Risk Premium and Yields: A GMM Estimation of Affine Term Structure Models

Abstract

I propose a new GMM-based estimation of affine models of the term structure. Through proper weighting of overidentifying moment conditions, the method allows us 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. An application to Brazilian yields shows that this modelling choice can lead to highly different decomposition of yields between term premia and interest expectations.

1. Introduction

Affine models of the term structure present the following restriction. Given the risk-neutral distribution of the underlying factors generating observed yields, you are free to select the data-generating distribution of these factors, or the expressions governing bond risk premia. But you cannot choose both. This restriction leads to two paths to estimating an affine model. You can follow Adrian et al. (2013) or Joslin et al. (2011), and estimate the data-generating dynamics of the factors by OLS; the model then gives you the expression for risk premia. Or you can follow Cochrane and Piazzesi (2009), and select price-of-risk parameters to reproduce unrestricted return-forecasting regressions; the model then gives you the required data-generating distribution of the underlying factors.

This paper proposes a new and simple estimation strategy for affine models that nests these two cases. The method contains two steps, each cast as a separate general method of moments optimization (GMM, Hansen (1982)). The first step selects parameters of the risk-neutral distribution to optimize the fit of the cross section of yields. The second step selects parameters of the data-generating distribution to match two groups of orthogonality conditions. The first group refers to the reduced-form shocks of the model, which must be orthogonal to model factors. Targeting only these moments is numerically equivalent to estimating data-generating parameters by OLS (as Adrian et al. (2013)). The second group refers to return-forecasting errors, which should also be orthogonal to model factors. Targeting only these moments is numerically equivalent to reproducing return-forecasting regressions, the strategy adopted by Cochrane and Piazzesi (2009). Hence, this second group of moments allows us to incorporate information relative to *bond risk premia* in the estimation of the affine model. By tuning the weighting matrix of the second GMM step, you can calibrate how much to incorporate.

One might ask: why care at all about reproducing realistic risk premia? First, economists often apply the affine model to decompose yields between expected interest and term premium. The term premium component is a sum of expected returns over time. Therefore, realistic dynamics of expected returns provide valuable information we might explore to discipline the decomposition. Second, risk premium moments can convey information about the long-term behavior of the factors. Modern empirical finance agrees that returns are, to some extent, predictable.[1] However, predictability might only show up over long investment horizons (Baker et al. (2003)), usually longer than a year.[2] Maximum likelihood and OLS estimation focus on minimizing one-step-ahead errors. Since the affine model is usually calibrated for monthly or quarterly data, these methods cannot explicitly target long-term predictability. On the other hand, in a GMM setup we can target predictability on any horizon. In practice, restrictions over long-term returns imply restrictions on the low-frequency dynamics of the underlying factors.

Does incorporating risk-premia information in the estimation lead to economically significant differences in the yield decompositions? I exemplify the GMM procedure using Brazilian data. In Brazil, two-year bond returns have been driven mainly by level movements in the term structure. OLS estimates fail to replicate this pattern. Consequently, decompositions of forward rates based on OLS estimation underestimate the term premium embedded in bond prices during times of elevated term structure level, if compared to an affine model that fully reproduces return-forecasting regressions. The differences are economically significant: the correlation between implied term premia series on two-year maturity bonds is 62%; the correlation between two-year-ahead expected interest, 3.5%.

[1] See Fama and French (1988), Cochrane (2008), Golez and Koudijs (2018), and many others, or Fama (1984), Cochrane and Piazzesi (2005), Andreasen et al. (2021) in the context of fixed-income markets.

[2] 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 persistent predicting variable and high-frequency noise that averages out over time.

2. The Affine Model

2.1. Notation

Let p_t^n be the log-price of an n-period maturity zero-coupon bond; $f_t^n = p_t^{n-1} - p_t^n$ the forward rate; $i_t = f_t^1$ the interest rate; and

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

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 $-p_t^m = i_t + f_t^2 + \cdots + f_t^m$. There is one risk premium $E_t rx_{t,t+m}^n$ to each maturity/investment horizon pair (n, m).

2.2. Forward Rate Decomposition

The forward rate f_t^n is the interest rate at which one can lock, in period t, a borrowing starting 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 one-period risk premia.[1]

[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.

2.3. Affine Model Restrictions

Following Ang and Piazzesi (2003) and Ang et al. (2007), consider a discrete-time version of the Duffie and Kan (1996) Gaussian affine model.[1] The state is a P-sized vector X_t. Its *data-generating* law of motion is

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

The affine model poses the discount factor

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

Vector $\lambda_t = \lambda_0 + \lambda_1 X_t$ captures the sensitivity of m_{t+1} to the shocks e_{t+1}. This discount factor leads to the interest rate solution

$$i_t = \delta_0 + \delta_1' X_t. \tag{3}$$

Additionally, the solution allows us to price bonds using a *risk-neutral* distribution, under which X_t follows:

$$X_t = \mu^* + \Phi^* X_{t-1} + e_t^* \qquad e_t^* \sim N(0, \Sigma), \tag{4}$$

where

$$\begin{aligned} \mu^* &= \mu - \Sigma\,\lambda_0, \\ \Phi^* &= \Phi - \Sigma\,\lambda_1. \end{aligned} \tag{5}$$

The solution for forward rates satisfies

$$f_t^n = v_n + E_t^* i_{t+n-1}, \tag{6}$$

where expectation E_t^* integrates using (4).[2] Therefore, if X_t follows (4), forward rates satisfy the expectations hypothesis: the term premium is

[1] This section presents the main restrictions of the model. The appendix presents details.

[2] $v_n = -\frac{1}{2} B_{n-1}' \Sigma B_{n-1}$ is a small Jensen inequality constant.

zero (except for the small v_n). Hence, its "risk-neutral" denomination. Importantly, (6) implies that, given (μ^*, Σ^*), yields do not depend on data-generating parameters (μ, Σ).

Finally, parameters (λ_0, λ_1) govern *risk premia* in the affine model. The solution for one-period expected returns is

$$E_t rx^n_{t,t+1} = v_n + \text{cov}_t(rx^n_{t,t+1}, e'_{t+1})\ \lambda_t. \tag{7}$$

The risk premium on a bond equals conditional covariances between return and shocks (the "quantity" of risk) and the discount rate sensitivity to these shocks λ_t (the "price" of risk). Since shocks e_{t+1} are homoscedastic, the covariance term is constant; thus all risk premia variation stems from risk price $\lambda_t = \lambda_0 + \lambda_1 X_t$.[1]

2.4. Summary of the Affine Model

In summary:

- (μ, Φ) govern the dynamics of X_t through (2);
- (μ^*, Φ^*) govern the cross-section of yields through (6);
- (λ_0, λ_1) govern risk premia through (7).

By (5), 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, which motivates my GMM approach.

3. Moment Conditions and Estimation

Our job is to estimate: data-generating parameters (μ, Φ), risk-neutral parameters (μ^*, Φ^*), price of risk (λ_0, λ_1), interest rate (δ_0, δ_1) and shock

[1] The solution for bond prices is linear in the state: $p^n_t = A_n + B'_n X_t$. Therefore, the covariance term in (7) equals $B'_{n-1}\Sigma$.

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

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

I also estimate interest parameters $(\hat{\delta}_0, \hat{\delta}_1)$ by running OLS on (3).[2]

Fixing $(\hat{\Sigma}, \hat{\delta}_0, \hat{\delta}_1)$, we focus on (μ, Φ), (μ^*, Φ^*), and (λ_0, λ_1). By (5), we can only choose two of these tuples. I proceed in two steps, each cast as a GMM problem. (For a generic moment condition $E[h_t(\theta)] = 0$ and positive-definite weighting matrix W, GMM estimates

$$\hat{\theta} = \ \text{Argmin}_{\theta} \quad g_\theta' \ W \ g_\theta \tag{8}$$

where $g_\theta = T^{-1}\sum_{t=1}^{T} h_t(\theta)$ is the sample counterpart of the moment condition.[3])

3.1. Step 1: Term Structure Pricing

Since we are ultimately interested in understanding the observed term structure, and because the forward rate decomposition (1) only holds for model-implied yields, we want the affine model to price them as close as possible to actual data. To optimize term structure fit, the first step of the estimation matches the following moment condition.

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

$$E\left[\epsilon_{n,t}^2\right] = 0 \qquad \text{where} \qquad \epsilon_{n,t} = f_t^n - \hat{f}_t^n, \tag{M1}$$

and $\hat{f}_t^n$ is the model-implied forward rate (6).

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

[2]We could also estimate (δ_0, δ_1) along with (μ^*, Φ^*) in the first GMM step below.

[3]See Hamilton (1994) for a textbook treatment of GMM estimators. In both steps of the estimation, I use numerical search algorithms to solve (8).

M1 holds in the affine model because pricing errors are identically zero $\epsilon_{n,t} = 0$. The first step of the estimation selects $(\hat{\mu}^*, \hat{\Phi}^*)$ to match M1 in the GMM optimization. By targeting it, we minimize squared pricing errors, like unrestricted OLS.[1]

3.2. Step 2: Data-Generating Dynamics and Risk Premia

Fix $(\hat{\mu}^*, \hat{\Phi}^*)$. The second step of the estimation selects (μ, Φ) and (λ_0, λ_1). (μ, Φ) governs the dynamics of model factors $E_t X_{t+m}$. (λ_0, λ_1) governs one-period risk premia, and thus *long-term* risk premia $E_t rx^n_{t,t+m}$.[2] Since (5) says we can only pick one of these tuples, and since we observe both X_t and $rx^n_{t,t+m}$, we need to decide which of these series we attempt to fit in the estimation. The second GMM problem targets two sets of orthogonality conditions that extract information from X_t and $rx^n_{t,t+m}$ separately.

Moment Condition 2 (Shock Orthogonality)**:**

$$E\left[e_{t+1}\right] = E\left[e_{t+1} X_t'\right] = 0. \tag{M2}$$

Moment Condition 3 (Return-Forecasting Error Orthogonality)**:**

$$E\left[\varepsilon^n_{t,t+m}\right] = E\left[\varepsilon^n_{t,t+m} X_t'\right] = 0, \tag{M3}$$

where $\varepsilon^n_{t,t+m} = rx^n_{t,t+m} - E_t rx^n_{t,t+m}$.

M2 holds in the affine model because e_{t+1} is an exogenous shock. X_t should not forecast it. M2 asks for an expected factor model $E_t X_{t+1}$ such that the resulting shock series $\hat{e}_{t+1}$ replicates this pattern. This orthogonality condition defines OLS estimates of (μ, Φ), the procedure advocated by Adrian et al. (2013).

However convenient, OLS estimation presents two caveats. First, it

[1] One is free to choose which bond maturities n to target and the weighting matrix. Including all maturities in sample might lead to substantial parameter variance due to large correlation between yields.

[2] The appendix shows that we can represent risk premia over any investment horizon $E_t rx^n_{t,t+m}$ as a linear combination of one-period risk premia $E_t rx^n_{t,t+1}$.

does not target empirical risk-premium patterns. Second, OLS estimates are highly sensitive to model specification. For instance, estimating a version of (2) in difference leads to wildly different conditional forecasts of X_t. In the case of near unit root processes, OLS leads to a downward bias on the system's persistence.[1]

M3 is the orthogonality condition that defines the OLS estimates of return-forecasting regressions

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

The affine model implies a similarly affine solution for excess returns.[2] M3 asks for a risk premium model $E_t rx^n_{t,t+m}$ such that return-forecasting errors $\hat{\varepsilon}^n_{t,t+m}$ are orthogonal to X_t; hence, a model that approximates the unrestricted return-forecasting regressions and thus incorporates risk premium information. Such is the strategy adopted by Cochrane and Piazzesi (2009).

In the one-period return case, $E_t rx^n_{t,t+1}$ is determined by (λ_0, λ_1) (equation (7)). In longer horizons, *when return predictability becomes clearer*, the connection with (λ_0, λ_1) blurs. But the principle of incorporating risk-premium information in the estimation of data-generating dynamics (μ, Φ) continues to apply. In that sense, a key advantage of estimating the model by GMM is that can target excess returns over investment horizons longer than one model period.

---

[1]See Yamamoto and Kunitomo (1984). This bias is particularly problematic in the context of affine models because risk-neutral dynamics Φ^* tends to have a near-unitary root. The data-generating distribution having a similar property is a plausible possibility.Cochrane and Piazzesi (2009) and Joslin et al. (2014) find near-unitary roots in the data-generating process. Several bias-correction methods have been proposed to address the issue. See Phillips and Yu (2005), Tang and Chen (2009) and Bauer et al. (2012).

[2]In particular:

$$rx^n_{t,t+m} = E_t rx^n_{t,t+m} + \varepsilon^n_{t,t+m} = E_t rx^n_{t,t+m} + \sum_{j=0}^{m-1} \Phi^j e_{t+m-j}.$$

In the affine model, return-forecasting errors are a linear combination of model shocks e_t. Therefore, M3 holds.

Weighting Matrix

The second estimation step selects $(\hat{\lambda}_0, \hat{\lambda}_1)$ through a GMM problem targeting *both* M2 and M3. We then infer $(\hat{\mu}, \hat{\Phi})$ from (5). I convert all moments to standard deviation units.

This second GMM problem contains overidentifying restrictions.[1] As such, the choice of a weighting matrix W is critical.[2] Consider the following one:

$$W_w = \begin{bmatrix} (1-w)\, \bar{I}_{M2} & 0 \\ 0 & w\, \bar{I}_{M3} \end{bmatrix}, \qquad w \in [0,1],$$

where $\bar{I}_{M2}$ and $\bar{I}_{M3}$ are identity matrices divided by their sizes (*i.e.* their diagonal sum to one), which match the number of moments in M2 and M3. W_w attributes a weight of w to return-forecasting orthogonality moments M3, and $1 - w$ to shock orthogonality M2. We can then interpret w as the *degree of risk premium information* incorporated in the estimation.

Next, I exemplify how the choice of w affects the results of the affine model.

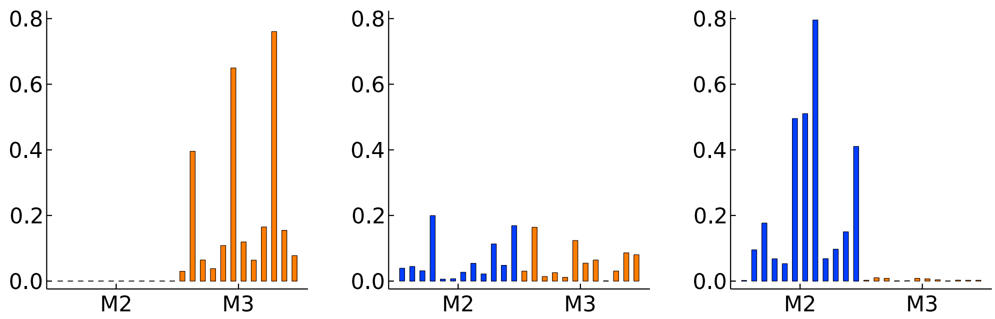
4. Application with Brazilian Data

I use Brazilian monthly data on zero-coupon bonds, spanning from 2009-M9 to 2023-M9.[3] For each month, I have the prices of bonds with maturities

[1]There are $P^2 + P$ parameters in (λ_0, λ_1), or in (μ, Φ). We match the $P^2 + P$ moment conditions in M2 and $P \times N$ moment conditions in M3 (where N is the number of targeted maturities).

[2]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). The efficient weighting matrix throws this trade-off under the carpet and asks that we pay more attention to combinations of individual moments with the lowest variance. Thus, while it can be an appropriate choice in other contexts, the efficient weighting matrix defeats the purpose of this paper.

[3]The Brazilian Financial and Capital Markets Association (ANBIMA) collects market price data of nominal bonds issued by the Brazilian federal government. 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. The model is very close to actual data. 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)



(a) $w = 0$ (OLS) **(b)** $w = 0.5$ **(c)** $w = 1$

Notes. Brazilian data. All moments in standard deviation units. Blue bars represent M2 errors; orange bars represent M3 errors.

Figure 1: Step 2: Estimated Moment Errors in Absolute Value

$n = 1, 2, \ldots, 40$ quarters. The appendix plots the data.

I assign as factors X_t the first three principal components of the cross section of forward rates.[1] They describe the *level*, *slope* and *curvature* of the term structure.

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

I choose signs so that the interest rate is increasing in all three factors.

The appendix reports $\hat{\Sigma}$, $(\hat{\delta}_0, \hat{\delta}_1)$, and the risk-neutral parameters $(\hat{\mu}^*, \hat{\Phi}^*)$ from step 1.[2] Moving to step 2, I target through M3 the two-year return ($m = 8$) on bonds with maturities $n = 12, 20, 28$ quarters. As you can see from table 1, two-year bond returns look quite predictable ($R^2 = 0.42$ with $n = 20$) and mostly driven by variation in the term structure *level*.

of 0.044% in the annualized yield to maturities. The closest analogue to this interpolation approach for the US is studied by Gürkaynak et al. (2007).

[1]Like in the US, these three factors capture over 99% of cross-section variation of Brazilian forward rates. 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 $\lambda_1 / \sum_i \lambda_i$. See Litterman and Scheinkman (1991), Knez et al. (1994), Dai and Singleton (2000).

[2]In step 1, I target maturities $n = 2, 6, \ldots, 38$ quarters.

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

Investment Horizon: $m = 8$ quarters
Bond Maturity: $n = 20$ quarters

Model	a	Level$_t$	Slope$_t$	Curve$_t$	R^2
Unrestricted OLS	3.8	5.8	-0.8	1.9	42.4
Affine Model, $w = 0$	3.7	2.6	-1.7	0.8	30.0
Affine Model, $w = 0.5$	3.6	5.1	-0.7	1.3	41.7
Affine Model, $w = 1$	3.8	5.8	-0.7	1.9	42.4

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 risk-premia moments M3 that define the OLS estimates of the return-forecasting regression. Columns a and R^2 reported in percentages.

Table 1: Return-Forecasting Regression: OLS Evidence vs Affine Model

Figure 1 plots the minimized sample moments in absolute value for three choices of w.[1] Higher bars mean larger moment errors.[2]

When $w = 0$, $(\hat{\mu}, \hat{\Phi})$ coincide with the OLS estimates of (2): the affine model matches M2 exactly. However, it misses risk premium moments M3 by as much as 75% of a standard deviation, since we do not incorporate any risk-premium information. Table 1 shows how this specification misses significant return predictability (over 10% R^2 points). In addition, the loading on the level factor is less than half what we get from unrestricted regressions.

As we increase w, the affine model better matches M3 and the return-forecasting regressions. The cost: model shocks e_{t+1} get increasingly less orthogonal to X_t. Figure 1 illustrates this trade-off imposed by the affine framework. The intermediary case $w = 0.5$ balances moment errors.

[1] Instead of $w = 1$, I actually use $w = 0.999$, since $w = 1$ leads to wild one-period ahead dynamics. For clarity of the argument, I call this case $w = 1$.

[2] To convert moments to standard deviation units, I divide moments involving X_t by 1%; M2 moments by 0.5%, which is about the standard deviation of e_t; and M3 moments by 5%.

4.1. Dynamics and Yield Decomposition

Figure 2 describes the affine model response to a level shock in the limit cases $w = 0$ (no risk premium information incorporated), and $w = 1$ (information fully incorporated). I focus on level shocks because they lead to the largest differences across w in this application.[1]

The top panels depict the impulse response function (IRF) of the two-year risk premium on $n = 20$ bonds in the affine mode and according to return-forecasting regressions. OLS estimation halves the shock effect on risk premia. The $w = 1$ model, by construction, replicates regression results.

The middle panels contain factors IRFs under data-generating dynamics $(\hat{\mu}, \hat{\Phi})$. It explains how greater risk premia comes about. Compared to OLS, in the $w = 1$ case the level factor response decays faster and switch signs, and the slope response is negative. Since interest is increasing in X_t, its response also declines faster (bottom panels, thick curve), which implies that bond prices recover faster. Hence the larger premium. You can see how fitting long-term risk premia changes the long-term behavior of the affine model.

The bottom panels also depict the time-0 response of the term structure (thin curve), which is the same in both specifications.[2] The term premia response in $t = 0$ is the difference between the curves.[3] If the shock effect on interest fades faster under $w = 1$, then the difference must be accounted by larger term premia - which is, in turn, consistent with higher risk premium called for by the return-forecasting regressions.

Figure 3 shows how enhanced return-predictability in the $w = 1$ specification changes the decomposition of two-year forward rates. In the OLS case, expected interest closely follows market yields. As we incorporate

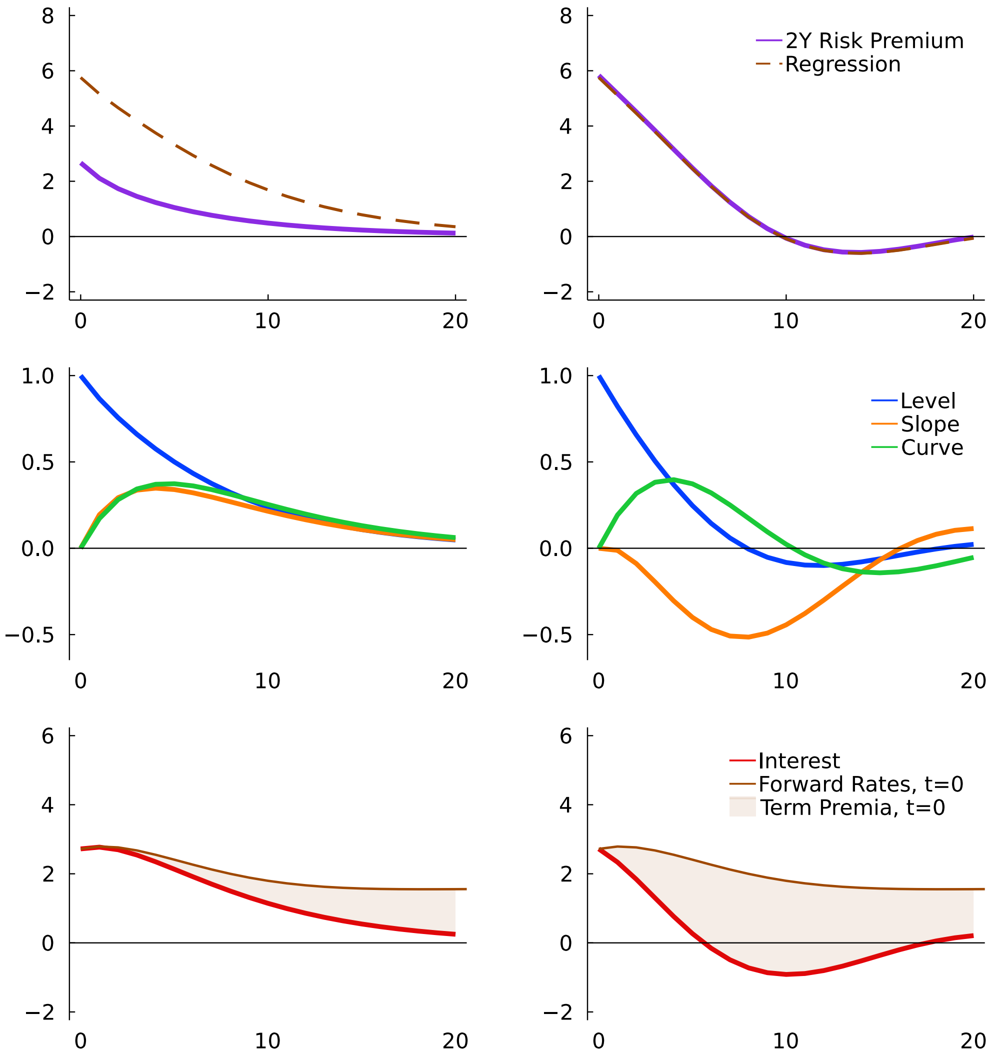
[1]The appendix plots similar figures for slope and curvature shocks.

[2]Recall that forward rates are pinned down by risk-neutral parameters (μ^*, Φ^*), estimated in step 1.

[3]By (1),

$$\Delta E_t tp^n_{t=0} = \Delta E_t f^n_{t=0} + \Delta E_t i_{t+n-1},$$

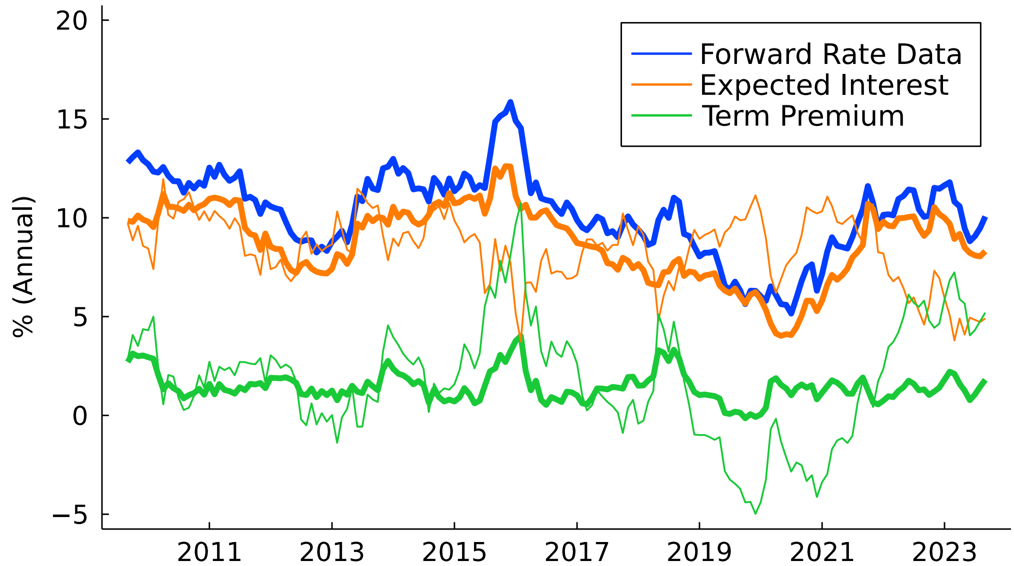
where $\Delta E_t = E_t - E_{t-1}$ denotes the revision in expectations following the shock.



(a) $w = 0$ (OLS) **(b)** $w = 1$

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

Figure 2: Model Response a Level Factor Shock

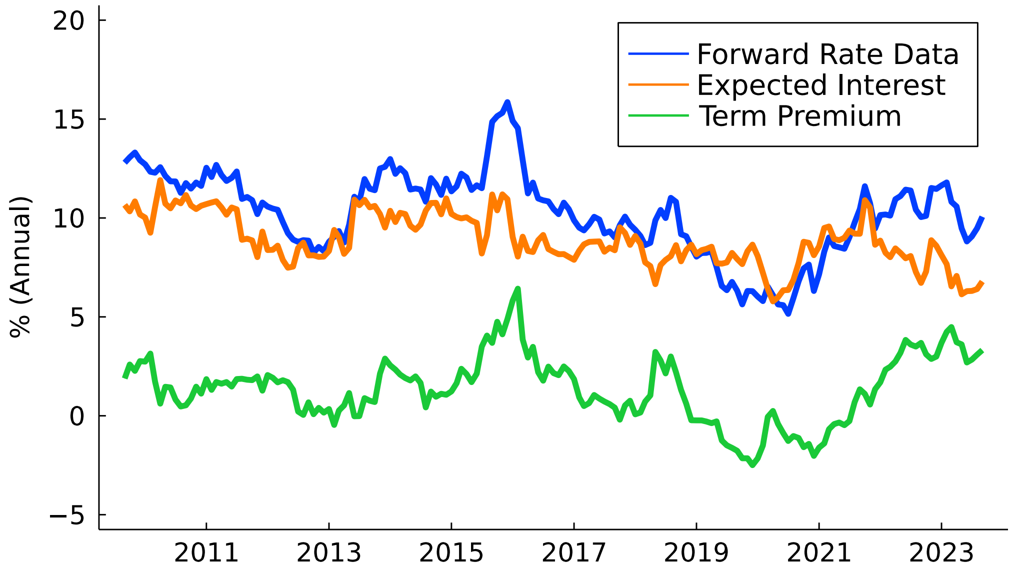


Notes. The figure plots the estimates of the forward rate decomposition (1), for two-year bond maturities $n = 8$. I multiply all rates by four to annualize them.

Figure 3: Forward Rate Decomposition of Two-Year Yields. Thick: $w = 0$ (OLS). Thin: $w = 1$.

risk premia in the estimation, interest reverts to average faster, particularly in 2015/16 and 2019, when the term structure level was respectively too high and too low. This is the pattern we see in the IRFs. Faster interest mean-reversion leads to additional return predictability, resulting in a more volatile term premium curve. The differences are highly significant economically. The correlation between term premium curves is 62%, and 3.5% between expected interest curves.

Lastly, a key advantage of the GMM method I propose is the possibility of only partially incorporating risk premia information. As one might expect, figure 4 shows how the decomposition implied by the intermediary case $w = 0.5$ is approximately an average of the two limit cases.



Notes. The figure plots the estimates of the forward rate decomposition (1), for two-year bond maturities $n = 8$. I multiply all rates by four to annualize them.

Figure 4: Forward Rate Decomposition of Two-Year Yields ($w = 0.5$)

5. Conclusion

This paper presents a new GMM-based estimation of affine models of the term structure that incorporates long-term bond risk premia. Using Brazilian data, I show that replicating realistic risk premia might require substantially different dynamics of bond prices and interest rates.

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.

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.

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. (2005). Bond Risk Premia. *American Economic Review*, 95(1):138–160.

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

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

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.

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

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

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.

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.

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.

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

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.

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. Derivation of the Gaussian Affine Model

This section presents the details of the 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:

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

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

Equation (A.1) holds for a general stochastic discount factor m_{t+1}. The m_{t+1} used by the affine model is given by (A.2):

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

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 (A.2), we get the linear solution to i_t given by (3) and

$$E_t rx^n_{t,t+1} = -(1/2)\text{var}_t(rx^n_{t,t+1}) + \text{cov}_t(rx^n_{t,t+1}, e'_{t+1})\lambda_t. \tag{A.3}$$

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 (A.3) 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 (A.3) implies a solution for equilibrium prices linear in the factors, given by

$$p^n_t = A_n + B_n \cdot X_t.$$

Coefficients A_n and B_n satisfy the well-known Riccati equations

$$\begin{aligned} A_{n-1} - A_n + B'_{n-1}\mu - \delta_0 &= -\frac{1}{2}B'_{n-1}\Sigma B_{n-1} + B'_{n-1}\Sigma\lambda_0 \\ B'_{n-1}\Phi - B'_n - \delta'_1 &= B'_{n-1}\Sigma\lambda_1 \end{aligned} \tag{A.4}$$

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 (A.4) 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 (A.4) and gather terms, we arrive at

$$\begin{aligned} A_{n-1} - A_n + B'_{n-1}(\mu - \Sigma\lambda_0) - \delta_0 &= v_n \\ B'_{n-1}(\Phi - \Sigma\lambda_1) - B'_n - \delta'_1 &= 0. \end{aligned} \tag{A.5}$$

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

[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.

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 (4), which performs that change of drift, as "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 (A.5):

$$\begin{aligned}
f_t^n &= A_n^f + B_n^{f\prime} X_t \\
&= v_n + \delta_0 - B'_{n-1}\mu^* + \delta_1' \Phi^{*n-1} X_t \\
&= v_n + \delta_0 + \delta_1' \left[(I + \Phi^* + \cdots + \Phi^{*n-1})\mu^* + \Phi^{*n-1} X_t \right] \\
&= v_n + \delta_0 + \delta_1' E_t^* X_{t+n-1} \\
&= v_n + E_t^* i_{t+n-1}.
\end{aligned}$$

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 (A.5). 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 $f_t^n = v_n + E_t^* i_{t+n-1}$.

A.1. Decomposition of Excess Returns over Long Horizons

The text claims we can decompose excess returns on horizons longer than one quarter between multiple one-period excess returns. Specifically:

$$E_t rx_{t,t+m}^n = \sum_{j=0}^{m-1} E_t rx_{t+j,t+1+j}^{n-j} - \sum_{j=0}^{m-1} E_t rx_{t+j,t+1+j}^{m-j} \tag{A.6}$$

I verify (A.6).

As indicated by (1), the term premium is a combination of expected one-period excess returns:

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

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 $f_t^2 = E_t i_{t+1} + E_t rx_{t+1}^2$.

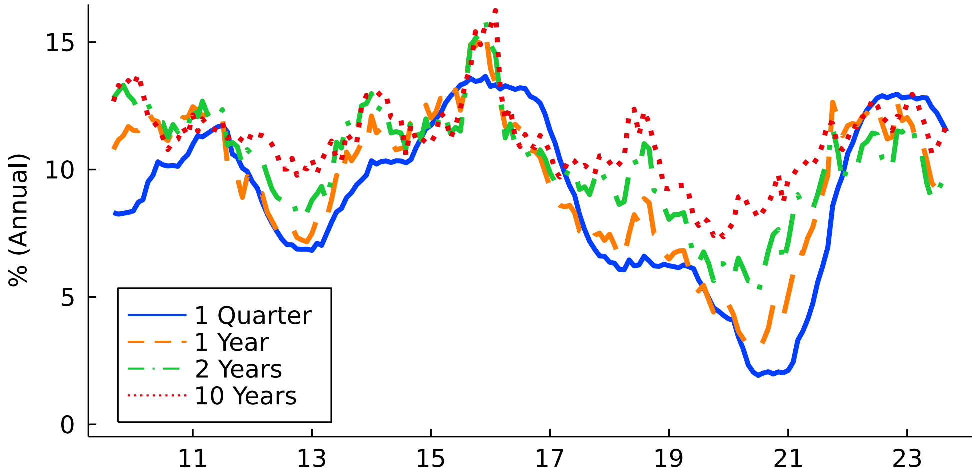
Returns over horizons longer than one quarter can be decomposed between individual one-period excess returns and term premia that adjust for the difference between future interest (which enter the definition of these one-period returns) and current forward rates (which enter the definition of the long-horizon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

$$
\begin{aligned}
rx_{t,t+m}^n &= p_{t+m}^{n-m} - p_t^n - \left[i_t + f_t^2 + \cdots + f_t^m\right] \\
&= [p_{t+m}^{n-m} - p_{t+m-1}^{n-m+1} - i_{t+m-1}] - (f_t^m - i_{t+m-1}) \\
&+ [p_{t+m-1}^{n-m+1} - p_{t+m-2}^{n-m+2} - i_{t+m-2}] - (f_t^{m-1} - i_{t+m-2}) \\
&+ \cdots + \\
&+ [p_{t+1}^{n-1} - p_t^n - i_t]
\end{aligned}
$$

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:

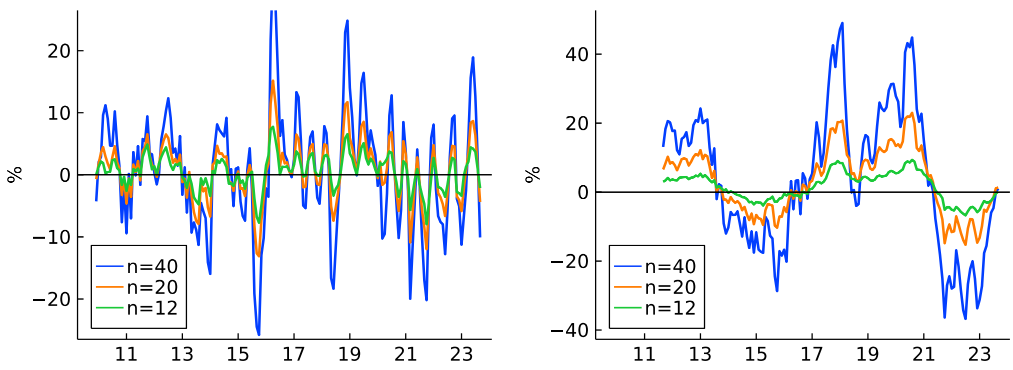
$$
\begin{aligned}
E_t rx_{t,t+m}^n = &\left[E_t rx_{t+m-1,t+m}^{n-m+1} + E_t rx_{t+m-2,t+m-1}^{n-m+2} + \cdots + E_t rx_{t,t+1}^n\right] \\
&- \left(tp_t^m + tp_t^{m-1} + \cdots + tp_t^2\right)
\end{aligned}
$$

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 (A.6).



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

Figure 5: Quarterly Forward Rates f_t^n (Data)

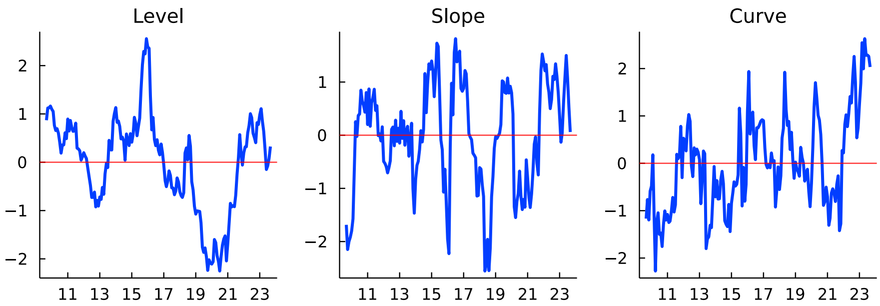


(a) 1-quarter holdings ($m = 1$)

(b) 2-year holdings ($m = 8$)

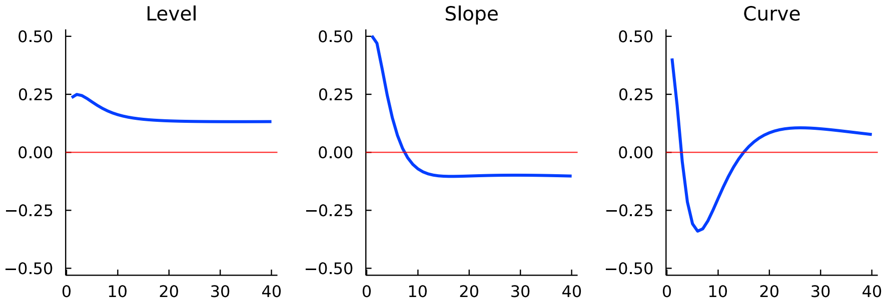
Notes. Each curve corresponds to the excess return on a bond of a given maturity $n \in (12, 20, 40)$.

Figure 6: Excess Returns $rx_{t,t+m}^n$ (Data)



Notes. The x-axis represents time.

Figure 7: 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 8: Principal Components' Eigenvectors

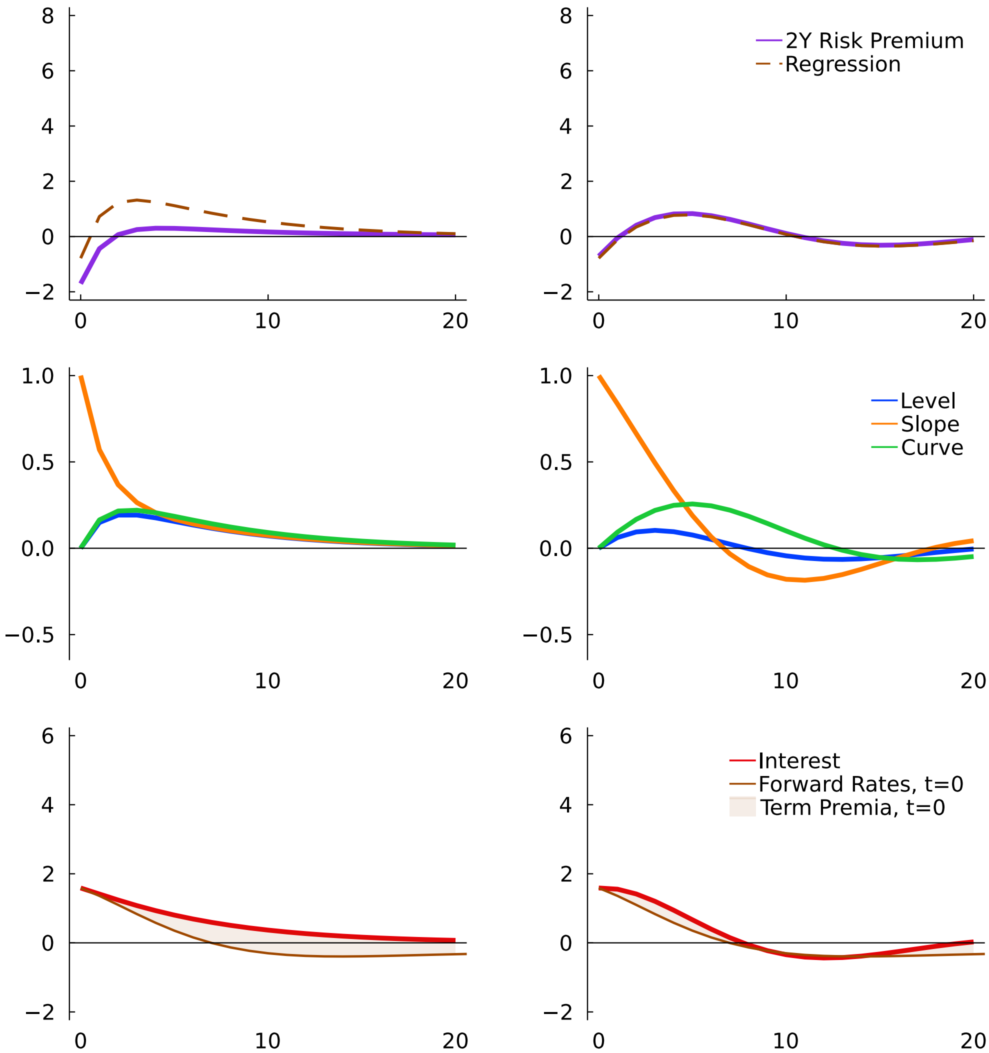
		Level	Slope	Curve	|Eig|
Risk-Neutral	$\hat{\mu}^*$		$\hat{\Phi}^*$		
Level	0.05	0.98	-0.03	-0.01	0.82
	(0.00)	*(0.01)*	*(0.01)*	*(0.01)*	
Slope	0.05	-0.01	0.86	-0.16	0.82
	(0.05)	*(0.06)*	*(0.07)*	*(0.03)*	
Curve	-0.08	0.22	0.14	0.77	0.997
	(0.29)	*(0.05)*	*(0.03)*	*(0.13)*	
Data-Gen. ($w = 0$)	$\hat{\mu}$		$\hat{\Phi}$		
Level	-0.01	0.87	0.15	-0.14	0.43
	(0.06)	*(0.06)*	*(0.09)*	*(0.06)*	
Slope	0.02	0.19	0.57	0.08	0.75
	(0.82)	*(0.07)*	*(0.08)*	*(0.09)*	
Curve	0.12	0.17	0.17	0.59	0.85
	(0.68)	*(0.10)*	*(0.07)*	*(0.15)*	
Data-Gen. ($w = 0.5$)	$\hat{\mu}$		$\hat{\Phi}$		
Level	0.02	0.86	0.13	-0.26	0.46
	(0.06)	*(0.06)*	*(0.09)*	*(0.07)*	
Slope	0.03	0.20	0.56	0.05	0.85
	(0.82)	*(0.07)*	*(0.07)*	*(0.09)*	
Curve	0.10	0.22	0.14	0.68	0.85
	(0.68)	*(0.09)*	*(0.10)*	*(0.14)*	
Data-Gen. ($w = 1$)	$\hat{\mu}$		$\hat{\Phi}$		
Level	-0.02	0.82	0.06	-0.09	0.75
	(0.07)	*(0.10)*	*(0.09)*	*(0.15)*	
Slope	0.01	-0.01	0.84	-0.36	0.90
	(0.82)	*(0.09)*	*(0.07)*	*(0.10)*	
Curve	0.06	0.19	0.09	0.83	0.90
	(0.67)	*(0.08)*	*(0.05)*	*(0.15)*	

Notes. μ column in percentage. $|\text{Eig}|$ reports the absolute values of the eigenvalues of Φ. GMM asymptotic standard errors in parenthesis (spectral density matrix estimated using Newey and West (1987) estimator with 24 lags).

Table 2: Affine Model Dynamics: GMM Estimation Results

		Level	Slope	Curve
	$\hat{\delta}_1$		$\hat{\Sigma} \times 100^2$	
$100 \times \hat{\delta}_0$	2.24			
Level	0.68	0.21	-0.22	-0.03
Slope	0.40	-0.22	0.56	-0.19
Curve	0.15	-0.03	-0.19	0.56

Table 3: Preliminary Estimates of $(\hat{\delta}_0, \hat{\delta}_1)$ and $\hat{\Sigma}$

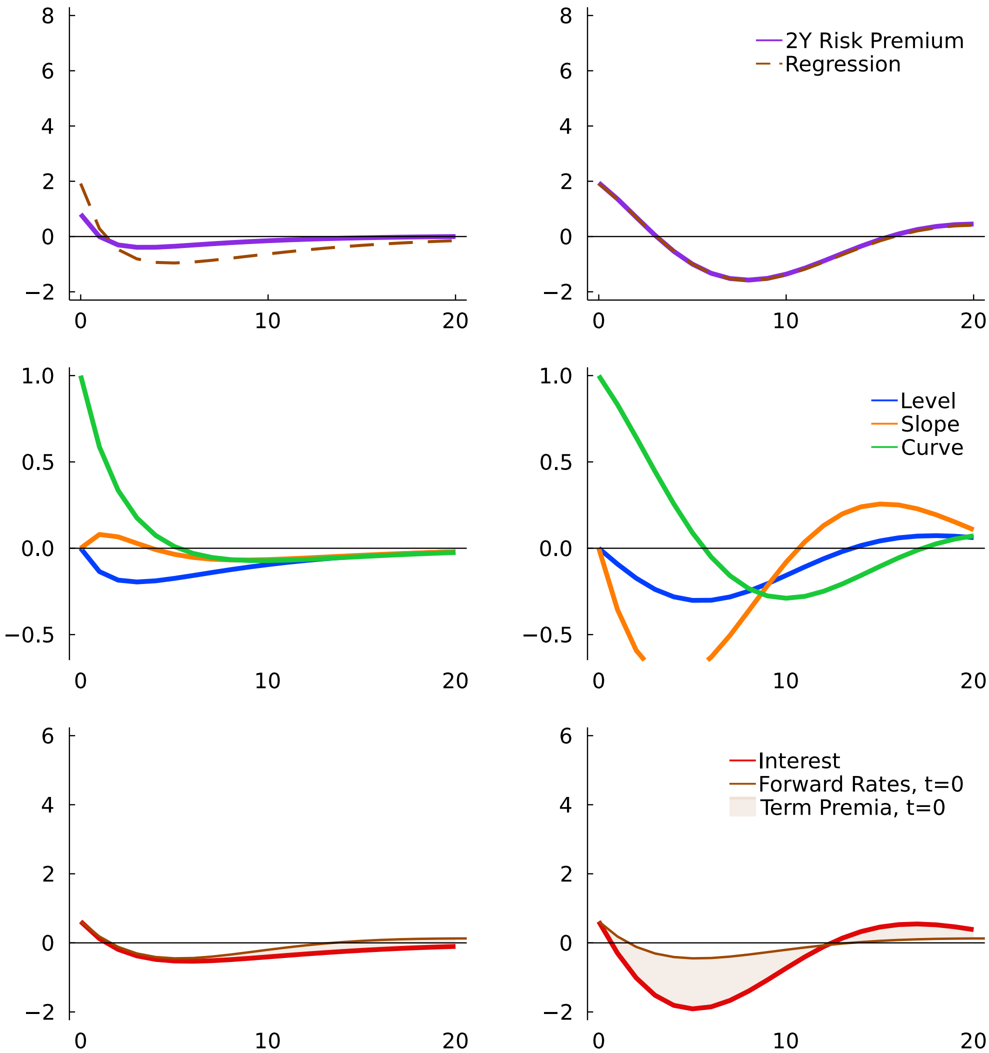


(a) $w = 0$ (OLS)

(b) $w = 1$

Notes. One period = one quarter. Top: risk premium on 2-year holdings of five-year bonds $E_t rx^{n=20}_{t,t+8}$ (also, the risk premium implied by the unrestricted return-forecasting regression (9)). Middle: factors X_t. Bottom: 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).

Figure 9: Model Response a Slope Factor Shock



(a) $w = 0$ (OLS)　　**(b)** $w = 1$

Notes. One period = one quarter. Top: risk premium on 2-year holdings of five-year bonds $E_t rx^{n=20}_{t,t+8}$ (also, the risk premium implied by the unrestricted return-forecasting regression (9)). Middle: factors X_t. Bottom: 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).

Figure 10: Model Response a Curvature Factor Shock