLS. I include the ten maturities $n = 2, 6, 10, \ldots, 38$.[2] The weighting matrix is the identity $W = I$. Despite having more free parameters than moment conditions, we cannot perfectly match any moment, since A_n^f and B_n^f are nonlinear functions of (μ^*, Φ^*). Still, the affine model produces small pricing errors. The RMSE is 0.202% (annualized), not much larger than the

[1] Parameters A_n^f and B_n^f satisfy $A_n^f = A_{n-1} - A_n$ and $B_n^f = B_{n-1} - B_n$. Their initial conditions are $A_1^f = \delta_0$ and $B_1^f = \delta_1$. See the appendix.

[2] Throwing away some yield data reduces parameter variance due to the large correlation between forward rates. Using the whole sample leads to near identical point estimates.

0.185% resulting from unrestricted OLS (which produces the smallest RMSE by construction). The appendix plots them.

Table 2 summarizes the results from the GMM estimation (terms in parentheses are standard deviations, see table notes). The top, left panel shows the estimated risk-neutral autoregressive matrix $\hat{\Phi}^*$ and the absolute value of its eigenvalues. The largest one is 0.997, pointing at near unit-root behavior. The first row of $\hat{\Phi}^*$ shows that this unit root stems from the level factor law of motion, which approximates an AR(1) (the loadings of Level_t on Slope_{t-1} and Curve_{t-1} are -0.03 and -0.01), with an autocorrelation of 0.98. The appendix plots the impulse-response functions (IRF) of the risk-neutral distribution to each shock and makes this point visually.

It is well known in the bond pricing literature that risk-neutral dynamics must be highly persistent. This is an implication of the fact that the first principal component of the term structure (the level factor) accounts for most variation of forward rates.[1] The affine model can only replicate that pattern if most innovations that raise interest rates increase long-term yields as well. In turn, risk-neutral pricing equation (8) ($f_t^n = \text{constant} + \delta_1' \Phi^{*n-1} X_t$) says that long-term rates can only increase if Φ^{*n-1} does not converge too fast, which is why it needs to have a root close to one.

4.2. Second Step: Physical Distribution and Risk Premia

The second step involves estimating the physical distribution (μ, Φ) and price-of-risk parameters (λ_0, λ_1). Given $(\hat{\mu}^*, \hat{\Phi}^*)$, the building of the risk-neutral distribution (7) implies we can only choose one of them. I start by assuming that factors' sample averages (zero by construction) coincide with their unconditional averages in the model, thus $\hat{\mu} = 0$. Equation (7) then gives $\hat{\lambda}_0 = -\hat{\Sigma}^{-1}\hat{\mu}^*$. You can regard this simplification as the limit case of an intermediary GMM step targeting both the sample averages of X_t and sample averages of expected returns.

Estimation of Φ and λ_1 follows from a second GMM problem, which attempts to match

[1] In our dataset, it accounts for over 90% of forward rates' quadratic variation.

	Level	Slope	Curve	|Eig|	Level	Slope	Curve
	Risk-Neutral Dynamics Φ^*				δ_1		
Level	0.98	-0.03	-0.01	0.82	0.68		
	(0.01)	*(0.01)*	*(0.01)*		*(0.00)*		
Slope	-0.01	0.86	-0.16	0.82	0.40		
	(0.06)	*(0.07)*	*(0.03)*		*(0.00)*		
Curve	0.22	0.14	0.77	0.997	0.15		
	(0.05)	*(0.03)*	*(0.13)*		*(0.00)*		
	Physical Dynamics Φ ($w = 0.01$)				$\Sigma\lambda_1 = \hat{\Phi} - \hat{\Phi}^*$		
Level	0.85	0.07	-0.09	0.68	-0.13	0.10	-0.07
	(0.15)	*(0.11)*	*(0.16)*		*(0.16)*	*(0.10)*	*(0.16)*
Slope	0.11	0.77	-0.28	0.88	0.12	-0.09	-0.12
	(0.07)	*(0.05)*	*(0.07)*		*(0.07)*	*(0.04)*	*(0.07)*
Curve	0.31	0.08	0.77	0.88	0.09	-0.06	-0.01
	(0.12)	*(0.10)*	*(0.18)*		*(0.11)*	*(0.11)*	*(0.14)*
	Physical Dynamics Φ ($w = 0.5$)				$\Sigma\lambda_1 = \hat{\Phi} - \hat{\Phi}^*$		
Level	0.84	0.13	-0.24	0.46	-0.14	0.17	-0.22
	(0.07)	*(0.10)*	*(0.07)*		*(0.07)*	*(0.07)*	*(0.06)*
Slope	0.19	0.56	0.05	0.82	0.20	-0.30	0.21
	(0.07)	*(0.07)*	*(0.09)*		*(0.07)*	*(0.07)*	*(0.09)*
Curve	0.21	0.14	0.67	0.83	-0.01	0.00	-0.11
	(0.09)	*(0.11)*	*(0.16)*		*(0.07)*	*(0.12)*	*(0.10)*
	Physical Dynamics Φ ($w = 1$, OLS)				$\Sigma\lambda_1 = \hat{\Phi} - \hat{\Phi}^*$		
Level	0.87	0.15	-0.14	0.43	-0.11	0.18	-0.12
	(0.06)	*(0.09)*	*(0.06)*		*(0.07)*	*(0.06)*	*(0.04)*
Slope	0.19	0.57	0.08	0.78	0.20	-0.29	0.24
	(0.07)	*(0.07)*	*(0.08)*		*(0.07)*	*(0.07)*	*(0.08)*
Curve	0.16	0.17	0.61	0.83	-0.06	0.03	-0.17
	(0.11)	*(0.08)*	*(0.15)*		*(0.10)*	*(0.08)*	*(0.09)*

Notes. Column |Eig| reports the absolute values of the eigenvalues of each autoregressive matrix. GMM asymptotic standard errors in parenthesis (spectral density matrix estimated using Newey and West (1987) estimator with 24 lags). I use the delta method to compute the the variance of Φ.

Table 2: Affine Model Parameters: GMM Estimation Results

two sets of orthogonality conditions that the e_t shocks in the affine model satisfy.

4.2.1. Matching OLS Moments

One way of estimating Φ is to simply run OLS on (2). We can then reverse-engineer $\lambda_1 = \Sigma^{-1}(\Phi - \Phi^*)$.[1] This is the approach adopted by Adrian et al. (2013). By construction, OLS sets the P^2 parameters in Φ to match the P^2 orthogonality conditions below.

Moment Condition 2 (OLS)**:**

$$E\left[e_t X'_{t-1}\right] = 0. \tag{M2}$$

Moment M2 holds in the affine model from the assumption that e_t is an i.i.d. exogenous random process. We should not be able to forecast it using X_t. M2 asks that we select Φ in such a way that the resulting sample shock series satisfies this property.

However convenient, OLS estimation presents two caveats. First, OLS does not target empirical patterns related to risk premium. For instance, nothing guarantees that the level factor will be the main driver of long-term premia, as empirical pattern 2 calls for. By matching M2 alone, we leave on the table potentially valuable information to discipline parameter selection. Moreover, we give up the model's ability to reproduce these risk-premium patterns. One might be interested in the correct specification of the risk premium by itself, but, more importantly, realistic risk premium dynamics are critical in the exercise of decomposition yields since the term premium is just a sum of expected returns (see (1)).

The second issue is that OLS estimates are highly sensitive to model specification, as Cochrane and Piazzesi (2009) demonstrate. Estimating a version of (2) in difference, for instance, leads to wildly different conditional forecasts of X, and therefore different estimates of the term premium. More deeply, in the case of near unit root processes, OLS estimates lead to a downward bias on the system's persistence (Yamamoto and Kunitomo (1984)). This is particularly problematic in our context because we know the risk-neutral dynamics Φ^*

[1] In that sense, the first-stage estimate of the risk-neutral dynamics imposes no restrictions on the conditional drift of X_t (Joslin et al. (2011) formalize this point).

must have a near-unitary root - thus the physical distribution having a similar property is a plausible possibility.[1]

4.2.2. *Matching Risk Premia Moments*

Another approach to complete the estimation of the affine model is to focus on risk premia. For instance, Cochrane and Piazzesi (2009) estimate λ_1 to approximate their return-forecasting regression on one-period returns, and then compute $\Phi = \Phi^* + \Sigma\lambda_1$. ($\lambda_1$ is directly connected to one-period returns through (6).) In our case, empirical findings 1 and 2 relate to returns over longer investment horizons, so the connection with λ_1 gets blurred. But the principle of using information about risk premia to infer physical Φ from risk-neutral Φ^* continues to apply. All we need is the appropriate moment condition.

Moment Condition 3 (Risk Premium)**:**

$$E\left[\varepsilon^n_{t,t+m} X'_{t-1}\right] = 0. \qquad \text{where } \varepsilon^n_{t,t+m} = rx^n_{t,t+m} - E_t rx^n_{t,t+m} \tag{M3}$$

For a given pair (m, n), M3 contains P moment conditions. In the context of return-forecasting regressions (9), M3 defines OLS estimates. In the context of the affine model, M3 represents an orthogonality condition that, again, holds because e_{t+m} is independent from X_{t-1}. To see this, write down the expression for excess returns:

$$\begin{aligned} rx^n_{t,t+m} &= p^{n-m}_{t+m} - p^n_t + p^m_t \\ &= \text{constant} + \left[B'_{n-m}\Phi^m - B'_n + B'_m\right] X_t + \underbrace{\sum_{j=0}^{m-1} \Phi^j e_{t+m-j}}_{\varepsilon^n_{t,t+m}}. \end{aligned} \tag{14}$$

The sum of shocks on the right-hand side corresponds to the return-forecasting error $\varepsilon^n_{t,t+m}$. Moment M3 therefore asks that we choose Φ^m such that this combination of sample shocks is

[1] Indeed, several bias-correction methods have been proposed to address this issue. See Phillips and Yu (2005), Tang and Chen (2009) and Bauer et al. (2012).

uncorrelated with X_{t-1}.

We need to choose m and n. I focus on replicating the properties of risk premia on a two-year investment horizon, $m = 8$. In the two-year case, return-forecasting regressions combine an elevated degree of predictability (pattern 1) with an nice pattern of risk premium driven mainly by the level of the term structure (pattern 2).[1] With $m > 1$, a key difference between M2 and M3 is that they constrain the drift of X over different time horizons. Moment M2 cares about Φ, the drift in one quarter; M3 cares about Φ^8, the drift in two years. It should now be clear how long-term return predictability can provide critical information for the dynamics of the term structure.

I target the return on bonds with maturities $n = 12, 20, 28$. These are the bond returns we saw in table 1. Selecting three bond maturities means M3 gives us $9 = P^2$ moment conditions, the same as M2.

Lastly, why bother with risk premia over *two-year* investment horizons when we saw that the term premium is a sum of *one-quarter* expected returns? The appendix shows that we can write long-term risk premia as a combination of one-period premia:

$$E_t rx^n_{t,t+m} = \sum_{j=0}^{m-1} E_t rx^{n-j}_{t+j,t+1+j} - \sum_{j=0}^{m-1} E_t rx^{m-j}_{t+j,t+1+j} \tag{15}$$

(The first sum groups the excess returns of re-investing the n-period bond over the next m periods. The second sum groups the difference between the cost of financing in period t and the expected cost of re-financing period-by-period at the prevailing interest rate.[2]) The point: moment M3 *does* convey information about short-term risk premia. Critically, it conveys information about risk premia *beyond the next period* $t + 1$, as (15) contains returns realizing up to $t + m$. This is valuable information to us: the term premium component of the forward rate decomposition (1) also depends on the expectation of one-period returns on horizons

[1] The empirical patterns also justify targeting the three-year horizon, but that would yield little predictability/interpretation gains at the high cost of losing twelve observations from a 169-data-point sample.

[2] The second sum is a really a sum of term premium terms - the difference between current forward rates and expected future interest.

longer than one quarter.

4.2.3. Moment Weighting and Matching

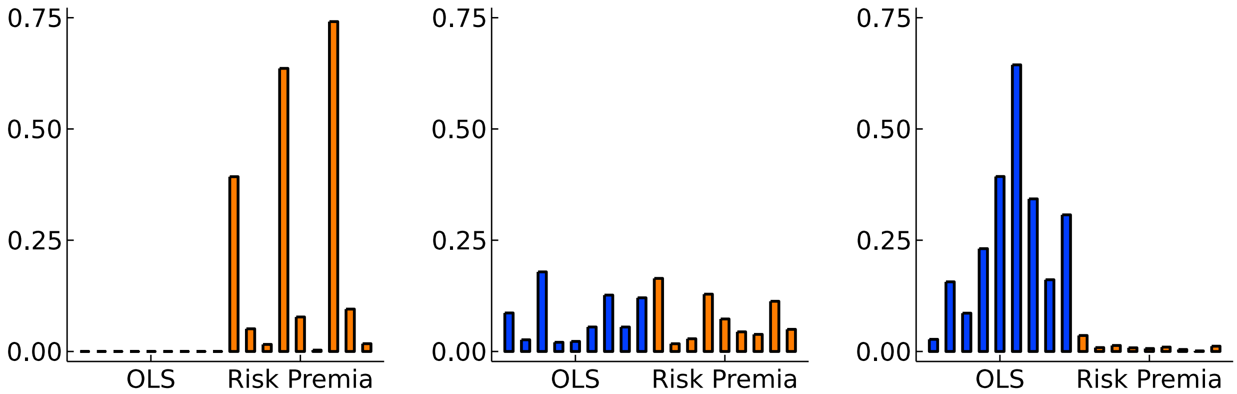
I compute $\hat{\lambda}_1$ as the solution to a second GMM estimation that targets OLS moments M2 and risk-premium moments M3. We therefore search for $P^2 = 9$ parameters to match $2P^2 = 18$ moment conditions. To write moments in comparable standard deviation units, I divide the sample counterpart of M2 by $1\% \times 0.5\%$ and that of M3 by $1\% \times 5\%$. These constants are close to the standard deviations of the underlying terms in each targeted moment. With $\hat{\lambda}_1$ estimated, I compute physical distribution parameter $\hat{\Phi} = \hat{\Phi}^* + \hat{\Sigma}\hat{\lambda}_1$, which completes the estimation of the affine model.

Given the presence of overidentifying restrictions, the choice of the weighting matrix W is critical. One possibility is to use the asymptotically efficient inverse of the spectral density matrix. Nevertheless, the whole point of using GMM for us is to bring to the center stage a prominent trade-off imposed by the affine structure: it can reproduce OLS (M2) *or* it can reproduce return forecasting regressions (M3); it cannot reproduce both. The efficient weighting matrix throws this trade-off under the carpet and simply asks that we pay more attention to combinations of individual moments - regardless to which group M2 or M3 they belong - with the lowest variance. Therefore, while it can be an appropriate choice in other contexts, the efficient weighting matrix defeats one of the purposes of this paper.

So consider instead the following diagonal weighting matrix:

$$W_w = \frac{1}{18} \begin{bmatrix} wI_9 & 0 \\ 0 & (1-w)I_9 \end{bmatrix},$$

where I_9 is the 9×9 identity and $w \in [0,1]$. Weighting matrix W_w attributes a combined weight of w to OLS moments M2, and $1 - w$ to risk-premium moments M3. Figure 5 plots the resulting minimized moments, in absolute value, for three different choices of w (in the language of the general GMM formulation (13), it plots $|g_{\hat{\theta}}|$). Higher bars mean larger



(a) $w = 1$ (OLS) **(b)** $w = 0.5$ **(c)** $w = 0.01$

Notes. All moments in standard deviation units. Blue bars represent OLS moments (M2); orange bars represent risk premia moments (M3). The weighting matrix gives a combined weight of w to the former, $1 - w$ to the latter.

Figure 5: Second-Step GMM: Estimated Moment Errors (Absolute Value)

moment errors. The blue bars correspond to OLS moments M2, the orange bars correspond to the return-forecasting moments M3.

Key Point 1: Under OLS-implied physical dynamics, the share of risk premia variance accounted for by the level factor is substantially lower (as much as half) than what is implied by return-forecasting regressions.

When $w = 1$ (panel 5a), $\hat{\Phi}$ coincides with the OLS estimates of the law of motion (2). Shocks e_t are orthogonal to X_{t-1}, and the affine model matches M2 with zero error. However, the model misses risk premium moments M3 by as much as 70% of a standard deviation.

Missing M3 means it does not explain observed returns as well. More importantly, it means that we get different risk premia. Table 3 shows how OLS estimates of the return-forecasting equation (9) compares to those implied by the affine model. It considers two investment horizons: $m = 8$ (two years, which we target) and $m = 12$ (three years, which we do not). Bond maturities n are the same as before.

The message of the table is clear: *estimating Φ by OLS leads to unrealistic risk premium behavior.* The loadings of long-term expected returns on Level_t are far too small compared to the loadings on the other two factors. Since they capture the variance decomposition of risk premia (equation (11)), these findings mean that OLS-implied premia violate empirical pattern 2. In the case $m = 8$, $n = 12$, the share of variance attributable to Level_t falls from 88% in the regression to 43% in the OLS-estimated affine model. The amount of predictability we leave on the table attributable *specifically* to the level factor is comparatively large. Indeed, the three outstanding moment error bars in panel 5a refer to the three conditions

$$E[\varepsilon^n_{t,t+8}\text{Level}_{t-1}] = 0 \qquad n = 12, 20, 28,$$

showing how one could extract additional information from the level factor to explain observed excess returns. Put differently, in the world we estimate using $w = 1$, returns would not look like what they do. We would not get pretty figure 4, as projecting returns with Level_t alone would lead to wildly different forecasts than using the whole state X_t.

As we lower w, the return-forecasting errors implied by the affine model get increasingly more orthogonal to X_t (lower orange bars in figure 5), but implied shocks e_t get increasingly *less* orthogonal (higher blue bars). Taken together, the three panels underscore the trade-off imposed by the affine structure in terms of empirical objectives. As expected, the case $w = 0.5$ balances these moment errors. When $w = 0.01$, we get very close to matching M3, and thus to reproducing return-forecasting regressions, as table 3 reports.[1] (We should not expect to find exactly the same figures, because the return-forecasting regressions (9) have a free intercept term, which the GMM solver does not - we have fixed μ.) On the other hand, we get farther from OLS dynamics, and the RMSE of sample errors grows.

There is no "right" choice of w. There might be statistical reasons to focus on one group of identifying moments, like OLS's notorious bias in the presence of unit roots (Cochrane and Piazzesi (2009)'s point). There might be numerical reasons, like the convenience of OLS as opposed to non-linear searches (Adrian et al. (2013)'s point). But from a *theoretical* perspective,

[1]Setting $w = 0$ further improves the fitting of M3, but leads to wild short-term dynamics. I therefore focus on the remainder of the paper on the $w = 0.01$ case.

		$E_t rx^n_{t,t+m} = a^{(n)} + b^{(n)} \cdot X_t$					
Investment Horizon (m):			Two Years			Three Years	
Bond Maturity: $n = 12$							
	w	Level_t	Slope_t	Curve_t	Level_t	Slope_t	Curve_t
OLS		2.53	-0.08	0.94			
Affine Model	0.01	2.27	-0.30	0.56		$(rx^{12}_{t,t+12} = 0)$	
Affine Model	0.5	1.67	-0.23	0.39			
Affine Model	1	0.60	-1.65	0.22			
Bond Maturity: $n = 20$							
	w	Level_t	Slope_t	Curve_t	Level_t	Slope_t	Curve_t
OLS		5.75	-0.78	1.92	5.28	-0.05	1.58
Affine Model	0.01	5.62	-1.06	1.35	4.23	-0.42	0.13
Affine Model	0.5	4.98	-0.80	1.07	3.57	-0.54	0.18
Affine Model	1	2.60	-1.72	0.85	2.10	-1.05	0.57
Bond Maturity: $n = 28$							
	w	Level_t	Slope_t	Curve_t	Level_t	Slope_t	Curve_t
OLS		8.54	-1.43	2.68	9.21	-0.56	3.05
Affine Model	0.01	8.35	-1.79	2.11	7.60	-0.96	0.55
Affine Model	0.5	8.15	-1.29	1.77	6.97	-1.02	0.39
Affine Model	1	4.86	-2.55	1.47	4.77	-1.77	1.02

Notes. This table compares the return-forecasting regression (9) estimated by OLS with the estimates implied by the affine model. Parameter w governs the weight we give to OLS moments M2 in detriment of risk-premia moments M3.

Table 3: Return-Forecasting Regression: Regression Evidence vs Affine Model

both M2 and M3 are residual orthogonality conditions that we equally expect to hold in "reality". So why do we have to choose? The root of the trade-off that motivates this paper is model uncertainty. Like any model, the affine framework is a simplification that cannot be expected to reproduce all features from reality. Figure 5 hides the many other moments we are *not* matching. With the "right" model (and a lot of data), we would be able to match both M2 and M3, and the trade-off would vanish. In that sense, selection of w is not so much a matter of right or wrong, as it is a matter of research preference and careful consideration of which aspects of reality we need to capture.

4.3. Estimated Factor Dynamics

We have seen where our estimates come from, we now look at what they tell us. The three bottom panels of table 2 report the physical distribution's autoregressive matrix $\hat{\Phi}$ for the same group of weighting parameters w, which I focus on for the rest of the paper: $w = 0.01$, $w = 0.5$ and $w = 1$ (OLS). Regardless of which we choose, physical dynamics is largely less persistent than in the risk-neutral case: the largest eigenvalue of the system drops from 0.997 to the range 0.83-0.88. The level factor no longer resembles an AR(1).

Figures 6, 7 and 8 summarize model dynamics for different choices of w. They contain IRFs to Level_t, Slope_t and Curve_t shocks, respectively (e.g., the level shock is $[1, 0, 0]$). The top graphs show the responses of model factors X_t.

The graphs in the middle depict the (annualized) forward rate decomposition:

$$tp_t^n = f_t^n - E_t i_{t+n-1}$$

The thick red line represents the response of interest rate over time i_{t+n}. The thin line plots the reaction of the term structure $f_{t=0}^n$ *in period zero when the shock hits*, as a function of maturity n. The difference between them (filled area) is the term premium $tp_{t=0}^n$ at period zero as a function of the bond's maturity n. Since we use the same risk-neutral distribution $(\hat{\mu}^*, \hat{\Phi}^*)$, the time-zero response of the term structure $f^n_{t=0}$ is identical across w's. The choice of w only affects its *decomposition*, not the response itself.

The bottom plot depicts the response of the two-year risk premium on the five-year maturity bond $E_t rx^{20}_{t,t+8}$ (solid line). For comparison, I also plot the risk premium implied by the unrestricted return-forecasting regression (9) (dashed).[1]

Key Point 2: As we estimate physical dynamics Φ giving more weight to risk-premium moments M3 rather than OLS moments M2 (i.e., as we choose lower w):

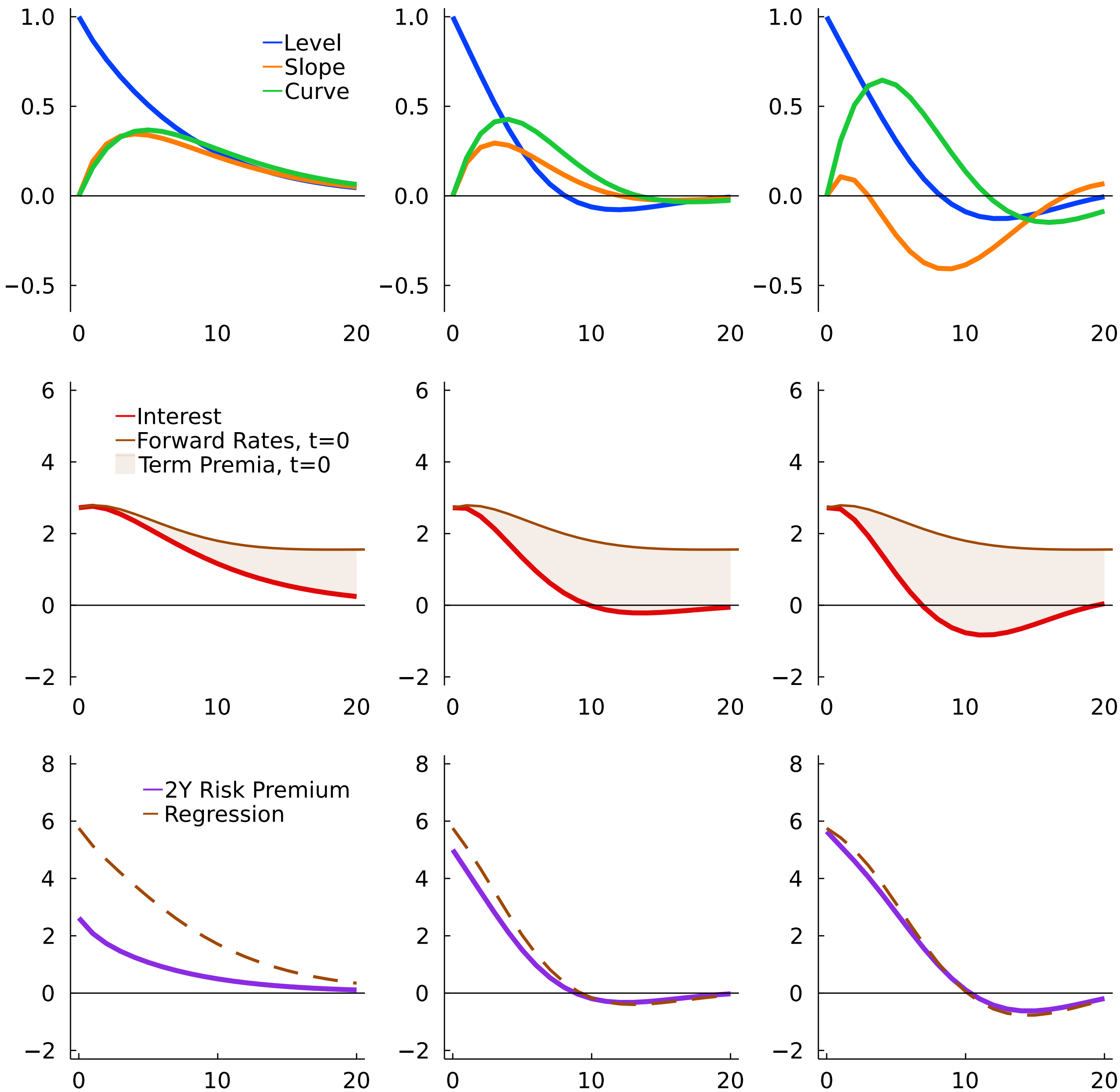
1. Interest rate dynamics presents *higher-frequency* and *amplified* cycles.

2. The model attributes an increasing portion of *level* movements in the term structure to term premia.

The left panel refers to $w = 1$, the dynamics we get from OLS. Comparison with the panels to the right shows that, as we ask the model to reproduce risk premium moments, the IRFs of X_t (top graphs) display higher-frequency and amplified cycles; and interest IRFs (middle graphs) inherit these properties. "Higher-frequency" means they change direction faster: with $w = 0.01$, *all* nine X_t IRFs switch signs at some point of the twenty-period transition, compared with two under OLS $w = 1$. "Amplified" means that the local peaks and bottoms of the IRFs tend to be larger in absolute value. In all, the responses acquire the look of an "S"-shape. On the other hand, in the $w = 1$ OLS case, the path of the variable shocked reminds us of an AR(1), especially in the Level_t and Slope_t shocks.

Higher-frequency and amplified cycles allow the affine model to replicate return-forecasting regressions. As we pick lower w, the risk premium IRFs (bottom graphs) better approximate the path implied by the regressions. From the figures, OLS dynamics $w = 1$ misses the return-forecasting regression by the largest margin after a level shock (panel 6a), as one would expect given key point 1. In that case, the risk premium $E_t rx^{20}_{t,t+8}$ jumps by 2.62 *vs* 5.75 in the return-forecasting regression. With $w = 0.01$, however, the level factor and the interest

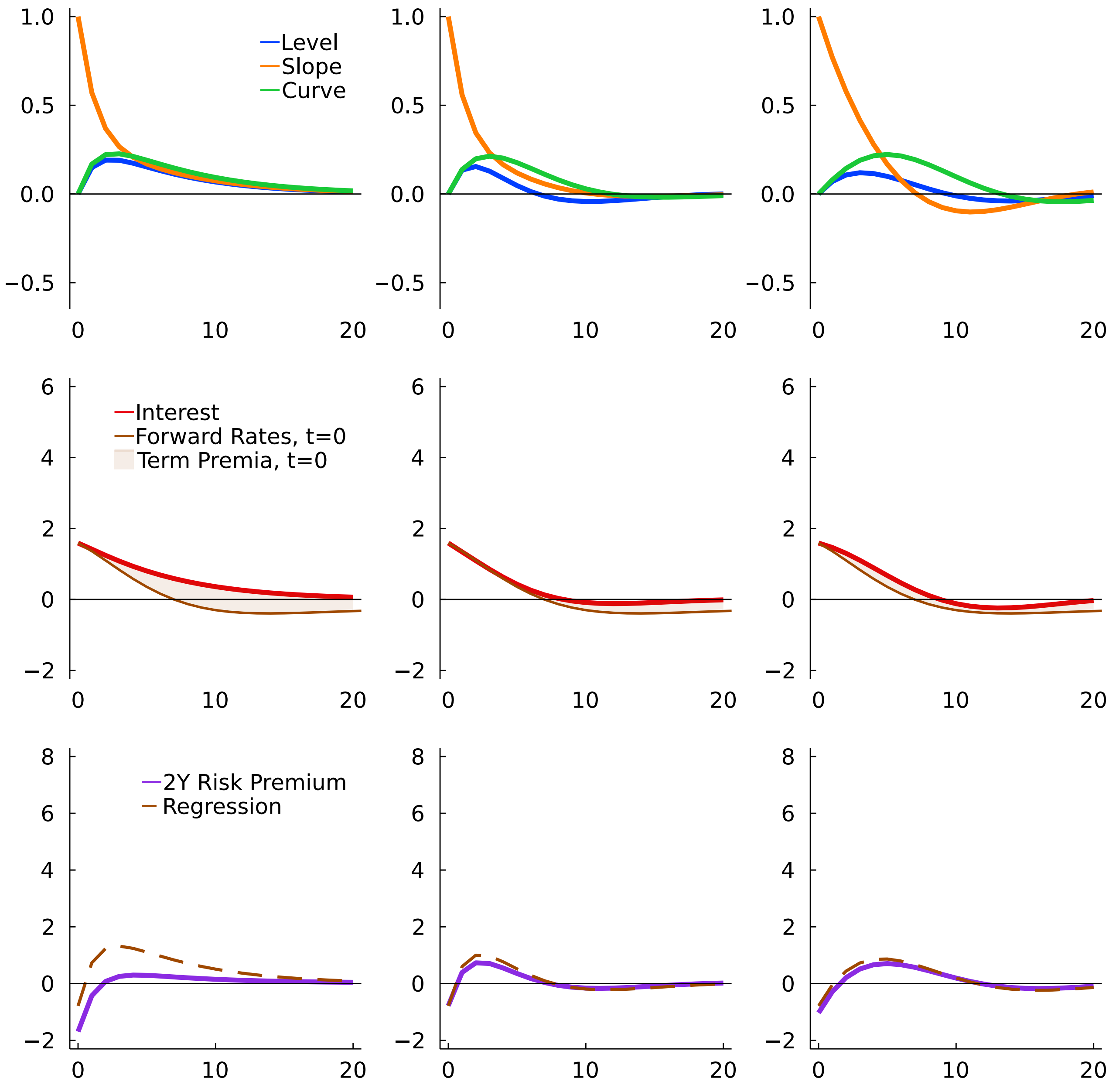
[1] The dashed line multiplies the $b^{(n)}_m$ estimated from (9) by the IRFs of X_t.



(a) $w = 1$ (OLS) **(b)** $w = 0.5$ **(c)** $w = 0.01$

Notes. One period = one quarter. Top: factors X_t. Middle: interest rate i_t, forward rate at period zero $f^n_{t=0}$, and the term premia $tp^n_{t=0}$ (both as functions of maturity n). Bottom: risk premium on 2-year holdings of five-year bonds $E_t rx^{20}_{t,t+8}$ (also, the risk premium implied by the unrestricted return-forecasting regression (9)).

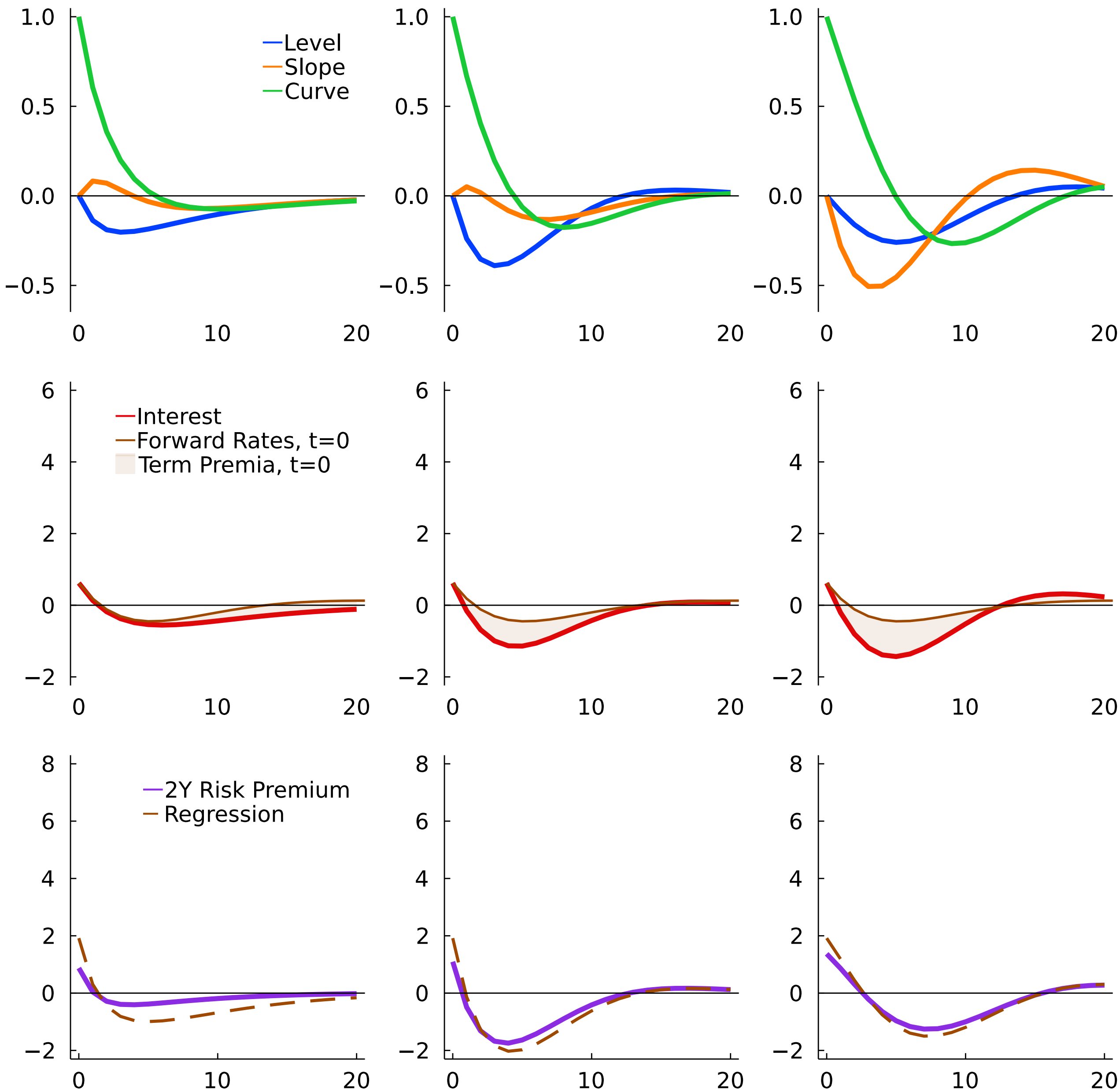
Figure 6: Physical Dynamics: IRF to a Level Factor Shock



(a) $w = 1$ (OLS) **(b)** $w = 0.5$ **(c)** $w = 0.01$

Notes. One period = one quarter. Top: factors X_t. Middle: interest rate i_t, forward rate at period zero $f^n_{t=0}$, and the term premia $tp^n_{t=0}$ (both as functions of maturity n). Bottom: risk premium on 2-year holdings of five-year bonds $E_t rx^{20}_{t,t+8}$ (also, the risk premium implied by the unrestricted return-forecasting regression (9)).

Figure 7: Physical Dynamics: IRF to a Slope Factor Shock



(a) $w = 1$ (OLS) **(b)** $w = 0.5$ **(c)** $w = 0.01$

Notes. One period = one quarter. Top: factors X_t. Middle: interest rate i_t, forward rate at period zero $f^n_{t=0}$, and the term premia $tp^n_{t=0}$ (both as functions of maturity n). Bottom: risk premium on 2-year holdings of five-year bonds $E_t rx^{20}_{t,t+8}$ (also, the risk premium implied by the unrestricted return-forecasting regression (9)).

Figure 8: Physical Dynamics: IRF to a Curvature Factor Shock

rate decline faster (*higher frequency*) and overshoot downward their long-term mean (*amplified cycle*). As a result, bond prices recover faster, which justify higher two-year expected returns in period zero.

Still on the level shock of figure 6, we see a large difference in projected *term premia* in period zero (filled area in the middle plot). Since expected interest declines faster in the $w = 0.01$ case, a larger portion of the initial jump of the term structure must be accounted for by term premia. You can also understand the increase in term premia as the result of agents inferring larger expected returns from the initial jump in the term structure level (equation (1) shows how the term premium is just the sum of quarterly premia).

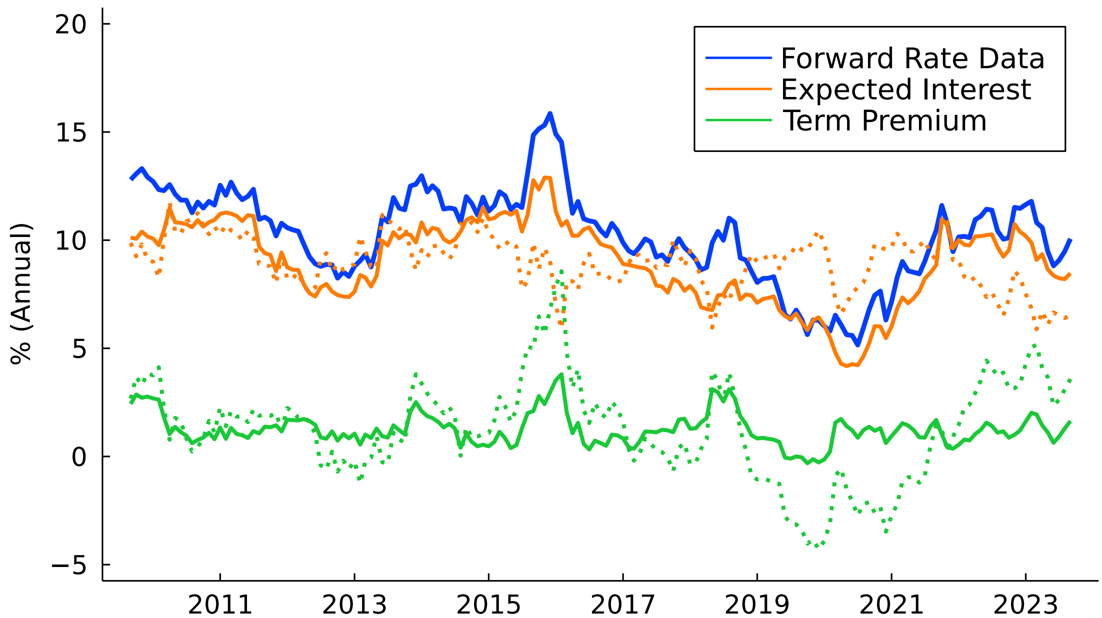
The difference in projected term premia by different choices of w is less clear for the other two shocks of the model. In response to a slope shock, we get very low term premia regardless of w. In response to a curve shock, we get some but not a lot of term premia with $w = 0.01$, and hardly any with OLS $w = 1$. This difference is reminiscent of the positive loading of risk premia on the curve factor (see table 3).

Finally, the IRFs present some "continuity" in weighting w: estimated IRFs with $w = 0.5$ look like an average of the OLS $w = 1$ and $w = 0.01$ cases.

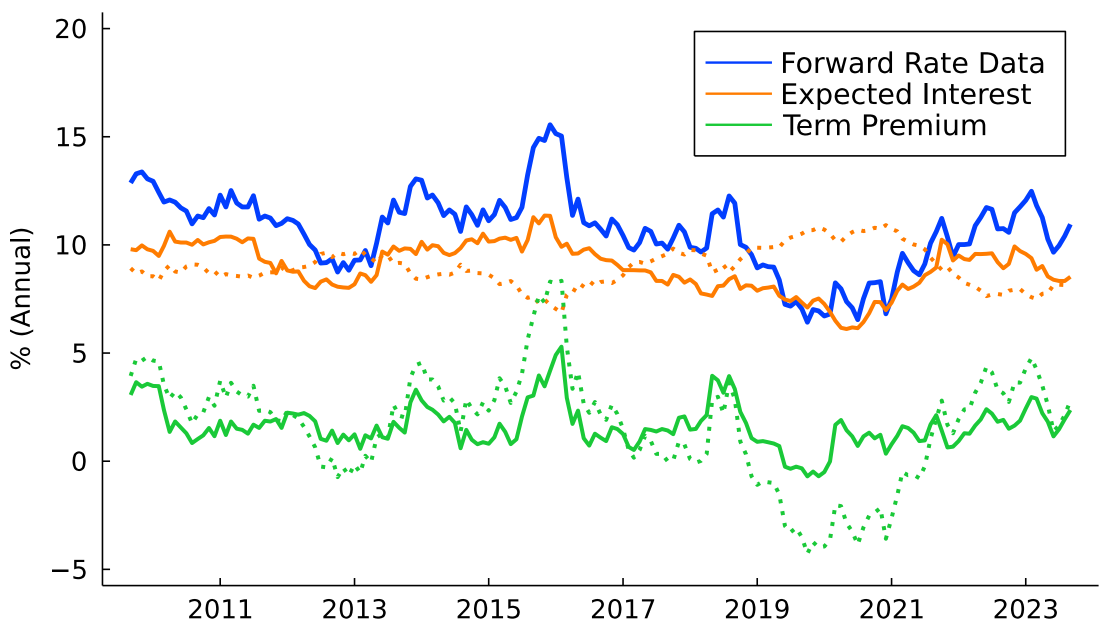
5. Forward Rate Decomposition

Figure 9 plots my estimates of the forward rate decomposition (1). Each plot contains the decomposition implied by the $w = 1$ OLS estimation of Φ (solid curves), and the $w = 0.01$ estimation that focuses on risk premium moments M3 (dashed curves). I ignore the $w = 0.5$ case here to avoid overcrowding the figures (the $w = 0.5$ series is approximately an average of the other two anyway).

The top panel decomposes two-year yields f_t^8; the bottom panel three-year yields f_t^{12}. The decomposition of one-year yields (unplotted) attributes most variation to expected interest, regardless of w. The decomposition of yields with maturities superior to five years attributes



(a) 2-Year Forward Rates ($n = 8$)



(b) 3-Year Forward Rates ($n = 12$)

Notes. Each panel plots the estimates of the forward rate decomposition (1), for maturities $n = 8$ and $n = 12$. I multiply all rates by four to annualize them. Parameter w governs the weight we give to OLS moments M2 in detriment of risk-premia moments M3.

Figure 9: Forward Rate Decomposition. Solid: $w = 1$ (OLS), Dashed: $w = 0.01$

most variation to term premia. The latter finding contrasts with most decompositions for the United States (such as Kim and Wright (2005), Cochrane and Piazzesi (2009), Adrian et al. (2013), Joslin et al. (2014)) which find that a substantial share of variation in *ten-year* yields is accounted by changes in expected interest. Part of the explanation for the difference between results for US and Brazil is that these studies use longer time series of US data, in which the interest rate presents a lot of low frequency dynamics that are absent with Brazilian data starting in 2009. As the term structure plot 1 shows, in the period 2009-2023, the interest rate fluctuates around its averages with cycles of about two years. It is not surprising that our estimates reflect this average cycle length.

The difference between the decompositions estimated with $w = 1$ (OLS) and $w = 0.01$ reflects key point 2 of the previous section. The two term premia series tend to differ more when the *level* of the term structure is unusually high or low (you can find the plot of X_t in the appendix). For example: in 2015/2016, Brazil experiences a financial crisis and the level of the term structure peaks. The two-year forward rate increases by 4%. The OLS estimate of Φ suggests that about half of that jump is explained by higher expected interest in 2017, the other half by the higher risk premia demanded by investors. The estimate of Φ using $w = 0.01$ leads to a model that reads considerably more return predictability from the high level of the term structure in that period. Investors must have been demanding much higher risk premia. Such higher premium however is only consistent if interest rates were expected to *fall* already by 2017, consistent with its IRF in figure 6c. The period 2019/2020 depicts a similar phenomenon, with an opposite sign.

The figures show how replicating the return-forecasting regressions can involve current and future expected interest moving *in opposite directions* at times. The regressions demand that we incorporate to the expected dynamics of the interest rate the ups and down observed in its time series. The decomposition of the three-year bond in panel 9b smooths out some of the noise and make that point clearer.

6. Decomposition of Monetary Policy Surprises

The IRFs of the physical distribution of X_t have shown that level shocks lead to largely different decompositions depending on the weight w we give to OLS moments M2 in detriment of risk premia moments M3. Curve and slope shocks, not so much. These are reduced-form disturbances with no structural interpretation of their own. In this section, I ask the more interesting question of how the choice of w impacts the decomposition of a domestic monetary policy shock.

I identify a monetary shock applying the high frequency approach that has become common in the monetary economics literature (for example: Cochrane and Piazzesi (2002), Hamilton (2008), Gertler and Karadi (2015), Nakamura and Steinsson (2018); Gomes et al. (2023) use high-frequency identification using Brazilian data). I measure monetary surprises by comparing the actual change in the overnight Selic rate (the Brazilian equivalent of the Fed Funds Rate) to the rate change implied by future contract agreements ("DI" futures). DI future contracts are traded in Brazil with high liquidity and in monthly maturities. I collect surprise data directly from Bloomberg.

Bloomberg's surprise series uses the futures' yield implied by the last traded contract at the day of the policy announcement. Monetary policy rate announcements by the Brazilian Central Bank happen after market hours, so market participants are unaware of the policy rate decision until the following day. The key idea behind high-frequency identification is that it surpasses the problem of endogeneity (the central bank might react to market yields just as market yields obviously react to the central bank; hence we cannot build a surprise series using market prices traded one month prior to the announcement.) The central bank however does not respond to market price changes that take place *on* the announcement day or the next.

Given this procedure, the resulting surprise figures cannot be directly compared to shocks of a *quarterly* model, since agents learn valuable economic information from the start of a

Dependent	η	p-value	R^2
Level	-0.03	0.76	-6.6
Slope	0.72	0.00	7.6
Curve	-0.23	0.58	0.8

Notes. The dependent variables are the changes in model factors ΔX_d following policy announcements. The explanatory variable is the surprise component of the change in Selic rates. The regressions have no intercept term.

Table 4: Policy Rate Announcement Regressions

quarter until the day of the announcement. When working with quarterly or monthly VARs, it is then common practice to add up the individual shocks in a given quarter. Our VAR, however, is formed by combinations of market-price yields. As such, we *can* observe them daily and directly measure them in "high frequency".[1]

Given an announcement date d, let s_d be the corresponding surprise, that is, the actual change in the Selic rate minus the change implied by DI future market. The average annualized absolute surprise is 0.0585%. I collect data on the term structure of forward rates at the day of the announcement, at the end of the following business day, and then compute the same principal component transformation used to build the original sample X_t. This gives two points: X_d and $X_{d+1\text{ day}}$. The surprise in the state X associated with the announcement is the change between these two points:

$$\Delta X_d = X_{d+1\text{ day}} - X_d.$$

Hence, ΔX_d represents the change in level, slope, and curvature of the term structure around day-s announcement. Some announcement dates coincide with policy rate announcements by the US Federal Reserve (they are both on Wednesdays), a strong confounder. I remove these announcements from the sample, which leaves 85 data points.

Next, I separately regress each factor's surprise series on policy rate surprises s_d, without

[1] In fact, most papers measure monetary shocks in terms of their effect on the term structure - see for instance the discussion of policy indicator *vs* policy instrument in Gertler and Karadi (2015).

a constant term:

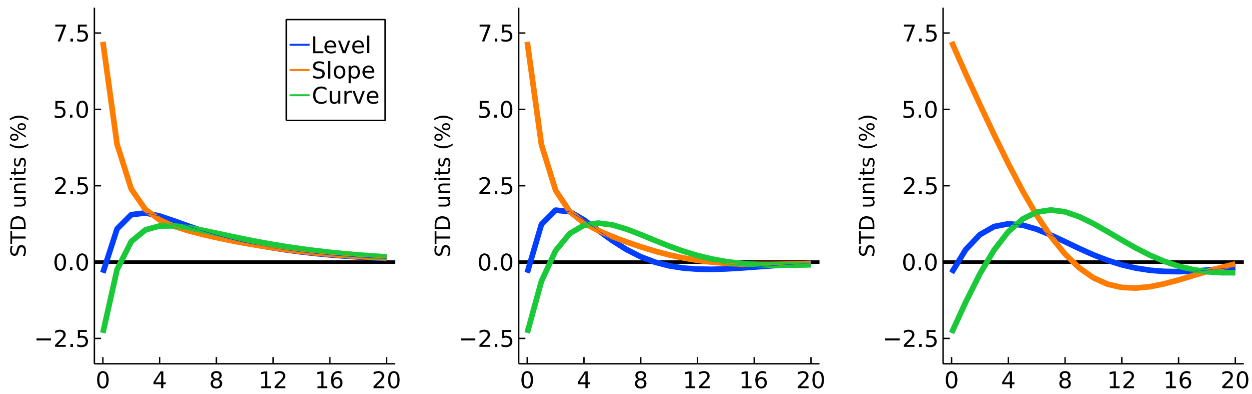
$$\Delta X_d = \eta \, s_d + \epsilon_d.$$

The estimated $\hat{\eta}$ vector determines the average effect of annualized policy surprises on the level, slope and curve of the term structure, given the announcement surprise. Table 4 reports results. It shows that monetary surprises typically affect the *slope* of the term structure. A 0.1% surprise corresponds to a 0.072% increase in the slope factor, or 7.2% of its sample standard deviation. The effects on the level and curve factors are negative, but imprecisely estimated. The projected effect on the quarterly interest rate according to the affine model is $4\,\delta_1 \cdot \hat{\eta} \times 0.1\% \approx 0.091\%$, which is not too far from the 0.1% surprise on the Selic rate.

Figure 10 plots the IRF of state X_t to the shock $e_{t=0} = \hat{\eta} \times 0.1\%$, a positive 0.1% surprise in the central bank announcement, which I henceforth call a "monetary shock". The period of the shock is period zero $t = 0$. I report the responses in standard deviation units to facilitate interpretation in terms of term structure shapes (figure 2). For example: in the quarter of the announcement, the slope of the term structure is 7.2% of one sample standard deviation higher than usual.

When we calculate the IRF using the physical dynamics estimated by OLS $w = 1$, the expected effect of a positive monetary shock is a protracted period of elevated term structure *slope*. The level of the term structure grows, but not as much, and not in the period of the announcement. The impact on the curvature is negative in the first two quarters, turns positive, and then slowly fades along with the other two factors. The positive effect of the monetary policy shock on the three factors is not surprising, since I normalize the principal components that build X_t to be increasing in the interest rate. Such positive effect takes a long time to fade: after the first two quarters, agents expect the three factors to stay above average for the five years described by the plots.

Qualitatively, the shape of the IRF is similar when $w = 0.5$ or $w = 0.01$. But in accordance with the higher-frequency property of the physical dynamics underscored by key point 2, agents expect the three factors to revert the positive impact of the monetary shock more



(a) $w = 1$ (OLS) **(b)** $w = 0.5$ **(c)** $w = 0.01$

Notes. One period = one quarter. The monetary shock corresponds to a 0.1% (annualized) surprise (the average absolute surprise on policy rate announcements is 0.0585%). I divide the factors by 1% to put them in standard deviation units.

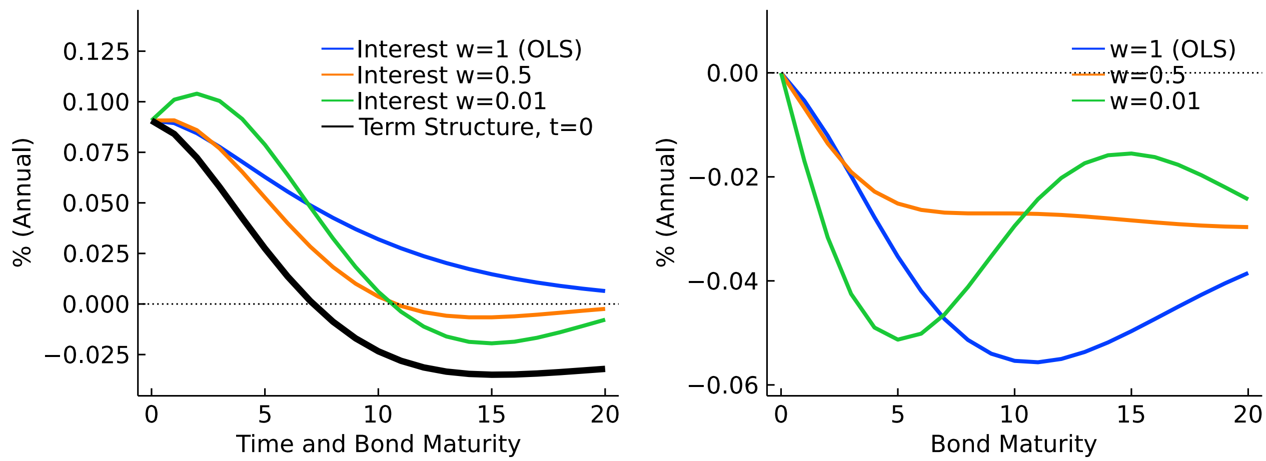
Figure 10: Physical Dynamics: IRF to a Monetary Policy Shock

quickly. Additionally, the IRF cycles show more amplitude. The curve factor peaks at a higher level, and, in the $w = 0.01$ case, the impact on the slope factor becomes negative after only two years of transition. These effects combine to give the "S"-shape look to the IRFs.

How do we translate these level, slope and curve effects into expected interest and risk premia effects?

Key Point 3: In response to a positive monetary policy shock:

1. The term structure of forward rates becomes more negatively inclined; the effect on its level is positive, but not as large and not in the announcement period.
2. Two-year risk premia decline.
3. In the announcement period, term premia embedded in the term structure decline.
4. As we estimate physical dynamics giving more weight to risk-premium moments M3, the positive effect of the shock on expected interest decays faster. For $w = 0.5$ and $w = 0.01$, it lasts ten quarters.



(a) Interest Rate and Term Structure

(b) Term Premia at Shock Period $t = 0$

Notes. One period = one quarter. The monetary shock corresponds to a 0.1% (annualized) surprise (the average absolute surprise on policy rate announcements is 0.0585%). The left figure plots the IRF of interest rates (as a function of time) and the response of the term structure $f_{t=0}^{n+1}$ at the shock period $t = 0$. The difference between the two is the term premia at period $t = 0$, plotted on the right. The vertical axis reports annualized percentage units.

Figure 11: Interest and Term Premia: Response to a Monetary Policy Shock

By the signs of the model-implied loadings of the return-forecasting regressions (table 3), the effects of the monetary shock $\hat{\eta}$ on all three factors indicate a *decline* in long-term risk premia. Indeed, the risk premium on a two-year holding of the five-year maturity bond $E_t rx_{t,t+8}^{20}$ decreases in period zero. It falls by -0.15%, -0.10% and -0.127% in the cases $w = 1$, $w = 0.5$ and $w = 0.01$, respectively. As expected, the case $w = 0.01$ best approximates the -0.121% effect implied by our return-forecasting regression (9).

Figure 11 translates the state vector responses into a decomposition of the whole term structure

$$f_t^n = E_t i_{t+n-1} + tp_t^n$$

in the shock period $t = 0$. The dark thick curve in panel 11a depicts the effect of the monetary shock on $f_{t=0}^{n+1}$ as a function of maturity n. The term structure becomes more negatively inclined, consistent with the rise in Slope$_t$. Short-term yields increase relative to long-term ones. The other three curves correspond to the IRF of the interest rates as a function of time

$t + n$, or the change in its expected path $E_t i_{t+n}$. Regardless of w, agents expect the interest rate to decline, but at a slower pace than that implied by the inclination of the term structure in period zero. Consequently, the period-zero term premium $tp^n_{t=0}$ (panel 11b) is negative, in all w cases. A negative term premium is consistent with the decline in long-term risk premia, and indicates that quarterly risk premia fall as well.[1]

The pace at which interest rate falls differs depending on w. In the $w = 1$ OLS model, the i_t IRF resembles an AR(1) innovation: the initial positive effect slowly decays towards zero. On the other hand, with $w = 0.5$ or $w = 0.01$, the effect of the monetary shock on interest turns negative after exactly 10 quarters. The IRFs increasingly inherit the "S"-shape look from the states' IRFs. With $w = 0.01$, the affine model even suggests an expected further *increase* of the interest rate in the first three quarters after the announcement. These models suggest a much more temporary effect of monetary policy shocks. In terms of the decomposition of the term structure at period zero, they agree with OLS on the *sign* of term premia innovations, but the shapes are somewhat distinct, reflecting different risk premia and of course different expected paths of the interest rate. In conclusion: the choice of which set of moments we give more weight to in the GMM estimation - M2 or M3 - again affects the results we derive from the affine model.

7. Conclusion

This paper studies the joint distribution of Brazilian interest yields and bond holding returns using a multifactor version of the Gaussian affine model of the term structure (Dai and Singleton (2000)). The affine model imposes restrictions that lead to a trade-off involving which aspects of the model we can match to the evidence. I propose a new GMM procedure tailored to deal with this trade-off. I focus on OLS and risk premium orthogonality conditions, which have both been used in previous research to estimate the affine model, often leading to

[1]Equation (1) expresses the term premium as a function of one-period risk premia.

conflicting conclusions. The GMM approach I present in this paper unifies these approaches and allows us to analyze intermediary cases.

To estimate the model, I have shown an important empirical pattern of bond risk premia in Brazil, not previously documented. At long investment horizons, such as two and three years, expected excess returns have been driven by changes in the level of the terms structure far more than its slope and curve. Straightforward OLS estimation of the physical dynamics of the model fails to generate this pattern, which might not be ideal in many contexts, such as the decomposition of the term structure. Indeed, I find that which set of orthogonality conditions we privilege in the estimation leads to different conclusions about the decomposition of the term structure and the effects of monetary policy shocks.

References

Adrian, T., Crump, R. K., and Moench, E. (2013). Pricing the term structure with linear regressions. *Journal of Financial Economics*, 110(1):110–138.

Akgiray, V., Baronyan, S., Sener, E., and Yılmaz, O. (2016). Predictability of Emerging Market Local Currency Bond Risk Premia. *Emerging Markets Finance and Trade*, 52(7):1627–1646.

Almeida, C., Ardison, K., and Kubudi, D. (2015). Approximating Risk Premium on a Parametric Arbitrage-Free Term Structure Model. SSRN Scholarly Paper ID 2535766, Social Science Research Network, Rochester, NY.

Almeida, C. and Faria, A. (2014). Forecasting the Brazilian Term Structure Using Macroeconomic Factors. *Brazilian Review of Econometrics*, 34(1):45–77.

Andreasen, M. M., Engsted, T., Møller, S. V., and Sander, M. (2021). The Yield Spread and Bond Return Predictability in Expansions and Recessions. *The Review of Financial Studies*, 34(6):2773–2812.

Ang, A., Dong, S., and Piazzesi, M. (2007). No-Arbitrage Taylor Rules. Working Paper 13448, National Bureau of Economic Research.

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

Baker, M., Greenwood, R., and Wurgler, J. (2003). The maturity of debt issues and predictable variation in bond returns. *Journal of Financial Economics*, 70(2):261–291.

Bauer, M. D., Rudebusch, G. D., and Wu, J. C. (2012). Correcting Estimation Bias in Dynamic Term Structure Models. *Journal of Business & Economic Statistics*, 30(3):454–467.

Bjork, T. and Christensen, B. J. (1999). Interest Rate Dynamics and Consistent Forward Rate Curves. *Mathematical Finance*, 9(4):323–348.

Boudoukh, J., Richardson, M., and Whitelaw, R. F. (2008). The Myth of Long-Horizon Predictability. *Review of Financial Studies*, 21(4):1577–1605.

Cochrane, J. and Piazzesi, M. (2009). Decomposing the yield curve. *AFA 2010 Atlanta Meetings Paper*.

Cochrane, J. H. (2008). The Dog That Did Not Bark: A Defense of Return Predictability. *The Review of Financial Studies*, 21(4):1533–1575. Publisher: [Oxford University Press, Society for Financial Studies].

Cochrane, J. H. and Piazzesi, M. (2002). The Fed and Interest Rates—A High-Frequency Identification. *American Economic Review*, 92(2):90–95.

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

Coroneo, L., Nyholm, K., and Vidova-Koleva, R. (2011). How arbitrage-free is the Nelson–Siegel model? *Journal of Empirical Finance*, 18(3):393–407.

Da Costa Filho, A. E. (2021). Identification of monetary shocks through the yield curve: Evidence for Brazil. *Brazilian Review of Finance*, 19(1):24–51.

Dai, Q. and Singleton, K. J. (2000). Specification Analysis of Affine Term Structure Models. *The Journal of Finance*, 55(5):1943–1978.

Devpura, N., Narayan, P. K., and Sharma, S. S. (2021). Bond return predictability: Evidence from 25 OECD countries. *Journal of International Financial Markets, Institutions and Money*, 75:101301.

Duffie, D. and Kan, R. (1996). A Yield-Factor Model of Interest Rates. *Mathematical Finance*, 6(4):379–406.

Fama, E. and Bliss, R. (1987). The Information in Long-Maturity Forward Rates. *The American Economic Review*, 77(4):680–692.

Fama, E. F. (1984). The information in the term structure. *Journal of Financial Economics*, 13(4):509–528.

Fama, E. F. and French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, 22(1):3–25.

Faria, A. and Almeida, C. (2018). A hybrid spline-based parametric model for the yield curve. *Journal of Economic Dynamics and Control*, 86:72–94.

Farmer, L. E., Schmidt, L., and Timmermann, A. (2023). Pockets of Predictability. *The Journal of Finance*, 78(3):1279–1341.

Gertler, M. and Karadi, P. (2015). Monetary Policy Surprises, Credit Costs, and Economic Activity. *American Economic Journal: Macroeconomics*, 7(1):44–76.

Golez, B. and Koudijs, P. (2018). Four centuries of return predictability. *Journal of Financial Economics*, 127(2):248–263.

Gomes, D., Iachan, F., Ruhe, A. P., and Santos, C. (2023). Monetary Policy and Labor Markets in a Developing Economy. *SSRN Electronic Journal*.

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

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

Hamilton, J. D. (2008). Daily monetary policy shocks and new home sales. *Journal of Monetary Economics*, 55(7):1171–1190.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50(4):1029.

Joslin, S., Priebsch, M., and Singleton, K. J. (2014). Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks. *The Journal of Finance*, 69(3):1197–1233.

Joslin, S., Singleton, K. J., and Zhu, H. (2011). A New Perspective on Gaussian Dynamic Term Structure Models. *Review of Financial Studies*, 24(3):926–970.

Kim, D. H. and Wright, J. H. (2005). An Arbitrage-Free Three-Factor Term Structure Model and the Recent Behavior of Long-Term Yields and Distant-Horizon Forward Rates. *SSRN Electronic Journal*.

Kirby, C. (1997). Measuring the Predictable Variation in Stock and Bond Returns. *Review of Financial Studies*, 10(3):579–630.

Knez, P. J., Litterman, R., and Scheinkman, J. (1994). Explorations Into Factors Explaining Money Market Returns. *The Journal of Finance*, 49(5):1861–1882.

Litterman, R. and Scheinkman, J. (1991). Common factors affecting bond returns. *Journal of Fixed Income*, 1(1):54–61.

Nakamura, E. and Steinsson, J. (2018). High-Frequency Identification of Monetary Non-Neutrality: The Information Effect*. *The Quarterly Journal of Economics*, 133(3):1283–1330.

Newey, W. K. and West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3):703.

Osmani, G. and Tabak, B. (2008). Characterizing the Brazilian term structure of interest rates. *International Journal of Monetary Economics and Finance*, 2(2):103–114.

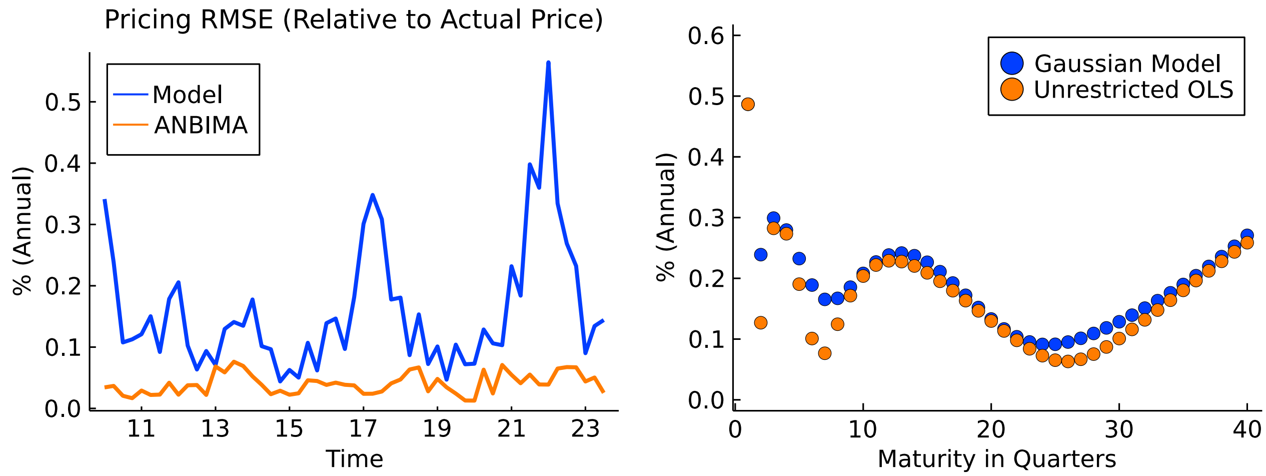
Phillips, P. C. B. and Yu, J. (2005). Jackknifing Bond Option Prices. *The Review of Economic Studies*, 18(2):707–742.

Svenson, L. (1994). Estimating and interpreting forward interest rates: Sweden 1992-4. *IMF Working Paper*, 94/114.

Tabak, B. M. and Andrade, S. C. d. (2003). Testing the Expectations Hypothesis in the Brazilian Term Structure of Interest Rates. *Brazilian Review of Finance*, 1(1):19–43.

Tang, C. Y. and Chen, S. X. (2009). Parameter estimation and bias correction for diffusion processes. *Journal of Econometrics*, 149(1):65–81.

Yamamoto, T. and Kunitomo, N. (1984). Asymptotic bias of the least squares estimator for multivariate autoregressive models. *Annals of the Institute of Statistical Mathematics*, 36(3):419–430.



(a) ... relative to actual prices (yields) **(b)** ... relative to OLS estimates (forward rates)

Notes. Panel 12a: RMSE against actual zero-coupon yields. Panel 12b: RMSE of forward rates in the Gaussian and OLS models. All errors multiplied by four to annualize.

Figure 12: Root Mean Squared Pricing Error...

A. Pricing Errors

A.1. Actual Prices

I compare ANBIMA data and the estimated Gaussian model's yields to actual price data. I use LTN (*Letra Financeira do Tesouro*) yields. These are zero-coupon bonds issued by the Brazilian government that always expire in the first business day of the months of January, April, July and October. The set of available bonds does not provide a complete spam of maturities, even if we consider only the months above and existence of coupon-paying ones. This is why I use ANBIMA's estimated term structure.

Given the quarterly model, I restrict the data sample to the first days of January, April, July and October (or the next business day). That guarantees that the maturities of outstanding bonds in any given time are a round number of quarters. For instance, on April 1st, a bond maturing in July 1st of the same year has a maturity of one quarter. On May 15th, that same bond would have a maturity of half a quarter, which I cannot price using the Gaussian model. Therefore I drop these observations.

The yield of the n-maturity bond is $y_{n,t} = -p_{n,t}/n$. Figure 12a plots the time series of RMSE using the ANBIMA data (orange) and Gaussian model (blue). Each point is the average across maturities. The averages across time and maturities are respectively 0.044% and 0.21% (both rates annualized). The difference between these figures can be interpreted as the constraints imposed by no-arbitrage - the Svenson model is not guaranteed not to produce arbitrage opportunities (Bjork and Christensen (1999), Coroneo et al. (2011)).

A.2. OLS Model

I estimate the linear model

$$f_{n,t} = a_n + b_n' X_t + \eta_t$$

for each maturity n separately by unrestricted ordinary least squares (OLS). Unrestricted OLS errors provide a natural benchmark to linear models, as by construction OLS minimizes squared errors. Figure 12b plots RMSE by maturity (that is, for each separate regression) in orange. The blue markers correspond to the average error produced by the Gaussian model. They correspond to the sample moment vector from the first-stage GMM moments conditions (M1) (the g in the general formulation (13)). The average error (across time and maturities) is: 0.185% in the OLS regression and 0.202% in the Gaussian model. Both of these are annualized. Figure (15) compares, in the same plot, forward rate data and model implied forward rates for four different maturities.

B. Derivation of the Gaussian Affine Model

This section presents the details of the Gaussian exponential-affine model. The vector of factors X_t evolves according to the law of motion (2) in the main text

$$X_t = \mu + \Phi X_{t-1} + e_t \qquad e_t \sim N(0, \Sigma).$$

The Gaussian shocks e_t give name to the model.

The one-period payoff of a bond with maturity n is the price of the same bond in the following period, when it turns into $n-1$ period bond. Bonds with maturity zero pay one unit of currency. Given a stochastic discount factor M_{t+1} for payoffs denominated in currency units, the price of the zero-coupon bond is given by

$$P_t^n = E_t M_{t+1} P_{t+1}^{n-1}, \qquad P_{t+1}^0 = 1$$

or, taking logs and defining $m_{t+1} = \log M_{t+1}$,

$$p_t^n = \log E_t \exp\left\{m_{t+1} + p_{t+1}^{n-1}\right\}, \qquad p_{t+1}^0 = 0.$$

Hence, a pricing theory amounts to picking m_{t+1}. We can re-write the pricing condition above in terms of excess returns:

$$0 = \log E_t \exp\left\{m_{t+1} + i_t + rx_{t,t+1}^n\right\}, \qquad rx_{t+1}^1 = 0, \tag{B.1}$$

where $i_t = f_t^1$, as defined in the main text.

Equation (B.1) holds for a general stochastic discount factor m_{t+1}. The m_{t+1} used by the Gaussian model is given by (3). Heteroskedasticity of m_{t+1} allows for time-varying prices of risk, which is precisely what we need to reproduce time-varying risk premia as observed in the data. Having the innovations to the discount factor come from a e_{t+1} term rather than X_{t+1} is necessary to generate an affine solution.

When applying our particular choice of a discount factor to (3), we get the linear solution to i_t given by (4) and

$$E_t rx_{t,t+1}^n = -(1/2)\mathrm{var}_t(rx_{t,t+1}^n) + \mathrm{cov}_t(rx_{t,t+1}^n, e_{t+1}')\lambda_t. \tag{B.2}$$

The covariance term represents the quantity of risk of each bond, or its "beta", which is why I refer to λ_t as the market price of risk.[1] Given the solution below, the conditional variance and covariance terms in (B.2) are both time invariant, and we can thus drop the t subscripts. All variation in the risk premium $E_t rx^n_{t,t+1}$ comes from λ_t.

Pricing condition (B.2) implies a solution for equilibrium prices linear in the factors, given by (5). Coefficients A_n and B_n satisfy the well-known Riccati equations

$$\begin{aligned} A_{n-1} - A_n + B'_{n-1}\mu - \delta_0 &= -\frac{1}{2}B'_{n-1}\Sigma B_{n-1} + B'_{n-1}\Sigma\lambda_0 \\ B'_{n-1}\Phi - B'_n - \delta'_1 &= B'_{n-1}\Sigma\lambda_1 \end{aligned} \tag{B.3}$$

which one can solve recursively, starting from $A_1 = -\delta_0$ and $B_1 = -\delta_1$. The terms on the left-hand side of the (B.3) correspond to the constant and time-varying parts of $E_t rx^n_{t,t+1}$. The terms on the right-hand side correspond to the covariance term $B_{n-1}\Sigma$ multiplied by λ_t. The main text defines $v_n = -\frac{1}{2}B'_{n-1}\Sigma B_{n-1}$.

If we subtract the right-hand side from both sides of the expressions in (B.3) and gather terms, we arrive at

$$\begin{aligned} A_{n-1} - A_n + B'_{n-1}(\mu - \Sigma\lambda_0) - \delta_0 &= v_n \\ B'_{n-1}(\Phi - \Sigma\lambda_1) - B'_n - \delta'_1 &= 0. \end{aligned} \tag{B.4}$$

Hence, we can find the coefficients A_n and B_n without regard to market price of risk parameters λ_0 and λ_1, *or* the physical distribution μ, Φ, by searching instead for the terms $\mu^* = \mu - \Sigma\lambda_0$ $\Phi^* = \Phi - \Sigma\lambda_1$ directly. Furthermore, if λ_t were zero, and thus we had a conditionally deterministic discount factor, the bond prices would be observationally equivalent to those in the model with risk, *if the drift of X_t was determined by parameters μ^* and Φ^**, instead of μ and Φ. This is why I refer to distribution (7), which performs that change of drift, as

[1] In the beta representation traditionally used in the empirical finance literature, an asset's beta is usually defined as the linear projection coefficient on the corresponding risk factor. Given the homoscedasticity of the innovations in the model, the coefficients emerge by simply left-multiplying λ_t by $\Sigma^{-1}\Sigma$. Then, $\text{cov}(rx^n, e')\Sigma^{-1} = \text{cov}(rx^n, e')E(ee')^{-1}$ becomes the "usual" beta, and $\Sigma\lambda_t$ becomes the market price of risk.

"risk-neutral".

Since log bond prices are linear in X_t, forward rates will also be linear, with $f_t^n = A_n^f + B_n^{f\prime} X_t$. The relationship between forward rates and prices $f_t^n = p_t^{n-1} - p_t^n$ implies $A_n^f = A_{n-1} - A_n$ and $B_n^f = B_{n-1} - B_n$. When $n = 1$, $A_1^f = \delta_0$ and $B_1^f = \delta_1$.

The main text claims that, under the risk-neutral distribution, forward rates coincide with expected interest (plus the Jensen term v_n). We can prove that with sheer force, using the risk-neutral version of the Riccati equations (B.4):

$$\begin{aligned}
f_t^n &= A_n^f + B_n^{f\prime} X_t \\
&= v_n + \delta_0 - B_{n-1}^{\prime} \mu^* + \delta_1^{\prime} \Phi^{*n-1} X_t \\
&= v_n + \delta_0 + \delta_1^{\prime} \left[(I + \Phi^* + \cdots + \Phi^{*n-1}) \mu^* + \Phi^{*n-1} X_t \right] \\
&= v_n + \delta_0 + \delta_1^{\prime} E_t^* X_{t+n-1} \\
&= v_n + E_t^* i_{t+n-1}.
\end{aligned}$$

A more elegant way to prove that equality is to write $E_t^* rx_{t,t+1}^n = v_t$, which is the real meaning of (B.4). Replacing the definition of excess returns gives

$$p_t^n = E_t^* p_{t+1}^{n-1} - [i_t + v_n] = -\sum_{j=0}^{n} E_t^* \left[i_{t+j} + v_{n-j} \right]$$

(the last equality follows from the fact that p_t^n converges to zero in n). Finally, $f_t^n = p_t^{n-1} - p_t^n$ gives $f_t^n = v_n + E_t^* i_{t+n-1}$.

B.1. Interpretation of the Risk-Neutral Distribution

How do we interpret the different behavior of the state variables under the actual and risk-neutral measures? In most economic models, "risk-neutral" measure is another name to probability times marginal utility. By factoring in marginal utility, one can do asset pricing by simply taking expectations of uncertain payoffs.

More formally, we can price a payoff $Y(e_{t+1})$ using the stochastic discount factor directly: $E_t\left[e^{m_{t+1}}Y(e_{t+1})\right]$. Alternatively, as I show below, we can price it by computing the expected payoff and discounting using the interest rate: $e^{-i_t}E_t^*\left[Y(e_{t+1})\right]$. But E_t^* changes the conditional distribution of the shock vector e_{t+1} by centering it around $-\Sigma\lambda_t$ (instead of zero). The re-definition of μ and Ψ in (7) leads to this change of conditional mean. By the pricing condition (6), the inequality $\lambda_{t,p} < 0$ means that $e_{t+1,p}$ is a "bad" shock: investors pay a premium for payoffs that provide insurance by correlating positively with it. Correspondingly, if the p-th element of $-\Sigma\lambda_t$ is positive, the risk-neutral distribution centers $e_{t+1,p}$ around a positive value.[1] In conclusion, we can price payoff $Y(e_{t+1})$ by simply computing its expected value if we make "bad" occurrences of e_{t+1} "more likely" under the probability density applied by the expectation operator. The altered law of motion for model factors (7) accomplishes that.

The exponential-affine model offers a lot of tractability to see the point algebraically. In a given period $t+1$, both the actual shock e_{t+1} and the artificial risk-neutral "shock" e^*_{t+1} are Gaussian with the same covariance matrix Σ. Their probability density functions therefore have the same normalizing constant $\mathcal{M}$; only their exponential kernel differ. Hence, we can factor in the discount factor m_{t+1} by changing such exponential kernel and put into evidence the density function of the risk-neutral measure:

$$\begin{aligned} 1 = E_t\left[e^{m_{t+1}}Y_{t+1}(e_{t+1})\right] &= \mathcal{M}\int \exp\{-(1/2)e'\Sigma^{-1}e\}\exp\{-i_t - \lambda_t'\Sigma\lambda_t - \lambda_t e\}Y_{t+1}(e)de \\ &= \mathcal{M}\int \exp\{-(1/2)(e-\phi_t)'\Sigma^{-1}(e-\phi_t)\}\exp\{-i_t\}Y_{t+1}(e)de. \end{aligned}$$

So, risk-neutral pricing here simply amounts to taking expected value of returns "as if" shocks e_{t+1} had a conditional mean equal to ϕ_t. The last equality then implies

$$\phi_t = -\Sigma\lambda_t,$$

and all we need now is to interpret ϕ_t. To that end, suppose period $t+1$ comes and we

[1] We multiply λ_t by Σ to account for shock correlation.

observe $e_{t+1} = \phi_t$. State variables will then be given by

$$X_{t+1} = \Phi X_t + \phi_t = \Phi X_t - \Sigma\lambda_t = \mu^* + \Phi^* X_t.$$

We have thus a link between the estimated risk-neutral dynamics - to which the IRFs above give a useful interpretation - and the discount factor m. What the equation above tells us is that *the discount factor, or marginal utility, places elevated weight on combinations of shocks ($\approx \phi_t$) that lead our states X to present the risk-neutral dynamics (μ^*, Φ^*).*

B.2. Decomposition of Excess Returns over Long Horizons

The text makes the claim that we can decompose excess returns on horizons longer than one quarter between multiple one-period excess returns (equation (15)). I verify this claim now.

I start with the observation that the term premium is a combination of expected excess returns, as indicated by (1):

$$\begin{aligned} tp_t^n &= E_t\left(rx_{t,t+1}^n - rx_{t,t+1}^{n-1}\right) + E_t\left(rx_{t+1,t+2}^{n-1} - rx_{t+1,t+2}^{n-2}\right) + \cdots + E_t rx_{t+n-2,t+n-1}^2 \\ &= \sum_{j=1}^{n-1} E_t rx_{t+j-1,t+j}^{n-j+1} - E_t rx_{t+j-1,t+j}^{n-j} \end{aligned}$$

Cochrane and Piazzesi (2009) demonstrate the equality above graphically. In the case $n = 2$ we get $tp_t^2 = E_t rx_{t+1}^2$, which gives the pretty expression $f_t^2 = E_t i_{t+1} + E_t rx_{t+1}^2$.

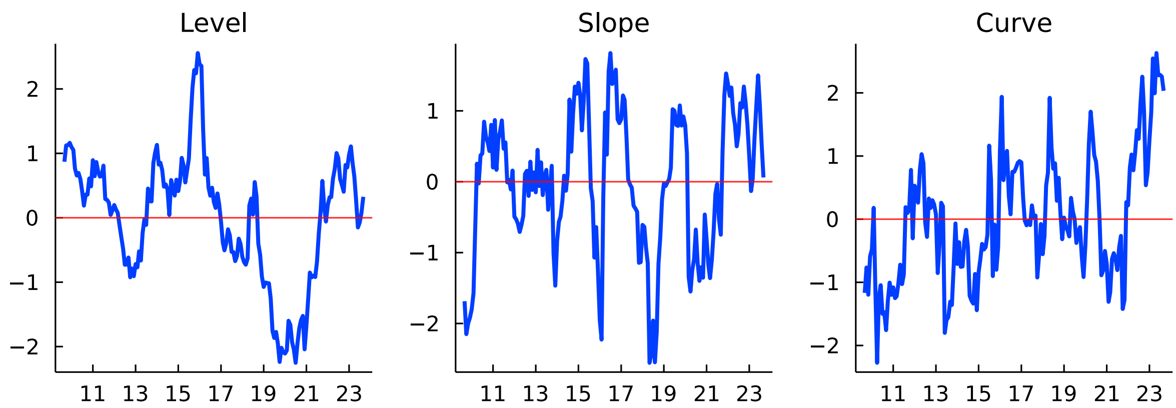
Returns over horizons longer than one quarter can be decomposed between individual one-period excess returns and term premia that adjust for the difference between future interest (which enter the definition of these one-period returns) and current forward rates (which enter the definition of the long-horizon return). The algebra:

$$
\begin{aligned}
rx^n_{t,t+m} &= p^{n-m}_{t+m} - p^n_t - \left[i_t + f^2_t + \cdots + f^m_t\right] \\
&= [p^{n-m}_{t+m} - p^{n-m+1}_{t+m-1} - i_{t+m-1}] - (f^m_t - i_{t+m-1}) \\
&+ [p^{n-m+1}_{t+m-1} - p^{n-m+2}_{t+m-2} - i_{t+m-2}] - (f^{m-1}_t - i_{t+m-2}) \\
&+ \cdots + \\
&+ [p^{n-1}_{t+1} - p^n_t - i_t]
\end{aligned}
$$

I use colored text to highlight variables that I add and subtract. When we take expectations, terms in brackets become one-period risk premia; terms in parentheses become term premia:

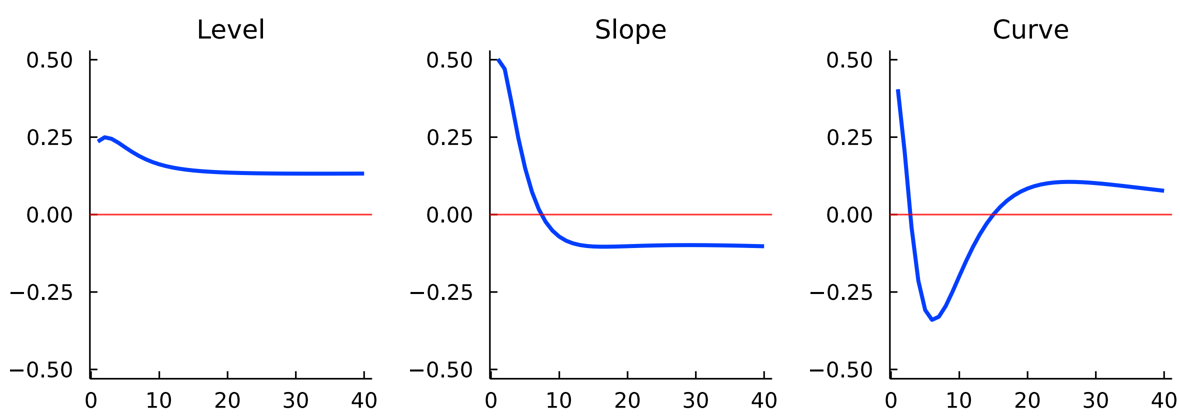
$$
\begin{aligned}
E_t rx^n_{t,t+m} = &\left[E_t rx^{n-m+1}_{t+m-1,t+m} + E_t rx^{n-m+2}_{t+m-2,t+m-1} + \cdots + E_t rx^n_{t,t+1}\right] \\
&- \left(tp^m_t + tp^{m-1}_t + \cdots + tp^2_t\right)
\end{aligned}
$$

We have shown that the term premia terms tp are the sum of one-period risk premia. The expression above therefore proves that the same is true for risk premia over horizons of more than a period. Replacing the expression for term premia, we get equation (15).



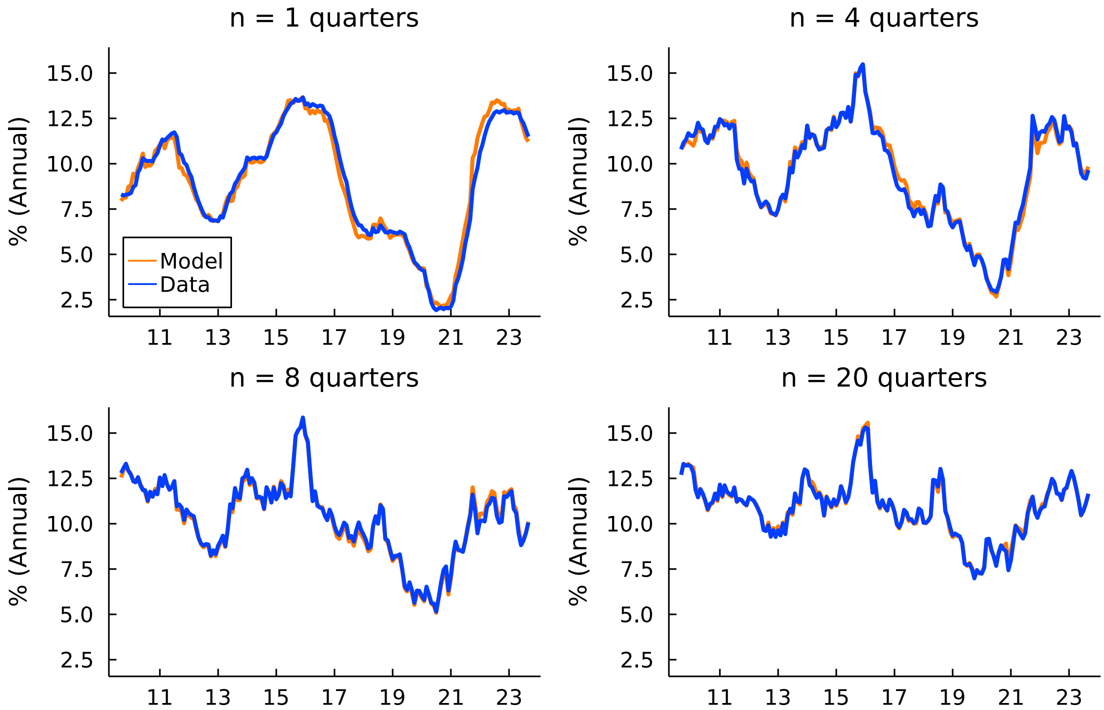
Notes. The x-axis represents time.

Figure 13: Model Factors (Time Series)



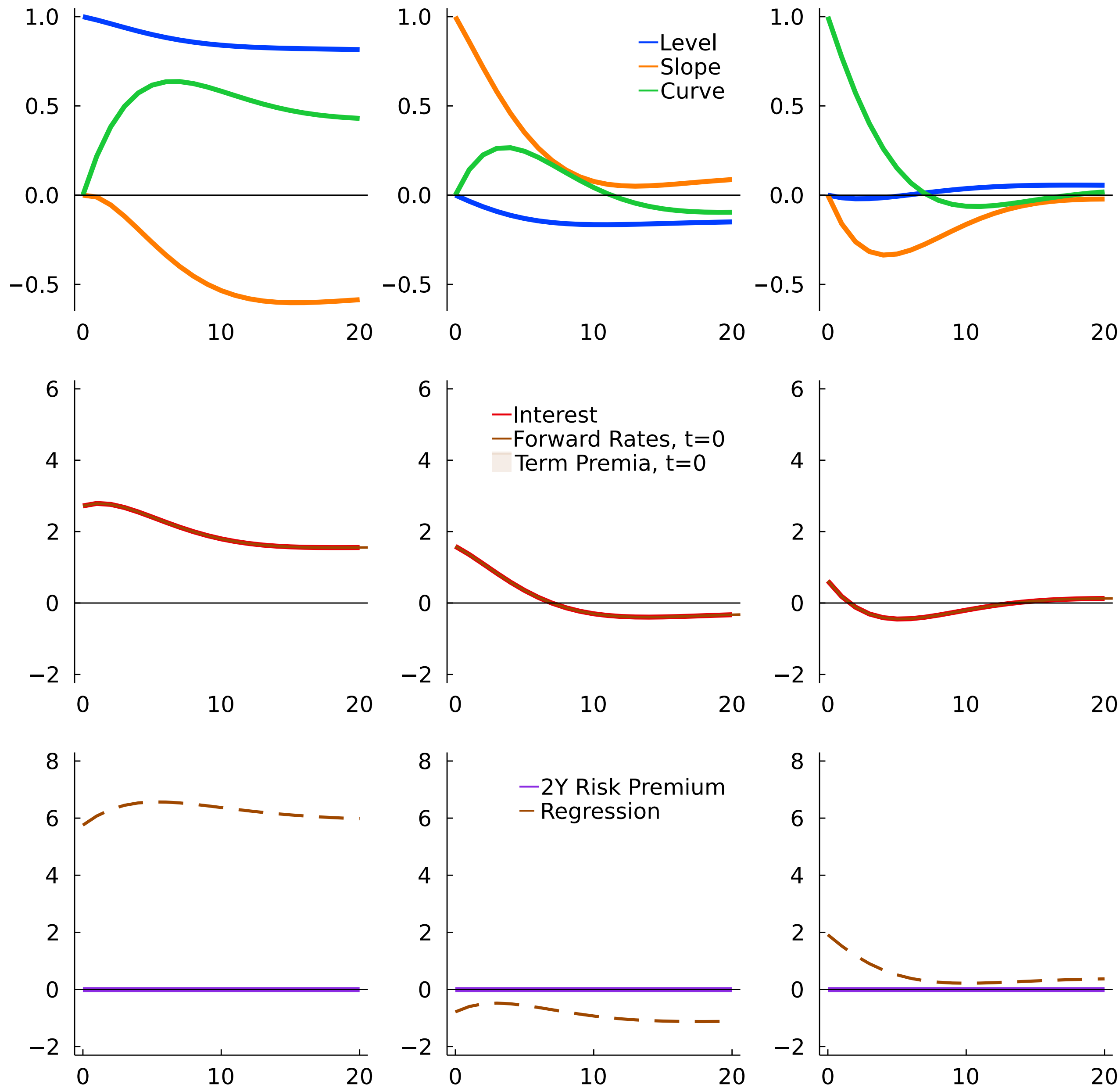
Notes. The x-axis represents maturity in quarters. The eigenvectors represent the OLS loadings of forward rates on each (unnormalized) factor, and the combination of forward rates that builds each factor.

Figure 14: Principal Components' Eigenvectors



Notes. The x-axis represents time.

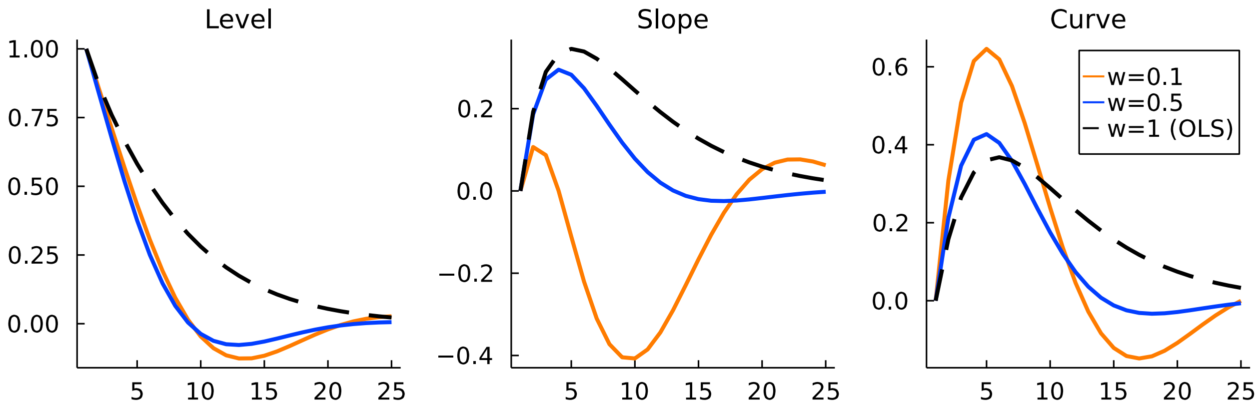
Figure 15: Forward Rates - Data and Model Pricing



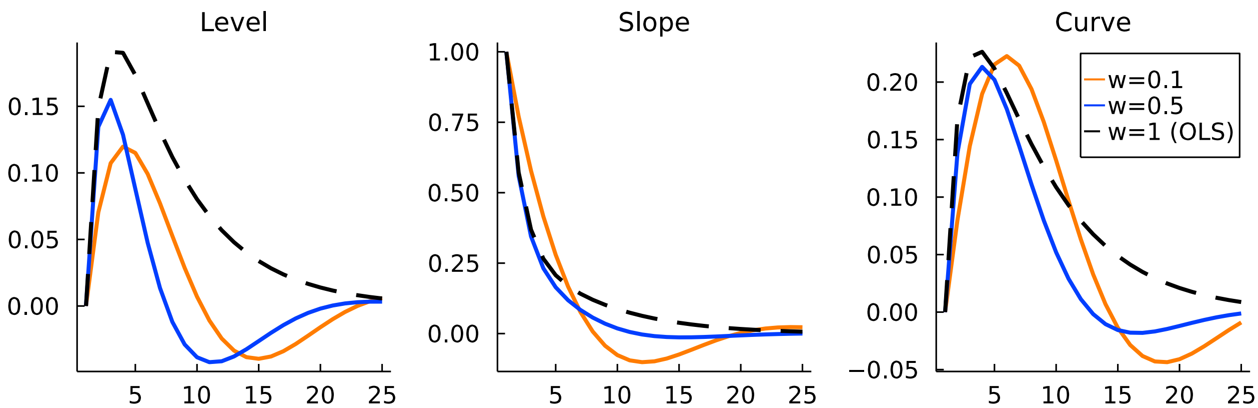
(a) Level shock **(b)** Slope shock **(c)** Curve shock

Notes. One period = one quarter. Top: factors X_t. Middle: interest rate i_t, forward rate at period zero $f^n_{t=0}$, and the term premia $tp^n_{t=0}$ (both as functions of maturity n). Bottom: risk premium on 2-year holdings of five-year bonds $E_t rx^{20}_{t,t+8}$ (also, the risk premium implied by the unrestricted return-forecasting regression (9)).

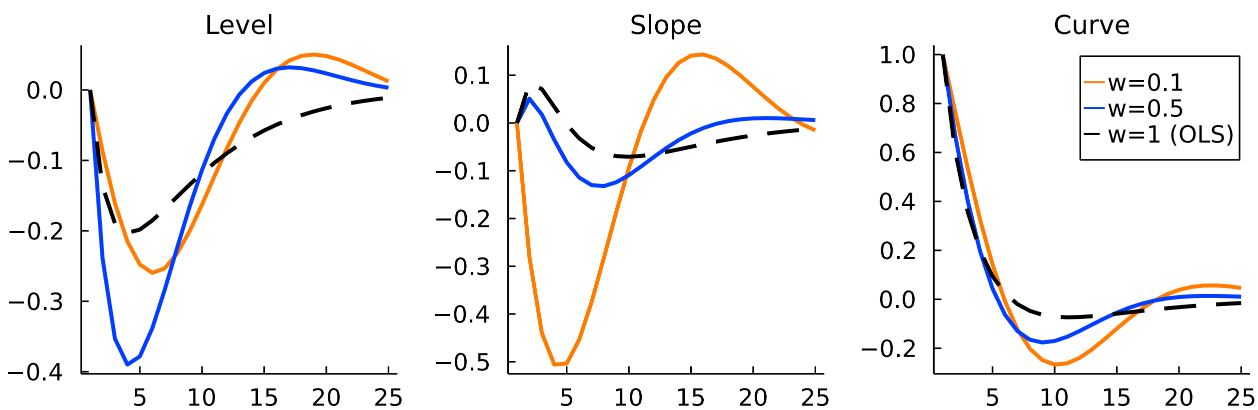
Figure 16: Risk-Neutral Dynamics: Impulse Response Functions



(a) Response to a Level Shock



(b) Response to a Slope Shock



(c) Response to a Curve Shock

Notes. Each of the nine plots contains three IRFs. The dashed line corresponds to the solution to the GMM problem using $w = 1$ (which coincides with straight OLS estimation of (2)). The blue corresponds to $w = 5$ (baseline), and the orange to $w = 0.01$. Parameter w governs the weight of OLS moments (M2) relative risk-premium moments (M3).

Figure 17: Physical Distribution: Comparison between IRFs with different w