Breaking the Tension: Almost Affine Term Structure Models with Stochastic Volatility*

Scott Joslin† Anh Le‡

This draft: July 13, 2025

Abstract

While affine term structure models have long provided a tractable and elegant framework for capturing interest rate dynamics, the literature has documented a tension between achieving a good fit to the cross-section of bond yields and capturing the time-series behavior of first and second moments of yields. We introduce a new class of no-arbitrage term structure models that incorporates stochastic volatility while relaxing the rigid structural constraints of traditional affine frameworks. Our approach of modeling volatility as an almost affine process preserves much of the tractability of affine models, including approximately linear bond pricing, while offering significantly greater flexibility in capturing the dynamics of both the conditional means and conditional volatility of yields. We demonstrate that our models outperform existing affine models along key dimensions. Interestingly, while existing affine models attribute as much as 80% of term premium variations to changes in yield volatility, our framework assigns a much smaller share – no more than 20% – to this channel.

Keywords: Term Structure Models, Stochastic Volatility, Term Premium

*We thank seminar participants at Penn State University, Foreign Trade University Hanoi, Shenzhen Finance Institute for helpful comments.

†University of Southern California, Marshall School of Business, sjoslin@usc.edu

‡Smeal College of Business, Pennsylvania State University, anh.le@psu.edu

1 Introduction

Affine term structure models (ATSMs), especially those incorporating stochastic volatility, have long provided a tractable and elegant framework for capturing interest rate dynamics. A longstanding empirical challenge for affine no-arbitrage term structure models is their difficulty in simultaneously capturing the cross-sectional and time-series dynamics of bond yields while also generating plausible time-variation in yields volatility. As observed by Dai and Singleton (2000) and Dai and Singleton (2002), this is the "tension in matching simultaneously the historical properties of the conditional means and variances of yields." Similarly, Duffee (2002) notes that the functional forms for risk premia are insufficiently flexible and, as a consequence, the overall goodness of fit "is increased by giving up flexibility in forecasting to acquire flexibility in fitting conditional variances."

These observations reflect a structural rigidity in affine term structure models with stochastic volatility that manifests in empirical trade-offs: while these models may price bonds accurately under the risk-neutral measure, they often fall short in describing the joint dynamics of yield levels, slopes, and volatilities under the physical measure. As highlighted in Joslin and Le (2021), the restrictions linking the volatility structure to the feedback dynamics of the state vector under both the physical and risk-neutral measures distort the model's ability to match key features of the data – most notably, the predictive relationship between the slope of the yield curve and future rate changes.

This tension is perhaps most transparently revealed through two regression-based diagnostics introduced by Dai and Singleton (2002) that focus on the ability of term structure models to match empirical patterns concerning conditional first moments of yields. The first test, based on observations of Campbell and Shiller (1991), assesses whether the model can match the failures of the expectation hypothesis found in the data. Specifically, a term structure model should match the pattern found in the data between the slope of the yield curve and future changes in interest rates. The second test builds on this idea assessing whether adjusting realized changes in yield by model-implied term premia can bring the predicted change back into line with the expectations hypothesis. Dai and Singleton (2002) find that the "the requirement that a DTSM match both is a powerful discriminator among models." They find in their data that while Gaussian term structure models are able to pass both of the tests, the general models with stochastic volatility are not able to do so. Their paper was one of the first to highlight this tension between first and second moments: adding stochastic volatility to the affine models allowed them to match conditional second moments but caused the models to fail first-moment tests.

In this paper, we introduce a new class of no-arbitrage term structure models which we term *almost affine models* that relax the structural constraints of standard affine models while preserving analytical tractability. The central innovation lies in modeling volatility as a nonlinear yet tractable function of the yield factors, such as through a max-linear transformation. This specification guarantees positive volatility without requiring volatility factors to be autonomous. Consequently, our framework permits richer interactions between volatility and the rest of the state vector, breaking the structural rigidity that often impairs the flexibility of traditional ATSMs.

Relaxing this type of constraint improves the performance of standard ATSMs along at least two important *ex ante* dimensions. First, our model eliminates the restrictive link between the feedback matrices under the physical and risk-neutral measures. This decoupling allows the estimation procedure to more accurately target the conditional means of bond yields without sacrificing cross-sectional pricing accuracy. Indeed, we show that our almost affine models – across various specifications involving different numbers of yield and volatility factors – consistently pass the two regression-based diagnostics introduced by Dai and Singleton (2002), matching the success of pure Gaussian term structure models in capturing the conditional means dynamics of bond yields.

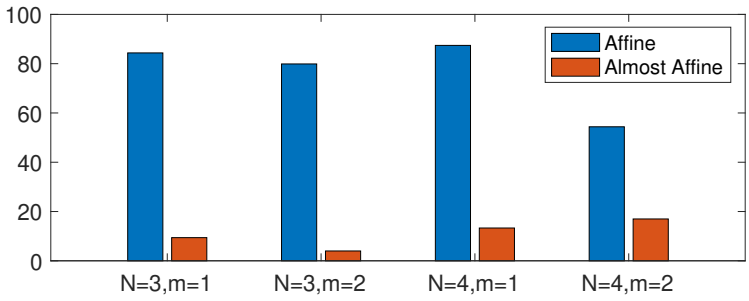
Second, volatility identification gains substantial flexibility. In conventional affine settings, as emphasized by Joslin and Le (2021), volatility instruments are subject to stringent eigenvector constraints that tie them simultaneously to both physical and risk-neutral feedback matrices. These restrictions significantly limit the model's ability to capture the empirical dynamics of yield volatility. In contrast, our framework enables a more flexible and unconstrained selection of volatility instruments. This not only facilitates better empirical fit to the volatility structure of yields but also has the potential to improve overall estimation efficiency – analogous to the well-known gains of generalized least squares over ordinary least squares.

Importantly, these gains in flexibility and empirical performances are achieved without sacrificing tractability. By construction, the almost affine framework preserves much of the analytical convenience of traditional affine models, as bond yields remain approximately linear in the state variables. Key to obtaining this approximation is our requirement that the max-linear transformation linking yield factors to volatility binds only infrequently inside the model. This design ensures theoretical admissibility of the volatility process while keeping the pricing implications close to those of a fully linear model. We show that yield approximation errors across the full maturity spectrum and over the entire sample period are economically negligible, typically on the order of one basis point or less.

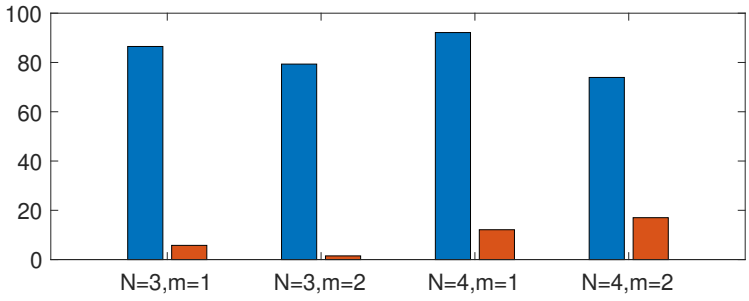
With their ability to price bonds accurately while simultaneously fitting both the conditional means and variances of bond yields, the proposed almost affine models offer a powerful framework for exploring fundamental questions about risk and return trade-offs in fixed income markets. When applied to the decomposition of term premium variation, the almost affine framework reveals a stark contrast with its conventional affine counterparts.

Figure 1 illustrates the proportion of time variation in term premiums attributable to interest rate volatility – the so-called "quantity of risk" channel – across models with different numbers of yield pricing factors (N) and volatility factors (m). For all model specifications and at both 5-year and 10-year horizons, the conventional affine models imply a near-mechanical link between volatility and term premium dynamics, attributing up to 80-90% of the variation to changes in yield volatility. By contrast, the almost affine models consistently attribute a much smaller share – no more than 20% – to the quantity of risk.

This striking divergence underscores a key implication: most of the variation in term premiums under the almost affine framework arises not from changes in volatility per se, but from shifts in the market price of risk, such as evolving risk aversion, investor sentiment, or



(a) 5-year horizon



(b) 10-year horizon

Figure 1: The share of interest rate volatility (in percentage) in explaining the time variation in term premiums implied by models with different number of yield pricing factors (N) and volatility factors (m).

macroeconomic conditions. The finding is robust across specifications (N = 3 or 4; m = 1 or 2), and across horizons, suggesting that relaxing the structural rigidity of affine models enables a more nuanced understanding of what drives bond risk premia.

Our paper contributes to the literature on no-arbitrage term structure models, and more specifically, to the strand that incorporates stochastic volatility into affine frameworks to better capture the joint dynamics of bond yields and risk premia. Foundational work by Duffie and Kan (1996), Dai and Singleton (2000), Dai and Singleton (2002), and Duffie, Pan, and Singleton (2000) established the analytical tractability and pricing implications of affine models, which have since become a cornerstone of the term structure modeling literature. Building on this foundation, many studies – including Collin-Dufresne and Goldstein (2002), Collin-Dufresne, Goldstein, and Jones (2008), Cheridito, Filipovic, and Kimmel (2007), and Duffee (2002) – have examined the role of stochastic volatility specification in improving the empirical performance of these models, particularly in explaining yield curve dynamics, bond return volatility, and the pricing of interest rate derivatives.

Two recent papers that are particularly relevant to our work are Joslin and Le (2021) and Doshi, Jacobs, and Liu (2024). Joslin and Le (2021) identify a fundamental structural rigidity in conventional affine term structure models – a source of restrictiveness that our

almost affine framework is specifically designed to address. While they explore several model variants that partially relax this constraint, those formulations are intentionally simplified for illustrative purposes. For instance, some of the models considered do not impose the no-arbitrage restrictions central to the ATSM literature, while others assume constant volatility under the risk-neutral measure despite allowing time-varying volatility under the physical measure. These simplifying assumptions, while helpful for exposition, limit the empirical applicability of those models. Our framework offers a fully specified, no-arbitrage model that accommodates flexible volatility dynamics under both measures.

Doshi, Jacobs, and Liu (2024) develop a no-arbitrage affine term structure model in which volatility factors follow GARCH-type dynamics. Strictly speaking, their model belongs to a restricted subclass of ATSMs in which innovations to the volatility factors are functionally tied to the innovations in the yield factors. In contrast to our paper, their focus is quite different. They emphasize the role of volatility identification in the estimation of ATSMs and explore the implications of their framework for pricing interest rate derivatives. Our analysis, by comparison, is centered on the interaction between yield predictability, bond risk premia and stochastic volatility. While they also investigate the implications of their model for the Campbell and Shiller (1991) regression, their approach appears to be in-sample, whereas our analysis – consistent with much of the term structure literature – examines the population-level implications of this regression.

The remainder of the paper is organized as follows. Section 2 revisits the sources of tension in standard ATSMs and motivates the need for a more flexible approach. Section 3 introduces the almost affine modeling framework and outlines its theoretical properties. Section 4 discusses the role of no-arbitrage and its implications for the model's dynamics. Section 5 presents our empirical implementation, including estimation results, model diagnostics. Section 6 provides analysis of the model-implied predictive regressions and term premium decomposition. Section 7 concludes.

2 The sources of tension in no-arbitrage affine term structure models

A longstanding empirical challenge for affine no-arbitrage term structure models is their difficulty in simultaneously capturing the cross-sectional and time-series dynamics of bond yields while also generating plausible time-variation in yields volatility. As observed by Dai and Singleton (2000) and Dai and Singleton (2002), this is the "tension in matching simultaneously the historical properties of the conditional means and variances of yields." Similarly, Duffee (2002) notes that the functional forms for risk premia are insufficiently flexible and, as a consequence, the overall goodness of fit "is increased by giving up flexibility in forecasting to acquire flexibility in fitting conditional variances."

This tension is perhaps most transparently revealed through two regression-based diagnostics introduced by Dai and Singleton (2002) as the LPY(i) and LPY(ii) regressions. In particular, LPY(i) assesses whether a model's implied slope coefficients in Campbell and

Shiller (1991) regressions – used to predict yield changes based on the slope of the yield curve – align with the empirical patterns observed across different maturities. LPY(ii) evaluates whether, after adjusting yield changes for model-implied term premia, the regression slopes converge to unity, as predicted by the expectation hypothesis.

The remainder of this section explores potential reasons why existing models with stochastic volatility fail to satisfy the LPY diagnostics. Our objective is to identify insights that can guide the design of a new stochastic volatility term structure model capable of passing both LPY(i) and LPY(ii) tests. These insights will directly inform the specification introduced in Section 3.

2.1 LPY(i): Yield changes and the slope of the yield curve

The LPY(i) regression focuses on the predictive power of the yield curve slope for future changes in yields, based on the Campbell and Shiller (1991) regression:

$$R_{t+1}^{n-1} - R_t^n = \alpha + \beta \cdot \frac{1}{n-1}(R_t^n - r_t) + \varepsilon_t \tag{1}$$

where R_t^n denotes the continuously compounding yield on a risk-free n-period zero coupon bond and $r_t = R_t^1$ is the one period short rate.

If the expectation hypothesis holds, the regression coefficient β should equal unity for all maturities n. However, empirical estimates of β typically deviate markedly from this prediction – often negative and declining with maturity – contradicting the expectation hypothesis. The negative coefficients indicate that a steeper yield curve (i.e. a higher slope) tends to predict declines in future yields. Moreover, the downward-sloping pattern of β across maturities is consistent with the mean-reverting nature of the slope: a higher slope today tends to lead to lower future slopes. This effect becomes more pronounced at longer maturities, as long-term yields by construction load more heavily on the slope factor. Consequently, the β coefficients become increasingly negative with maturity.

To replicate this empirical pattern, a model must possess sufficient flexibility to allow current slopes to predict future movements in level factors. It must also be able to generate the appropriate degree of mean reversion in the slope factor to match the observed term structure of β coefficients.

In practice, standard no-arbitrage affine models with stochastic volatility often struggle to reproduce the empirical regression coefficients. Interestingly, models that come closest to matching the observed patterns are pure Gaussian specifications, which assume constant volatility. As is shown by Dai and Singleton (2002), introducing higher dimensional volatility dynamics tends to hinder, rather than help, the model's ability to capture this feature of the data.

Joslin and Le (2021) show that the mechanism underlying this tension can be understood by studying what Dai and Singleton (2000) call *admissibility restrictions*. These are restrictions necessary to ensure that the volatility factors remain strictly positive with probability one. In an affine setting, admissibility rules out any linear dependence of the volatility forecasts on non-volatility factors because such linear dependence can imply negative volatility forecasts.

For example, an admissible affine model with one volatility factor, V_t, and one conditional Gaussian factor, G_t, cannot allow the conditional expectation from taking the linear form $E_t[V_{t+1}] = a + bV_t + cG_t$ for any non-zero c since this implies a positive probability that the volatility forecast, $E_t[V_{t+1}]$, might be negative. This means that, to remain strictly positive, the conditional means of the volatility variables can only depend on their own lagged values. In other words, the volatility factors must have an autonomous structure.

To see how an autonomous structure for the volatility factors can be restrictive, recall that, except for knife-edge cases, one can always express the volatility factor as a linear function of the principal components of yields (level, slope in our example). Let $\mathcal{P}_t$ denote a 2-element vector obtained by stacking up the slope, level factors, we write:

$$V_t = \alpha + \beta \mathcal{P}_t.$$

Additionally, due to the affine structure of the model, the conditional means of $\mathcal{P}_t$ must take a linear form:

$$E_t[\mathcal{P}_{t+1}] = K_0 + K_1 \mathcal{P}_t.$$

For V_t to be autonomous, certain restrictions need to be imposed on the feedback matrix K_1 and the loading vector β. For example, consider fixing β at $(1, 0)$ so that V_t corresponds to the first entry of $\mathcal{P}_t$ – the level factor. In this case, the autonomous restriction requires that the level factor follow an AR(1) process. This requires that the (1,2) entry of the feedback matrix K_1 constrained to zero, precluding the slope from predicting the level. Such a constraint may be overly restrictive, as matching Campbell and Shiller (1991) evidence does require that the slope factor, in fact, forecast future changes in level.

More generally, it can be shown that the restriction between K_1 and β takes the form: $\beta K_1 = c\beta$ where c is some scalar meaning that β must be a left eigenvector of the feedback matrix K_1.[1]

A risk-neutral analogue of this condition must also hold, as admissibility must be satisfied under both the physical and the risk-neutral measures. That is, β must also be a left eigenvector of the risk-neutral feedback matrix $K_1^{\mathbb{Q}}$, where $E_t^{\mathbb{Q}}[\mathcal{P}_{t+1}] = K_1^{\mathbb{Q}} \mathcal{P}_t + \text{constant}$.[2] Enforcing admissibility, therefore, amounts to requiring K_1 and $K_1^{\mathbb{Q}}$ to share a common left eigenvector.

This requirement introduces a fundamental tension. As Joslin and Le (2021) emphasize, the cross-section of bond prices provides sharp identification of risk-neutral dynamics, meaning that $K_1^{\mathbb{Q}}$ is tightly estimated with little reliance on time-series data. When the common-eigenvector restriction is binding, as often is the case in the data, then the estimate of the physical dynamics K_1 must accommodate the empirically accurate $K_1^{\mathbb{Q}}$, even at the cost of poor time-series fit. The distorted time-series estimates in turn lead the model to fail in reproducing the observed pattern of β coefficients in the Campbell and Shiller (1991) regression.

[1] The AR(1) structure of V_t requires that $E_t[V_{t+1}] = cV_t + \text{constant}$ for some scalar c. Substitute V by $\beta\mathcal{P}$, it follows that $\beta E_t[\mathcal{P}_{t+1}] = c\beta\mathcal{P}_t + \text{constant}$. Using $E_t[\mathcal{P}_{t+1}] = K_0 + K_1\mathcal{P}_t$, we obtain $\beta K_1 = c\beta$.

[2] Since volatility must remain positive under both probability measures, V_t must be autonomous under the equivalent-martingale measure as well. Hence, the loading vector β must be a left eigenvector of $K_1^{\mathbb{Q}}$.

This insight suggests that a promising strategy for passing the LPY(i) diagnostic is to relax the tight linkage between K_1 and $K_1^{\mathbb{Q}}$. In fact, in pure Gaussian term structure models – where time variation in volatility is turned off – the admissibility constraint is non-binding. As a result, the eigenvector restriction that links K_1 and $K_1^{\mathbb{Q}}$ is absent. These models have been particularly successful at reproducing the empirical LPY(i) patterns. However, their assumption of constant volatility is clearly counter-factual. The key challenge, then, is to construct a no-arbitrage model that allows for realistic volatility dynamics while retaining enough flexibility to match the observed LPY(i) coefficients. We propose such a model in Section 3.

2.2 LPY(ii): premium adjusted yield changes and the slope of the yield curve

Since the empirical LPY(i) pattern indicates a departure from the expectation hypothesis, Dai and Singleton (2002) argue that properly adjusting bond yields for time-varying risk premia should bring the regression coefficients β closer to unity, as predicted by the expectation hypothesis. Building on this logic, they introduce a premium-adjusted version of the LPY(i) regression – referred to as the LPY(ii) regression:

$$R_{t+1}^{n-1} - R_t^n + \frac{nTP_t^n - (n-1)TP_{t+1}^{n-1}}{n-1} = \alpha + \beta \cdot \frac{1}{n-1}(R_t^n - r_t) + \varepsilon_t \tag{2}$$

where the term premium TP_t^n is defined as

$$TP_t^n = R_t^n - \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{E}_t[r_{t+i}]. \tag{3}$$

If the estimated β coefficients in equation (2) are close to unity, it provides evidence that the model is correctly capturing time-varying risk premia.

In differentiating the contributions of the two diagnostics, Dai and Singleton (2002) state the following:

> *Matching LPY(i) says that the DTSM describes the historical behavior of yields under the actual measure P; while matching LPY(ii) says that the DTSM essentially has the dynamics "right" under the risk-neutral measure Q used in pricing bonds. We show that these properties are not equivalent, and that the requirement that a DTSM match both is a powerful discriminator among models.*

To understand why matching LPY(ii) implies that a model generates the "correct" risk-neutral dynamics, it is essential to recall the fundamental linkage among risk premia, historical dynamics, and risk-neutral dynamics. Intuitively, risk premia arise from the difference between the $\mathbb{P}$- and $\mathbb{Q}$-dynamics. Therefore, if a model accurately captures any two of the three components, the third is essentially determined. For instance, if the model correctly specifies the historical dynamics – thereby matching the LPY(i) regression – and also generates

realistic term premia – thus matching LPY(ii) – then it must necessarily produce the correct $\mathbb{Q}$-dynamics. In this sense, the interpretation that success in capturing LPY(ii) reflects accurate $\mathbb{Q}$-dynamics implicitly relies on the assumption that the model already successfully matches LPY(i).

But what should we make of a model that matches LPY(ii) while failing to match LPY(i)? This situation arises in the estimated $A_1(3)$ model reported by Dai and Singleton (2002). While the model succeeded in producing LPY(ii) coefficients that were somewhat close to unity, it failed to reproduce the empirical LPY(i) regression coefficients, indicating a poor fit for the historical dynamics. This raises an important question: can evidence from LPY(ii) be meaningfully interpreted on its own, or does its validity necessarily depend on the model first passing LPY(i)?

To this extent, we propose a new perspective on studying the LPY(ii) regression. First, note that with some algebra we can rewrite the left hand side of LPY(ii) as:[3]

$$R_{t+1}^{n-1} - R_t^n + \frac{nTP_t^n - (n-1)TP_{t+1}^{n-1}}{n-1} = \frac{1}{n-1}(R_t^n - r_t) + \frac{1}{n-1}\sum_{i=1}^{n-1}(\mathbb{E}_{t+1} - \mathbb{E}_t)[r_{t+i}]. \quad (4)$$

Equation (4) suggests that the premium adjusted yield change used in LPY(ii) is simply the sum of the yield curve slope and a forecast revision term. As a result, an alternative implementation of LPY(ii) can be obtained through the following regression:

$$\frac{1}{n-1}\sum_{i=1}^{n-1}(\mathbb{E}_{t+1} - \mathbb{E}_t)[r_{t+i}] = \alpha + (\beta - 1)\cdot\frac{1}{n-1}(R_t^n - r_t) + \varepsilon_t. \quad (5)$$

In this form, we see that LPY(ii) is really a regression of slope predicting revisions in the forecasts. This representation provides a seemingly straightforward interpretation of the LPY(ii) evidence. When the estimated β coefficients deviate from unity, it indicates that revisions in short-rate forecasts are predictable from the slope of the yield curve. This predictability implies that the model-implied forecasts are inconsistent with full-information rational expectations, as they can be improved by incorporating information embedded in the slope. Conversely, when the β estimates are close to unity, it suggests that the model has efficiently utilized the slope factor in forming its short-rate forecasts.

[3]The log excess return on an n-period bond can be expressed as:

$$\log \text{excess return}_{t+1}^n = nR_t^n - (n-1)R_{t+1}^{n-1} - r_t.$$

This expression can be rewritten as:

$$nTP_t^n - (n-1)TP_{t+1}^{n-1} + \sum_{i=1}^{n-1}(\mathbb{E}_t - \mathbb{E}_{t+1})[r_{t+i}].$$

To derive this second expression, we substitute the decomposition of yields into their term premium and expected short-rate components, as given by equation (3), into the original expression. The equivalence between the two lines yields the identity shown in equation (4).

Nevertheless, this simple interpretation of LPY(ii) is complicated by the fact that, in practice, equation (5) uses a model-implied forecast revision on the right-hand side, while the regressor (the slope) is directly observed from the data.[4] Consequently, we must consider the possibility that the model may not adequately capture the slope factor, and thus may misrepresent the right-hand side variable in (5).

We consider the following possibilities:

1. Suppose the state vector X_t does not accurately capture the slope factor. Further, assume that forecasts of X_{t+n} are fully flexible. In this case, conditioning on the slope will generally improve forecast accuracy and allow for the prediction of forecast revisions. Specifically, the difference $E_{t+1}[X_{t+s}] - E_t[X_{t+s}]$ is not predictable using information in X_t alone, but becomes predictable when incorporating the slope, since the slope lies outside the information set spanned by X_t.

   This scenario is particularly relevant for macro-finance term structure models that include perfectly observed macroeconomic variables and a single latent factor. These models are studied in Joslin, Le, and Singleton (2013). Due to the presence of only one latent variable, such models typically succeed in capturing the level factor but fail to accurately price slopes. As a result, while these models may match the LPY(i) regression, they are generally unable to match the LPY(ii) regression.

2. Suppose a model features a state vector X_t that accurately captures the slope factor. Further, assume that forecasts of X_{t+n} are fully flexible, meaning there are no restrictions on identifying K_1. Under these conditions, the model's forecasts of the short rate will optimally incorporate information from the slope factor. This optimality implies that slopes cannot be used to predict revisions in short-rate forecasts, and therefore, the LPY(ii) will hold. In this case, the estimated β coefficients will be close to unity.

3. Suppose the model adequately captures the slope factor but fails to generate fully flexible forecasts, for example due to the left eigenvector constraint discussed in Joslin and Le (2021). In this case, the implications for LPY(ii) are ambiguous. Generally, LPY(ii) should not work since one would expect the inefficient model-implied forecasts revisions can be improved by incorporating slopes. This is the case for many $A_M(N)$ models with $M > 1$ as reported by Dai and Singleton (2002).

   However, there is yet another possibility. The model may produce incorrect forecasts, but the forecast revisions–despite being incorrect–may still appear uncorrelated with the slope, simply because the slope is not informative about the model's misspecified forecast errors.

   To illustrate this possibility, suppose the estimated model implies the following empirical pattern:

$$\frac{1}{n-1}\sum_{i=1}^{n-1}(\mathbb{E}_{t+1} - \mathbb{E}_t)[r_{t+i}] = a + b \cdot \text{curvature}_t + \text{noise}_{t+1}. \tag{6}$$

[4] This issue does not arise with LPY(i), since the LPY(i) regression is conducted entirely within the model in population.

This specification clearly violates the full-information rational expectations principle, as forecast revisions are predictable (by curvature). However, when we implement the LPY(ii) regression in equation (5) in the context of the model in equation (6), we effectively project curvature on slope. Since curvature and slope are orthogonal in sample, by construction, this projection yields a zero coefficient on slope–i.e., $(\beta - 1) = 0$, implying $\beta = 1$.

In this case, the model will appear consistent with LPY(ii) not because it generates efficient forecasts, but because the slope factor is empirically uncorrelated with the variable that contains predictive power for its forecast revisions.

Turning back to the $A_1(3)$ model reported in Dai and Singleton (2002), it is notable that LPY(i) does not work but LPY(ii) appears to work reasonably well. How can this be? In the LPY(i) regression, the model's estimated coefficients are close to unity, indicating that the slope factor does not meaningfully predict future level factors. That is, the left-eigenvector constraint induced by the admissibility restrictions makes it so that that the feedback from slope to level is much weaker than required by the data. However, this very disconnect between today's slope and future levels may explain why LPY(ii) appears to work. The model optimal forecast of the change in level does not depend on the slope factor. Consequently, when we regress revisions in the model's optimal forecasts of the short rate on the slope, the estimated coefficient is zero. This leads LPY(ii) to "work" in the model even as LPY(i) fails.

This example highlights a rather simple but important point: LPY(ii) singularly focuses on the empirical slope factor in testing the efficiency of a model's implied forecasts. However, a more comprehensive diagnostic can be obtained by incorporating all the relevant factors. Specifically, consider the regression:

$$\frac{1}{n-1}\sum_{i=1}^{n-1}(\mathbb{E}_{t+1} - \mathbb{E}_t)[r_{t+i}] = a + b \cdot \mathcal{P}_t + \text{noise}_{t+1} \tag{7}$$

where $\mathcal{P}_t$ denotes the state vector, stacking up all the relevant principal components of bond yields. The appeal of this regression lies in its unambiguous interpretation. As long as $\mathcal{P}_t$ is reasonably accurately priced (which is typically the case with $N \geq 3$), any finding that $b \neq 0$ constitutes direct evidence against the efficiency of the model-implied forecasts. On the flip side, failure to reject $b = 0$ is clear evidence supporting the model.

In the next section, we formally develop a new no-arbitrage model. A key innovation of our setup is the modeling of volatility as a nonlinear function of the state variables, which ensures positive volatility without imposing an autonomous structure on any component of the state vector. Drawing on the insights of Joslin and Le (2021), we argue that this feature helps resolve the type of tension that has plagued many existing models. We confirm this conjecture empirically and show that the proposed model offers improvements over traditional affine frameworks across several important dimensions. In particular, it delivers meaningful time-varying volatility of yields without sacrificing forecasting flexibility.

3 Model

3.1 The general setup

The state variable X_t is an $N \times 1$ vector and is assumed to follow the following dynamics under the risk-neutral ($\mathbb{Q}$) measure:

$$X_{t+1} = K_{0X}^{\mathbb{Q}} + K_{1X}^{\mathbb{Q}} X_t + \epsilon_{t+1}^{\mathbb{Q}} \text{ with } \epsilon_{t+1}^{\mathbb{Q}} \sim N(0, \Sigma_{0,X} + \sum_{i}^{M} \Sigma_{i,X} Z_{i,t}), \tag{8}$$

where the volatility factor Z_t is an $M \times 1$ vector, related to X_t through:

$$Z_t = \max(0, \alpha_X + \beta_X X_t). \tag{9}$$

We require that $\Sigma_{i,X}$ $(i = 0..M)$ be positive semidefinite. With Z_t bounded below at zero, this requirement ensures that the conditional variance of X_{t+1} will be strictly positive definite.

Assuming that the short rate is linear in the state variables:

$$r_t = \delta_0 + \delta_X \cdot X_t, \tag{10}$$

the time-t price of an n-period bond is given by:

$$P_{n,t} = E_t^{\mathbb{Q}}[exp(-r_t - r_{t+1} - ... - r_{t+n-1})]. \tag{11}$$

To complete the model, we assume that the time series dynamics of X_t is given by:

$$X_{t+1} = K_{0X}^{\mathbb{P}} + K_{1X}^{\mathbb{P}} X_t + \epsilon_{t+1} \text{ with } \epsilon_{t+1} \sim N(0, \Sigma_{0X} + \sum_{i}^{M} \Sigma_{i,X} Z_{i,t}). \tag{12}$$

Several remarks about our modeling choices are in order. First, volatility in this model – driven by $\max(0, \alpha_X + \beta_X X_t)$ – is guaranteed to remain non-negative for all realizations of the state vector X_t. As a result, admissibility conditions are naturally satisfied without any additional structure on the conditional dynamics of of X_t (under either $\mathbb{P}$ or $\mathbb{Q}$). In contrast to many existing affine models with stochastic volatility, our specification does not impose any restrictions linking the volatility loading vector, β_X, to the feedback matrices, $K_{1X}^{\mathbb{P}}$ and $K_{1X}^{\mathbb{Q}}$. This separation offers greater modeling flexibility, enabling the conditional mean of yields to be fit without compromising the model's ability to capture conditional volatility.

Second, our framework nests the Gaussian term structure models as a special case. Specifically, by setting $\Sigma_{i,X}(i = 1..M)$, α_X, and β_X, to zeros, the time varying component of volatility is turned off. With constant volatility, our model reduces exactly to the Gaussian setup of Joslin, Singleton, and Zhu (2011). This nesting property will allow us to conduct formal statistical tests that compare the fit of our model relative to the Gaussian benchmark.

Third, because volatility is modeled nonlinearly, the conditional variances in our setup are no longer strictly affine functions of the state variables. This departure from linearity precludes strictly closed-form solutions for bond prices. In the next subsection, we introduce an approximation method that yields bond prices that are exponentially affine in the state vector X_t, thus preserving analytical tractability to a useful extent.

3.2 Approximate bond prices

The only source of nonlinearity in our model arises from the specification of Z_t:

$$Z_t = \max(0, \alpha_X + \beta_X X_t). \tag{13}$$

To simplify the analysis, we begin by assuming that $\alpha_X + \beta_X X_t$ rarely falls below zero, such that the nonlinearity introduced by the $\max(\cdot)$ operator is seldom binding. Under this assumption, a natural approximation is to treat Z_t as a linear function of the state vector X_t, i.e.,

$$Z_t \approx \alpha_X + \beta_X X_t. \tag{14}$$

This approximation restores linearity to the conditional means and variances of X_t, effectively returning us to the affine class of models where standard bond pricing results apply. In particular, it can be shown that the n-period bond yield is approximately linear in the state variables:

$$y_{n,t} \approx A_n + B_n X_t, \tag{15}$$

where A_n and B_n are determined by standard recursive equations.[5]

If, however, the probability that $\alpha_X + \beta_X X_t$ becomes negative is non-negligible, then bond yields implied by the model will be substantially nonlinear in the state vector, rendering the linear approximation inadequate. To avoid this complication, we restrict the parameter space to ensure that the probability of negative realizations of $\alpha_X + \beta_X X_t$ is small. This allows us to maintain the approximate linear form in (15), which greatly facilitates empirical implementation.

We acknowledge that this modeling choice limits us to a subset of the general class of models introduced in Section 3.1. Thus, our approach reflects a tradeoff between implementation tractability and generality. Importantly, we will demonstrate that this tradeoff is well-justified: the models we study, although "almost affine," offer substantially greater richness along several dimensions than many existing affine term structure models. Fully exploring the more nonlinear regions of the model's parameter space is potentially promising but left for future research.

In practice, there are various ways to ensure that the probability of negative realizations of $\alpha_X + \beta_X X_t$ remains low. For instance, one approach is to constrain model parameters so that such realizations occur less than, say, 1% of the time. For tractability, we estimate our models under the constraint that $\alpha_X + \beta_X X_t$ remains nonnegative in-sample. Additionally, to avoid undesirable behaviors under the pricing measure, we require that the $\mathbb{Q}$-dynamics be non-explosive.

Similar approximate arguments have been used in many related contexts, though their accuracy can vary depending on the specific application. In our case, we systematically assess

[5] To quickly see this, note that X_t is conditionally Gaussian thus its conditional Laplace transform is exponentially linear in its means and variances. Furthermore, the first two conditional moments of X_{t+1} are both linear in X_t, thus we can write: $E_t^{\mathbb{Q}}[exp(u \cdot X_{t+1})] = exp(a(u) + b(u) \cdot X_t)$ where $a(.)$ and $b(.)$ depend on other risk neutral parameters. Standard risk-neutral pricing then shows that:

$$A_n = \delta_0 + A_{n-1} - a(-B_{n-1}), \text{ and } B_n = \delta_X - b(-B_{n-1}).$$

the quality of the approximation and find that, across a range of parameterizations and yield maturities, the approximation errors are economically negligible – typically on the order of 1 to 2 basis points.

3.3 Econometric identification

A key feature of the model introduced in Section 3.1 is the flexibility to shift, rotate, and scale the state vector X_t to obtain observationally equivalent variants of the model.[6] We exploit this property to impose normalizations that enable econometric identification. Importantly, the ability to linearly transform the state vector is independent of whether we adopt a linear or nonlinear specification. Hence, the identification strategy applies even to nonlinear variants of the model.

To be specific, let Θ denote the full set of model parameters:

$$\Theta = \left(K_{0X}^{(j)}, K_{1X}^{(j)}, \Sigma_{i,X}, \alpha_X, \beta_X, \delta_0, \delta_X\right),$$

where $i = 0, \ldots, M$ and $j \in \{\mathbb{P}, \mathbb{Q}\}$. Consider a linear transformation of the state vector of the form

$$\widetilde{X}_t = U_0 + U_1 X_t.$$

Since the state space is unbounded, this transformation is unrestricted—just as in standard Gaussian models—implying no sign or zero constraints on U_0 and U_1. The transformed parameter set, $\widetilde{\Theta} = \left(\widetilde{K}_{0X}^{(j)}, \widetilde{K}_{1X}^{(j)}, \widetilde{\Sigma}_{i,X}, \widetilde{\alpha}_X, \widetilde{\beta}_X, \widetilde{\delta}_0, \widetilde{\delta}_X\right)$, is related to the original parameters via the following mappings:

$$\widetilde{K}_{0X}^{(j)} = U_0 + U_1 K_{0X}^{(j)} - (U_1 K_{1X}^{(j)} U_1^{-1}) U_0, \tag{16}$$

$$\widetilde{K}_{1X}^{(j)} = U_1 K_{1X}^{(j)} U_1^{-1}, \tag{17}$$

$$\widetilde{\Sigma}_{i,X} = U_1 \Sigma_{i,X} U_1', \quad \text{for } i = 0, \ldots, M, \tag{18}$$

$$\widetilde{\alpha}_X = \alpha_X - \beta_X U_1^{-1} U_0, \quad \widetilde{\beta}_X = \beta_X U_1^{-1}, \tag{19}$$

$$\widetilde{\delta}_0 = \delta_0 - \delta_X U_1^{-1} U_0, \quad \widetilde{\delta}_X = \delta_X U_1^{-1}. \tag{20}$$

Econometric identification is achieved by imposing restrictions that eliminate this invariance such that further shifts, rotations, or scalings of X_t would violate the constraints. Following Joslin, Singleton, and Zhu (2011), we impose the following normalizations on the risk-neutral dynamics:

1. $K_0^{\mathbb{Q}} = 0$;
2. $K_1^{\mathbb{Q}}$ is in Jordan canonical form;
3. δ_X is a vector of ones.

[6] Observationally equivalent models are models whose implied yields are identical state by state. For details, see Duffie and Kan (1996) and Dai and Singleton (2000).

The first condition centers the risk-neutral dynamics at the origin. The second prevents arbitrary rotations of the state vector, while the third fixes the scale of the state variables. With these normalizations, the long-run risk-neutral mean of X_t is zero, and the intercept in the short rate equation,

$$r_t = r_\infty^{\mathbb{Q}} + \iota' X_t, \tag{21}$$

where ι is a vector of ones, corresponds to the long-run risk-neutral mean of the short rate, which we denote as $r_\infty^{\mathbb{Q}} \equiv \delta_0$.

Up to this point, our normalization scheme mirrors that of Joslin, Singleton, and Zhu (2011) for Gaussian models. However, our setting includes a volatility component Z_t, which requires additional restrictions. In particular, Z_t can be freely scaled: for instance, doubling both α_X and β_X while halving all $\Sigma_{i,X}$ (for $i = 1, \ldots, M$) leaves the conditional variance of X_t unchanged, state by state. Thus, to ensure full identification, we must also fix the scale of Z_t.

The precise way we normalize Z_t will be discussed in Section 5, where we describe our empirical implementation.

3.4 Notation

To distinguish our framework from existing models and emphasize its "almost affine" structure, we refer to our specification as $AA_M(N)$ where N and M are the dimensions of state variables X_t and the volatility factors Z_t, respectively. In contrast, to reference the existing no-arbitrage affine models, we use the standard notation $A_M(N)$. In particular, $A_0(N)$ refers to Gaussian affine models with constant volatility while $A_M(N)$ with $M > 0$ corresponds to affine models with stochastic volatility.

Throughout the remainder of the paper, we will work extensively with the approximately linear yield pricing equation introduced in equation (15). For brevity, we will refer to this as the "linear pricing equation," omitting the qualifier "approximately." It should be understood, however, that this expression is an approximation. We will later verify that the resulting approximation errors are economically negligible.

4 The no-arbitrage constraints and yield dynamics

One important issue that has been studied extensively in the term structure literature is whether the imposition of the no-arbitrage conditions can lead to improvement in forecasting performance. In the context of Gaussian term structure models, Joslin, Singleton, and Zhu (2011) and Duffee (2011) derive an important result: that the no-arbitrage constraints are essentially inconsequential for model-implied yield forecasts. By contrast, Joslin and Le (2021) find that for the $A_M(N)$ models with $M > 0$, the no-arbitrage conditions have an overarching impact on the dynamics of yields. In the remainder of this section, we study the implications of no-arbitrage for our model.

4.1 Rotation from latent states to the principal components of bond yields

The role of no-arbitrage in term structure models is reflected through a set of cross-sectional restrictions imposed on the yield pricing equation. Consider a vector version of equation (15): $Y_t = A_X + B_X X_t$ where Y_t is any vector of yields. Under the no-arbitrage condition, the loading matrices A_X and B_X are not freely estimated parameters. Instead, they are constrained by a set of cross-equation restrictions that are functions of the risk-neutral parameters $\Theta^{\mathbb{Q}}$, which govern the risk-neutral dynamics of X_t.

In models where the states are latent, these latent factors are typically inferred from the cross-section of yields using the pricing equation above. Specifically, the latent state X_t can be extracted from any N linear combinations of yields and the associated loading functions, $A_X(\Theta^{\mathbb{Q}})$ and $B_X(\Theta^{\mathbb{Q}})$. As a result, the risk-neutral parameters $\Theta^{\mathbb{Q}}$, and thus the no-arbitrage condition itself, directly influence how the latent states are identified from the data.

Joslin, Singleton, and Zhu (2011) show that one can conveniently study the role of no-arbitrage by rotating the model to an observationally equivalent representation in which the state variables are portfolios of yields. When the states are observable in this way, their identification becomes unambiguous, and the no-arbitrage restrictions affect only the cross-sectional structure of the pricing equation, not the identification of the states themselves.

This rotation is made possible by the linearity of yields in the state vector. Let $\mathcal{P}_t = WY_t$ be a vector of N-yield portfolios, where W is a fixed matrix of loadings. Then $\mathcal{P}_t$ remains a linear function of X_t: $\mathcal{P}_t = WA_X + WB_X X_t$. By applying the linear transformation outlined in equations (16-20) (with $U_0 = WA_X$ and $U_1 = WB_X$), we rotate the model from latent states X_t to the observable yields portfolios $\mathcal{P}_t$.

In this rotated representation, the pricing equation becomes:

$$Y_t = A_{\mathcal{P}} + B_{\mathcal{P}}\mathcal{P}_t \tag{22}$$

and the state dynamics of $\mathcal{P}_t$ follow:

$$\mathcal{P}_{t+1} = K^{\mathbb{P}}_{0\mathcal{P}} + K^{\mathbb{P}}_{1\mathcal{P}}\mathcal{P}_t + \epsilon_{\mathcal{P},t} \text{ with } \epsilon_{\mathcal{P},t} \sim N(0, \Sigma_{0\mathcal{P}} + \sum_{i}^{M} \Sigma_{i,\mathcal{P}} Z_{i,t}). \tag{23}$$

where the volatility factor is given by $Z_t = \max(0, \alpha_{\mathcal{P}} + \beta_{\mathcal{P}}\mathcal{P}_t)$.[7]

Joslin, Le, and Singleton (2012) propose a particularly useful choice for the loading matrix W: selecting $\mathcal{P}_t$ to correspond to the first N principal components of bond yields. They show that this choice results in highly accurate identification of $\mathcal{P}_t$ even when individual yields are measured with noise. The intuition is that the principal components of bond yields represent "well-diversified" portfolios of yields, which naturally average out idiosyncratic observational errors associated with individual bonds. Indeed, Joslin, Le, and Singleton (2012) demonstrate that estimates obtained under the assumption that $\mathcal{P}_t$ is observed perfectly

[7] It can be shown that $B_{\mathcal{P}} = B_X(WB_X)^{-1}$ and $A_{\mathcal{P}} = A_X - B_{\mathcal{P}}WA_X$. Other parameters are derived using equations provided by (16-20).

are nearly identical to those obtained via Kalman filtering when allowing for noisy yield observations.[8]

Motivated by these findings, we adopt this approach in our analysis. For the remainder of the paper, we assume that $\mathcal{P}_t$ consists of the first N principal components of the yield curve. We further assume that these portfolios are observed without measurement error.

4.2 The impact of no-arbitrage on model-implied forecasts

Due to the VAR(1) structure of the time series dynamics of $\mathcal{P}_t$ in (23), the conditional forecasts of $\mathcal{P}_{t+h}$ at any horizon h, are dependent on the conditional means parameters, $K^{\mathbb{P}}_{0\mathcal{P}}$ and $K^{\mathbb{P}}_{1\mathcal{P}}$, the current value of $\mathcal{P}_t$ and the horizon h. Since $\mathcal{P}_t$ is observed, the impact of no-arbitrage on model-implied forecasts of the factors $\mathcal{P}_{t+h}$ must be through its impact on the estimates of $K^{\mathbb{P}}_{0\mathcal{P}}$ and $K^{\mathbb{P}}_{1\mathcal{P}}$.

It is important to note that $K^{\mathbb{P}}_{0\mathcal{P}}$ and $K^{\mathbb{P}}_{1\mathcal{P}}$ are pure time-series parameters. These parameters only govern the time-series evolution of the factors $\mathcal{P}_t$ but do not appear in the pricing equation (22). As such, they are completely inconsequential for the pricing performance of the model can be estimated from the time series dynamics of $\mathcal{P}_t$, independent of the no-arbitrage condition. Specifically, equation (23) implies that $K^{\mathbb{P}}_{0\mathcal{P}}$ and $K^{\mathbb{P}}_{1\mathcal{P}}$ can be estimated via a generalized least squares (GLS) regression of $\mathcal{P}_{t+1}$ onto $\mathcal{P}_t$, accounting for the heteroskedasticity in the innovations. Denote the conditional variance of the time series innovation $\epsilon_{\mathcal{P},t}$ by Σ_t $(= \Sigma_{0\mathcal{P}} + \sum_i^M \Sigma_{i,\mathcal{P}} Z_{i,t})$, the GLS estimates can be obtained as:

$$vec\left(\widehat{[K^{\mathbb{P}}_{0\mathcal{P}}, K^{\mathbb{P}}_{1\mathcal{P}}]}\right) = E_T[\tilde{\mathcal{P}}_t\tilde{\mathcal{P}}_t' \otimes \Sigma_t^{-1}]^{-1} E_T[\tilde{\mathcal{P}}_t\mathcal{P}_{t+1}' \otimes \Sigma_t^{-1}]\, vec(I_N), \tag{24}$$

where $\tilde{\mathcal{P}}_t = (1, \mathcal{P}_t')'$; $E_T[.]$ denotes sample averages; and I_N denotes an $N \times N$ identity matrix.

Equation (24) generalizes the ordinary least squares (OLS) estimator to account for time-varying volatility. In the special case where volatility is constant (by setting $\Sigma_{i\mathcal{P}}$ $(i = 1..M)$ or $\beta_{\mathcal{P}}$ to zeros), the expression simplifies considerably. Due to the constancy of Σ_t, the two inverse variance terms (Σ_t^{-1}) can be brought outside of the $E_T[.]$ operators and perfectly cancel, yielding the standard OLS formula::

$$\widehat{[K^{\mathbb{P}}_{0\mathcal{P}}, K^{\mathbb{P}}_{1\mathcal{P}}]} = E_T[\mathcal{P}_{t+1}\tilde{\mathcal{P}}_t'] E_T[\tilde{\mathcal{P}}_t\tilde{\mathcal{P}}_t']^{-1}. \tag{25}$$

Equation (25) underpins one of the main results of Joslin, Singleton, and Zhu (2011) in the constant volatility setting: that the no-arbitrage constraint has no bearing on the forecasts of yield factors. Since the risk-neutral parameters in $\Theta^{\mathbb{Q}}$ do not appear in (25), the no-arbitrage restriction cannot influence the estimates of $K^{\mathbb{P}}_{0\mathcal{P}}$ and $K^{\mathbb{P}}_{1\mathcal{P}}$ and hence has no effect on the model's factor forecasts.

However, when volatility is time-varying, the invariance result no longer holds. Equation (24) shows that the estimation of $K^{\mathbb{P}}_{0\mathcal{P}}$ and $K^{\mathbb{P}}_{1\mathcal{P}}$ is influenced by the time variation in Σ_t. The

[8] One caveat is that Joslin, Le, and Singleton (2012) study Gaussian term structure models with constant volatility. However, their argument (that "well-diversified" portfolios of yields benefit from the cancellation of observational errors of individual yields) is likely to hold for different volatility structures.

parameters driving this variation, such as $\Sigma_{i,\mathcal{P}}$'s and $\beta_{\mathcal{P}}$, also enter the pricing equation (22). Therefore, enforcing the no-arbitrage constraint, which imposes cross-sectional restrictions through equation (22), provides a source of identification for these variance parameters. In turn, this can have an impact on the estimates of $K_{0\mathcal{P}}^{\mathbb{P}}$ and $K_{1\mathcal{P}}^{\mathbb{P}}$.

How economically significant is this impact? The answer depends on how strongly the cross-sectional pricing restrictions identify the variance parameters. While this is ultimately an empirical question, a reasonable conjecture is that the no-arbitrage condition plays a relatively minor role. The rationale is that the variance parameters typically affect yield pricing through the Jensen effect whose economic magnitude is often small.

Thus far, our discussion has focused on forecasts of the state vector $\mathcal{P}_t$. To connect this discussion to the forecasts of yields, we turn to the pricing equation (22), which relates yields Y_t to $\mathcal{P}_t$. For example, the h-period-ahead forecast of the state vector $E_t[\mathcal{P}_{t+h}]$ can be translated to forecasts of yields simply by $A_{\mathcal{P}} + B_{\mathcal{P}} E_t[\mathcal{P}_{t+h}]$. This relationship implies that the no-arbitrage condition can influence yield forecasts either through its impact on the factor forecasts or through the pricing loadings $A_{\mathcal{P}}$ and $B_{\mathcal{P}}$. However, prior work (e.g., Duffee (2011)) has shown that the no-arbitrage constraint typically has only a limited effect on the estimates of these loadings. In practice, the loadings implied by no-arbitrage affine models tend to be very similar to those obtained from unconstrained regressions of yields on the factors.

4.3 Summary and comparison to other no-arbitrage affine models

We conclude this section by summarizing our findings and drawing comparisons between our framework and other no-arbitrage affine term structure models. As discussed in Section 2, standard affine models with stochastic volatility are subject to a structural constraint implied by no-arbitrage: the feedback matrices governing the dynamics under the $\mathbb{P}$ and $\mathbb{Q}$ measures must share a common left-eigenvector basis. Joslin and Le (2021) show that this condition places a strong restriction on the estimation of the time-series dynamics, limiting the flexibility of such models in capturing empirical regularities.

In contrast, our model ($AA_M(N)$) shares key features with the Gaussian affine class ($A_0(N)$) which assumes constant volatility. Both models are free from the left-eigenvector restriction: the feedback matrices $K_{1\mathcal{P}}^{\mathbb{P}}$ and $K_{1\mathcal{P}}^{\mathbb{Q}}$ are completely decoupled. More precisely, when the principal components of bond yields are used as state variables, the risk-neutral parameters have no influence on the conditional means under the physical measure, and vice versa. The parameters $K_{0\mathcal{P}}^{\mathbb{P}}$ and $K_{1\mathcal{P}}^{\mathbb{P}}$ do not appear in the pricing equations of either model, underscoring the separation between pricing and time-series dynamics.

In both the $AA_M(N)$ and $A_0(N)$ frameworks, the variance parameters do affect bond pricing through the Jensen effects. Consequently, the no-arbitrage constraint serves as one source of identification for these variance parameters. In the $AA_M(N)$ model, since the time-series parameters $K_{0\mathcal{P}}^{\mathbb{P}}$ and $K_{1\mathcal{P}}^{\mathbb{P}}$ are estimated via weighted least squares, where the weights are determined by the conditional variances, the no-arbitrage constraint can potentially have an indirect effect on the estimates of the conditional means, through the time-variation of the variances.

Model	Main channel
$A_0(N)$	No effect
$A_M(N)$ with $M > 1$	$K^{\mathbb{P}}_{1\mathcal{P}}$ and $K^{\mathbb{Q}}_{1\mathcal{P}}$ must share a common left-eigenvector basis
$AA_M(N)$ with $M > 1$	Jensen effect through the variance parameters

Table 1: How does no-arbitrage affect ML estimates of conditional means parameters?

However, in the Gaussian case where conditional variances are constant, the weighted least squares estimation simplifies to ordinary least squares. In this special case, even the indirect channel through which the no-arbitrage constraint might affect the time-series dynamics disappears entirely. A summary of our discussion in this section is provided in Table 1.

5 Empirical implementation

5.1 Data

For our empirical analysis, we employ the zero yields dataset constructed by Le and Singleton (2024). Using the CRSP database, which contains price information for individual U.S. Treasury coupon bonds, they implement the Fama-Bliss bootstrap method, following the procedure outlined in the CRSP manual, to generate a consistent panel of zero-coupon bond yields with maturities extending up to ten years. The sample spans the period from 1973 through 2024. To ensure a degree of homogeneity in the underlying instruments, standard filters are applied to exclude illiquid securities and bonds with embedded options.

For estimation, we focus on zero-coupon yields with maturities of 6 months, 1 year, 2 years, 3 years, 5 years, 7 years, and 10 years. Bonds with these maturities that are among the most actively traded in the Treasury market. Our analysis is conducted at a monthly frequency, using end-of-month observations for each yield maturity. The summary statistics of the zero yields used in estimation are provided in Table 2.

5.2 Estimation strategy

We obtain parameter estimates of our models via maximum likelihood estimation. Recall our assumption that the lower order principal components of bond yields, $\mathcal{P}_t$ are priced perfectly. We denote the remaining higher order principal components of bond yields by $\mathcal{P}^e_t$, and its corresponding loadings matrix by W^e such that $\mathcal{P}^e_t = W^e Y_t$.

Applying the pricing equation (22), $\mathcal{P}^e_t$ can be written as a linear function of $\mathcal{P}_t$: $\mathcal{P}^e_t = W^e(A_{\mathcal{P}} + B_{\mathcal{P}}\mathcal{P}_t)$. For simplicity, we adopt the common assumption that these yield portfolios are observed with independent and identically distributed Gaussian errors. Using the superscript ^{o} to denote observed quantities, we write:

$$\mathcal{P}^{e,o}_t = \mathcal{P}^e_t + e_t \text{ with } e_t \sim N(0, \sigma^2_e I). \tag{26}$$

	6-month	1-year	2-year	3-year	5-year	7-year	10-year
mean	4.66	4.84	5.03	5.21	5.51	5.73	5.92
std	3.57	3.55	3.51	3.45	3.31	3.20	3.05
min	0.02	0.06	0.11	0.12	0.23	0.38	0.53
10th percentile	0.13	0.23	0.48	0.82	1.38	1.74	2.08
median	4.93	4.97	4.92	5.06	5.37	5.60	5.69
90th percentile	9.14	9.50	9.59	9.80	9.98	10.12	10.18
max	16.26	15.81	15.64	15.54	15.28	15.04	15.04
autocorr	99.13	99.15	99.26	99.36	99.38	99.40	99.41

Table 2: Summary statistics of zero bond yields used in estimation. All numbers are in percentages.

The likelihood of the data (up to constant terms) can be written as:

$$\mathcal{L} = \sum_t \left(-\frac{1}{2}||\Sigma_t^{-1/2}(\mathcal{P}_{t+1} - K_{0\mathcal{P}}^{\mathbb{P}} - K_{1\mathcal{P}}^{\mathbb{P}}\mathcal{P}_t)||_2^2 - \frac{1}{2}log|\Sigma_t| \right.$$
$$\left. -\frac{1}{2}||(\mathcal{P}_t^{e,o} - \mathcal{P}_t^e)/\sigma_e||_2^2 - \frac{1}{2}log|\sigma_e^2 I| \right). \tag{27}$$

The first component of $\mathcal{L}$ (the first line of (27)) captures the model's time series fit, while the second component (the second line of (27)) reflects the cross-sectional fit. We obtain the estimates of the model's parameters by maximizing $\mathcal{L}$. Additional implementation details are provided in Appendix A.

5.3 Parameter estimates

In this section, we present the parameter estimates of our almost affine models alongside those of their corresponding affine counterparts. To conserve space, we focus on the three-factor specifications and relegate the estimates for the four-factor models to Appendix B.

For clarity and organization, we report the estimates of the conditional mean and conditional variance parameters separately. Table 3 summarizes the conditional mean parameters for three affine models – $A_0(3)$, $A_1(3)$, and $A_2(3)$ – as well as two almost affine models, $AA_1(3)$ and $AA_2(3)$.

A notable feature of Table 3 is the striking similarity in the estimates of the risk-neutral feedback matrix ($K_1^{\mathbb{Q}}$) across all five models. This invariance is consistent with findings reported in Joslin and Le (2021), highlighting the fact that the cross-section of bond yields, and hence the risk-neutral dynamics, is very strongly identified in the data.

Another key observation concerns the (1,2) entry of the physical feedback matrix ($K_1^{\mathbb{P}}$). In both affine models with stochastic volatility – $A_1(3)$ and $A_2(3)$ – this entry is positive, suggesting that the slope factor positively predicts the future level of interest rates. This

	Risk-neutral parameters				Time-series parameters			
	$K_0^{\mathbb{Q}}$	$K_1^{\mathbb{Q}}$			$K_0^{\mathbb{P}}$	$K_1^{\mathbb{P}}$		
	0.023	0.999	0.093	0.182	0.125	0.994	-0.019	0.027
$A_0(3)$	-0.009	-0.002	0.962	-0.180	-0.019	0.002	0.961	-0.138
	-0.012	-0.001	-0.001	0.951	-0.066	-0.000	0.002	0.866
	0.016	1.000	0.092	0.179	0.032	0.995	0.027	0.038
$A_1(3)$	-0.013	-0.002	0.963	-0.179	-0.041	-0.001	0.970	-0.206
	-0.011	-0.001	-0.001	0.952	-0.021	-0.001	-0.007	0.917
	0.018	1.000	0.092	0.181	0.083	0.997	-0.001	0.103
$AA_1(3)$	-0.011	-0.002	0.962	-0.179	0.018	0.001	0.954	-0.113
	-0.012	-0.001	-0.001	0.951	-0.019	-0.001	-0.008	0.920
	0.015	1.000	0.094	0.179	0.100	0.994	0.049	0.234
$A_2(3)$	-0.011	-0.002	0.962	-0.177	-0.005	-0.001	0.974	-0.131
	-0.010	-0.001	-0.002	0.953	-0.050	-0.001	0.003	0.894
	0.020	1.000	0.091	0.179	0.119	0.997	-0.027	0.080
$AA_2(3)$	-0.012	-0.002	0.963	-0.177	0.010	0.001	0.958	-0.116
	-0.013	-0.001	-0.001	0.952	-0.042	-0.000	-0.004	0.894

Table 3: 3-factor models: Conditional mean parameter estimates.

implication stands in contrast to the empirical evidence documented by Campbell and Shiller (1991), who find that higher slopes tend to predict lower future levels. As discussed in Joslin and Le (2021), this counter-factual result can be traced to a structural constraint imposed by the no-arbitrage condition in an affine setup, which forces $K_1^{\mathbb{Q}}$ and $K_1^{\mathbb{P}}$ to share some common left eigenvector basis.

Crucially, our almost affine models ($AA_1(3)$ and $AA_2(3)$) are designed to relax this constraint. The same is true for the purely Gaussian model $A_0(3)$. Accordingly, the (1,2) entries in the physical feedback matrices of these models are negative, aligning with the empirically observed negative predictive relationship between slope and future level. As we will demonstrate later, this structural flexibility enables our almost affine models to better capture the pattern of Campbell and Shiller (1991) regression coefficients observed in the data.

Table 4 reports the estimated parameters governing the conditional variance of bond yields implied by our models. Of particular interest are the parameters α and β, which directly identify the volatility factors.

In the affine models, the volatility factor takes the form $\alpha + \beta \mathcal{P}_t$, whereas in our almost affine models, it is specified as $\max(\alpha + \beta \mathcal{P}_t, 0)$. While we impose the condition during estimation that the $\max(\cdot)$ operator never binds in-sample – making the almost affine models

Panel A: 3-factor models with 1 volatility factor						
	$A_1(3)$			$AA_1(3)$		
α	-0.95			-0.31		
β	1.05	2.86	-6.82	1.00	-1.30	-1.62
$chol(\Sigma_0) \times 100$	24.20			36.88		
	8.15	0.09		20.22	6.63	
	7.11	-0.74	0.00	1.19	-6.48	0.00
$chol(\Sigma_1) \times 100$	18.06			22.40		
	-0.93	6.62		-3.25	5.35	
	0.17	0.02	2.98	0.86	1.82	3.72
Panel B: 3-factor models with 2 volatility factors						
	$A_2(3)$			$AA_2(3)$		
α	-1.45			-0.18		
	8.04			4.61		
β	1.06	2.98	-7.28	1.00	-1.37	-1.53
	0.99	1.97	7.42	0.99	-1.46	1.86
$chol(\Sigma_0) \times 100$	4.57			38.25		
	1.25	0.21		22.82	8.71	
	0.97	-0.62	0.00	3.29	-9.61	0.00
$chol(\Sigma_1) \times 100$	15.08			20.86		
	-2.73	5.63		-4.75	0.34	
	-1.15	-1.33	0.93	-0.14	-1.54	2.08
$chol(\Sigma_2) \times 100$	11.22			6.44		
	2.97	0.32		2.60	0.21	
	2.66	-0.94	0.00	1.92	-0.23	0.00

Table 4: 3-factor models: conditional variance parameter estimates.

appear numerically similar to the affine models – the distinction between the two classes of models is fundamental and practically consequential.

As emphasized by Joslin and Le (2021), in affine models the volatility loading vector β must satisfy a left eigenvector constraint with respect to both the risk-neutral and physical

feedback matrices. For instance, in the $A_1(3)$ model, since the risk-neutral matrix $K_1^{\mathbb{Q}}$ is tightly identified by the cross-section of bond yields, the admissible choices for β are essentially limited to the three left eigenvectors of $K_1^{\mathbb{Q}}$. This severely restricts the model's flexibility in selecting empirically optimal volatility structures.

By contrast, our almost affine models are free from this restriction. Without the left eigenvector constraint, β can be chosen from an infinite set of 3 by 1 vectors of loadings, allowing the model to better align with the empirical volatility dynamics observed in the data. While in some settings, imposing theoretically motivated restrictions may aid identification, in this context, the eigenvector constraint lacks clear economic justification and instead hampers the model's ability to identify the volatility factors that offer the best empirical fit.

This distinction is clearly reflected in the parameter estimates reported in Table 4. Panel A compares the β vectors estimated for the $A_1(3)$ and $AA_1(3)$ models. The affine model, out of the three choices of left eigenvectors of $K_1^{\mathbb{Q}}$, picks a β that combines the level and slope factors with positive weights. In contrast, the almost affine model – unconstrained in its choice – assigns a positive weight to the level factor and a negative weight to the slope factor. A similar pattern emerges in Panel B, which reports results for the two-volatility-factor models $A_2(3)$ and $AA_2(3)$. The increased flexibility of the almost affine models leads to materially different volatility factor structures that better reflect the dynamics present in the data.

The differences in volatility factor structures are further corroborated by Figure 2, which plots the one-month-ahead conditional volatility of the 5-year and 10-year yields as implied by the affine models alongside those generated by the almost affine models.

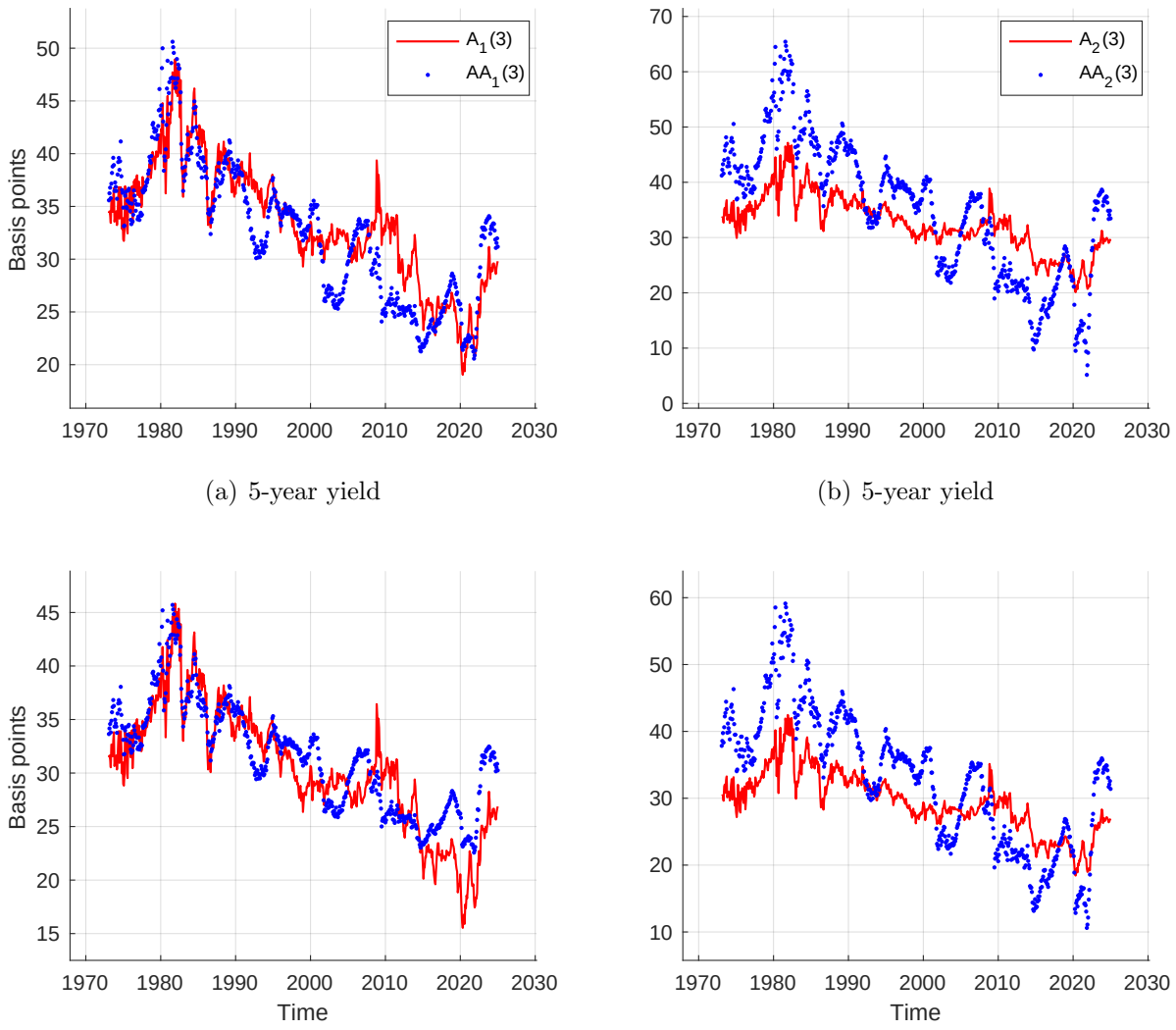
While both model classes capture broadly similar long-term trends in yield volatility, closer inspection reveals notable distinctions –especially in the two-volatility-factor specifications. The almost affine model $AA_2(3)$, benefiting from greater flexibility in the specification of volatility factors, produces yield volatilities that exhibit significantly richer time variation. In contrast, the affine counterpart $A_2(3)$, constrained by the left eigenvector condition, yields a comparatively smoother and more muted volatility profile. These differences are particularly evident during historically turbulent periods such as the early 1980s, early 2000s, and around 2020. In these episodes, the $AA_2(3)$ model exhibits sharper volatility adjustments, indicating a greater capacity to capture abrupt shifts in the macro-financial environment.

These features underscore the empirical advantage of relaxing the structural constraints embedded in affine models. By permitting a more flexible mapping between the yield factors and the volatility process, the almost affine specification more effectively captures the dynamic behavior of yield volatilities across the term structure, particularly during periods of heightened uncertainty.

5.4 Approximation quality

As discussed in Section 3.2, our almost affine models imply that yields are approximately linear in the state variables, i.e.,

$$Y_t \approx A_{\mathcal{P}} + B_{\mathcal{P}} \mathcal{P}_t.$$



(a) 5-year yield

(b) 5-year yield

(c) 10-year yield

(d) 10-year yield

Figure 2: One month ahead yield volatility in basis points implied by 3-factor models with volatility

A key question is how accurate this linear approximation is in practice. To assess this, we take each estimated almost affine model and simulate 50,000 paths under the risk-neutral measure at the start of each calendar year in our sample. Using these simulated paths, we compute the n-period zero coupon bond prices based on the expression:

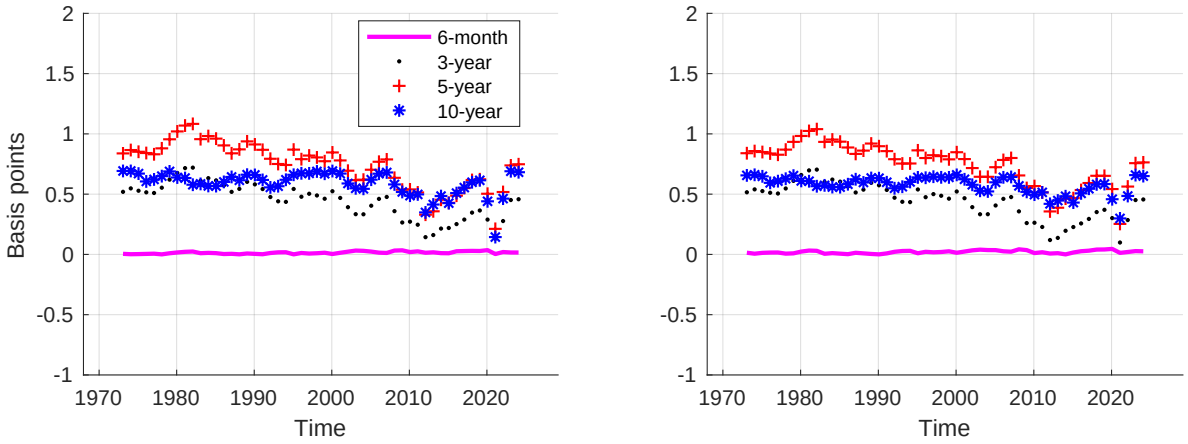
$$E_t^{\mathbb{Q}}[exp(-r_t - r_{t+1} - ... - r_{t+n-1})],$$

and then compute the corresponding n-period zero-coupon yield. Given the large number of simulations, the associated Monte Carlo noise is negligible, allowing us to treat the simulated yields as accurate benchmarks for the true yields implied by the model. We then compare

these benchmark yields with those generated by the linear approximation and interpret the absolute differences as approximation errors.

	$AA_1(3)$				$AA_2(3)$			
	6-month	3-year	5-year	10-year	6-month	3-year	5-year	10-year
mean	0.01	0.44	0.74	0.59	0.02	0.44	0.75	0.58
std	0.01	0.15	0.20	0.10	0.01	0.15	0.18	0.08
min	0.00	0.13	0.21	0.14	0.00	0.10	0.25	0.30
10th percentile	0.00	0.24	0.48	0.46	0.01	0.23	0.50	0.48
median	0.01	0.47	0.78	0.61	0.02	0.47	0.79	0.60
90th percentile	0.03	0.62	0.96	0.69	0.04	0.61	0.93	0.65
max	0.04	0.72	1.08	0.69	0.05	0.70	1.04	0.66

Table 5: 3-factor Almost Affine models: summary statistics of linear approximation absolute errors in basis points.



(a) $AA_1(3)$ (b) $AA_2(3)$

Figure 3: Time series of linear approximation absolute errors in basis points implied by 3-factor Almost Affine models.

Table 5 reports summary statistics of the absolute approximation errors (in basis points) across four maturities – 6-month, 3-year, 5-year, and 10-year – under the $AA_1(3)$ and $AA_2(3)$ models. Across both specifications, the approximation errors are remarkably small, typically well below 1 basis point. For short maturities, such as the 6-month yield, errors are essentially negligible, with mean values around 0.01-0.02 basis points and maximum errors no greater than 0.05 basis points.

As expected, the approximation error increases slightly with maturity, peaking at the 5-year horizon, where the mean error reaches approximately 0.74–0.75 basis points. Even

so, the 90th percentile and maximum errors remain under 1.1 basis points, confirming the economic insignificance of the approximation errors across the entire yield curve.

Interestingly, approximation accuracy is highly comparable between the $AA_1(3)$ and $AA_2(3)$ models. The addition of a second volatility factor in $AA_2(3)$ does not worsen the quality of the approximation. In fact, the 10-year pricing errors in this model are slightly smaller at the upper tail of the error distribution.

These findings are reinforced by Figure 3, which displays the time series of approximation errors across maturities for both models. Errors are stable over time and show no signs of clustering or drift. Some volatility in approximation error is visible for the longer-maturity yields (5-year and 10-year), but even in these cases, the fluctuations in approximation errors are modest. For the 6-month yield, errors hover near zero throughout the entire sample.

Importantly, the approximation remains robust even during historically volatile periods, such as the early 1980s and the 2020 COVID-19 crisis. The small magnitude and consistent behavior of the approximation errors across regimes demonstrate the reliability and numerical stability of the linearized pricing solution used in our almost affine models.

Table 5 and Figure 3 together validate the linear approximation as a highly accurate and tractable solution method for pricing bonds in our framework. The negligible magnitude of the errors ensures that our empirical findings are not driven by approximation artifacts.

5.5 Statistical model selection criteria

	$A_0(3)$	$A_1(3)$	$AA_1(3)$	$A_2(3)$	$AA_2(3)$
llk	21693	162	331	164	381
# of parameters	23	1	9	1	18
AIC	-43339	-322	-644^*	-325	-727^*
AICc	-43338	-322	-642^*	-325	-723^*
BIC	-43237	-317	-604^*	-321	-647^*

Table 6: 3-factor models: log likelihood scores, number of parameters, and fitness scores relative to the constant-volatility model $A_0(3)$.

Table 6 reports the log-likelihood values, Akaike Information Criterion (AIC), corrected AIC (AICc), and Bayesian Information Criterion (BIC) for the pure Gaussian model $A_0(3)$. For the affine and almost affine models, we present the *improvements* in these scores relative to $A_0(3)$, to highlight the gains from incorporating stochastic volatility and structural flexibility.

Across all criteria, introducing time variation in volatility leads to substantial improvements over the constant-volatility benchmark. Even the most parsimonious affine model with one volatility factor ($A_1(3)$) significantly outperforms $A_0(3)$, underscoring the empirical importance of incorporating stochastic volatility.

However, holding the number of volatility factors fixed, the almost affine models consistently achieve the best performance. For both one-factor and two-factor specifications, the

almost affine variants ($AA_1(3)$ and $AA_2(3)$) dominate their affine counterparts in terms of likelihood-based metrics.

While it is true that the almost affine models require a greater number of parameters – particularly when compared to the highly constrained affine models – both AIC and BIC explicitly penalize model complexity. The fact that $AA_1(3)$ and $AA_2(3)$ still emerge with the lowest AIC, AICc, and BIC scores highlights the strength of their empirical fit despite the increase in model flexibility. Among the two almost affine models, $AA_2(3)$ appears to have the edge, producing the best overall fit according to all three criteria.

The results for the four-factor models, reported in Section B.5, are largely consistent. While the almost affine models continue to outperform their affine counterparts, all selection criteria display a strong preference for specifications with only one volatility factor. This suggests that introducing a second volatility factor may yield diminishing returns in terms of statistical fit, even within the more flexible modeling framework.

It is important to emphasize, however, that these model selection metrics are purely statistical. In the next section, we shift focus to the economic implications of the almost affine framework, evaluating how these models perform relative to their affine counterparts in capturing economically meaningful yield dynamics.

6 Predictive regressions and term premium decomposition

6.1 LPY(i) regressions: conditional mean dynamics

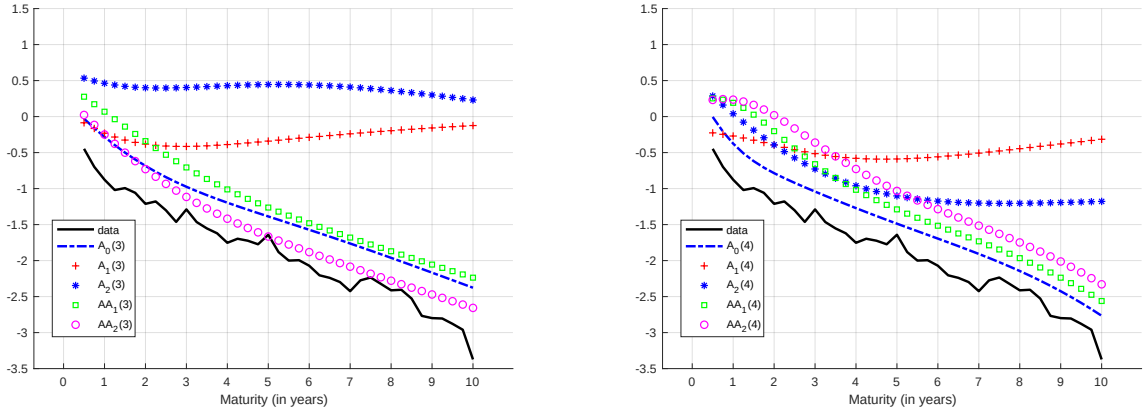


Figure 4: LPY(i) regression coefficients

Figure 4 plots the coefficients from the LPY(i) regressions in Equation (1), comparing model-implied values from various estimated affine and almost affine models to the empirical coefficients (solid black line). These regressions serve as a valuable diagnostic for assessing how well the models capture the conditional mean dynamics of bond yields – specifically, the

predictive relationship between yield slopes and future changes in yields across the maturity spectrum.

Consistent with previous findings (e.g., Dai and Singleton (2002)), the pure affine models struggle to replicate the empirical LPY(i) pattern, particularly at longer maturities. The model-implied coefficients from the $A_1(3)$ and $A_2(3)$ specifications tend to be overly flat and fail to match the steadily declining profile observed in the data. This shortcoming reflects the well-documented rigidity of the affine structure under no-arbitrage constraints – a limitation examined in detail by both Dai and Singleton (2002) and Joslin and Le (2021).

By contrast, the almost affine models perform significantly better. Both $AA_1(3)$ and $AA_2(3)$ closely track the empirical LPY(i) coefficients across the maturity range and perform on par with the pure Gaussian benchmark $A_0(3)$, which itself provides a strong fit to the empirical pattern. These results highlight the value of relaxing the common eigenvector constraint imposed by traditional affine models, thereby enabling the almost affine specifications to better capture realistic predictive dynamics in the yield curve.

The four-factor affine models, $A_1(4)$ and $A_2(4)$, show clear improvements over their three-factor counterparts. For maturities up to five years, their implied LPY(i) coefficients are broadly comparable to those produced by the almost affine and Gaussian models, suggesting that the inclusion of a fourth yield factor increases flexibility in capturing short- and medium-term yield predictability. However, beyond the five-year horizon, their performance again deteriorates, deviating from the empirical pattern. This highlights the importance of evaluating LPY(i) regressions across the full maturity spectrum; diagnostics focused only on short maturities – as is occasionally seen in prior studies – may yield an incomplete or even misleading picture of a model's adequacy.

An additional observation concerns the role of volatility factor dimensionality. In the three-factor affine models, introducing a second volatility factor appears to degrade the model's ability to match the empirical LPY(i) pattern. In contrast, within the four-factor affine framework, the addition of a second volatility factor yields modest improvements, particularly at longer maturities. This contrast suggests that the common left eigenvector constraint imposed by pure affine models can lead to a nuanced and potentially complex interaction between model dimensionality and volatility structure in shaping the dynamics of yields.

In contrast, the four-factor almost affine models, $AA_1(4)$ and $AA_2(4)$, align remarkably well with the empirical LPY(i) coefficients across virtually all maturities. Their consistent performance further demonstrates the benefits of loosening the structural constraints inherent in traditional affine specifications.

The superior performance of the almost affine models can be attributed to two main factors. First, they relax the admissibility-induced constraint that forces the risk-neutral and physical feedback matrices to share a common left eigenvector basis. By design, this structural relaxation allows the almost affine models far greater flexibility in specifying the time-series dynamics of bond yields. See Joslin and Le (2021) for a more detailed discussion of this point.

Second, and slightly more subtly, the almost affine models afford greater freedom in

selecting the volatility factors. Specifically, the volatility loading vector β is unrestricted and can be chosen freely from the full N-dimensional space. This flexibility enables the models to better capture the volatility dynamics of yields. Within the context of stochastic volatility term structure models, improved identification of yield volatility enhances the efficiency of the overall estimation process. In turn, this gives rise to sharper identification of model parameters, including those governing the conditional mean dynamics. This mechanism is analogous to the efficiency gains achieved in generalized least squares relative to ordinary least squares, where optimal weighting of more informative signals leads to more precise parameter estimates.

6.2 LPY(ii) regressions: short rate forecasts efficiency

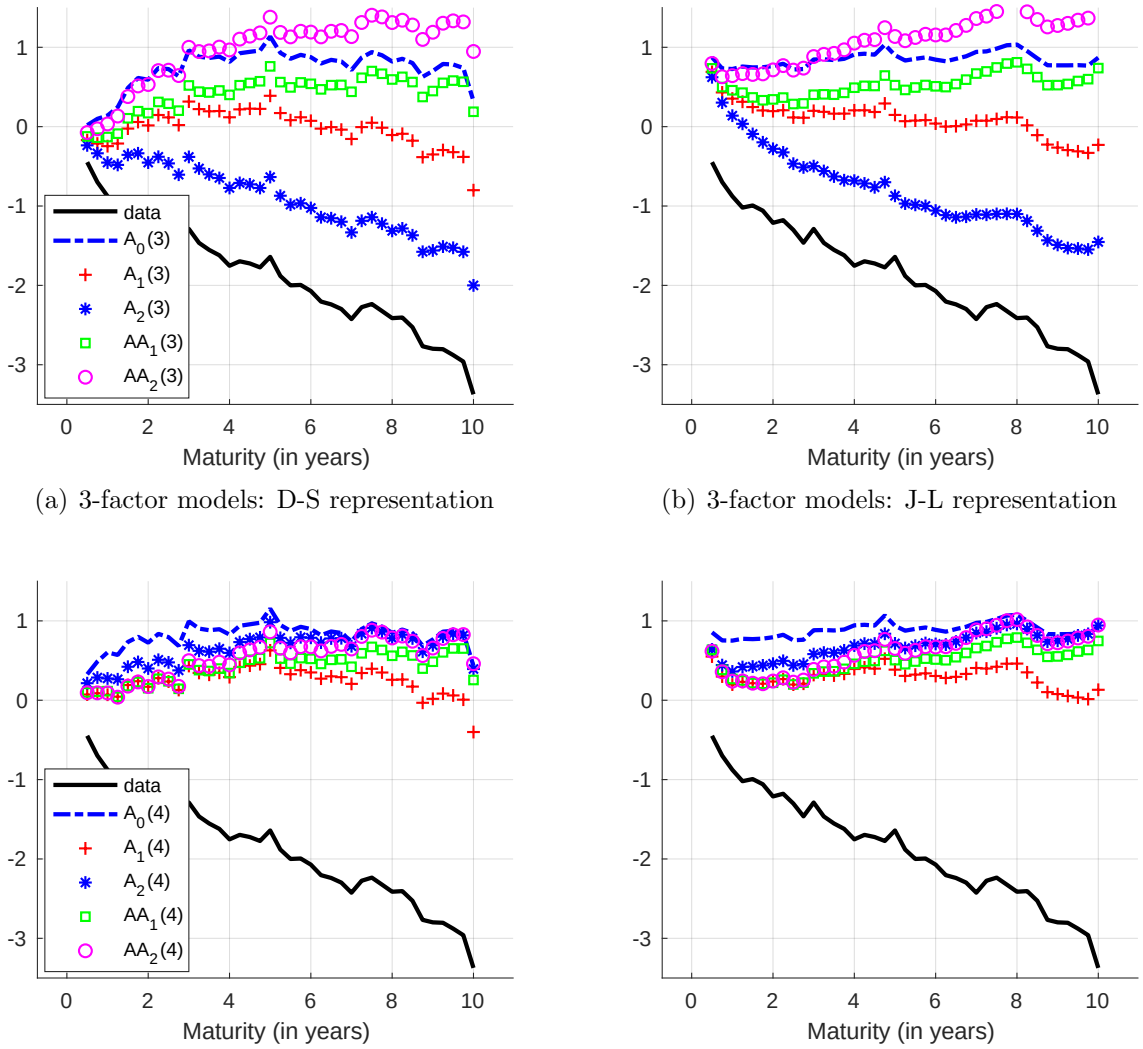
Panel A: 3-factor models			
	PC1	PC2	PC3
$A_0(3)$	0.991	0.878	0.129
$A_1(3)$	0.412	0.004***	0.032**
$A_2(3)$	0.202	0.047**	0.128
$AA_1(3)$	0.814	0.299	0.313
$AA_2(3)$	0.736	0.873	0.340
Panel B: 4-factor models			
	PC1	PC2	PC3	PC4
$A_0(4)$	0.998	0.952	0.246	0.001***
$A_1(4)$	0.784	0.411	0.322	0.000***
$A_2(4)$	0.053*	0.680	0.026**	0.008***
$AA_1(4)$	0.933	0.513	0.174	0.524
$AA_2(4)$	0.869	0.533	0.021**	0.015**

Table 7: Efficiency test of short rate forecasts

Figure 5 presents the coefficient estimates from the LPY(ii) regressions, based on Equation (3), alongside those from our proposed alternative formulation in Equation (5). To distinguish between the two, we refer to the original specification from Dai and Singleton (2002) as the D-S representation, shown in subplots (a) and (c), and to our alternative formulation as the J-L representation, shown in subplots (b) and (d).

Recall that the LPY(ii) regression is based on yields net of term premiums; thus, if a model accurately captures term premium dynamics, the expectations hypothesis should hold, and the regression coefficients should equal one across all maturities. Comparing the D-S and J-L representations reveals that they are nearly identical, with any differences attributable to yield pricing errors, which are typically small.[9] This near-equivalence allows us to rely on the

[9]In fact, when pricing errors are zero, the two specifications produce identical coefficients.



(a) 3-factor models: D-S representation

(b) 3-factor models: J-L representation

(c) 4-factor models: D-S representation

(d) 4-factor models: J-L representation

Figure 5: LPY(ii) regression coefficients

J-L representation and interpret LPY(ii) regressions as reflecting the efficiency of a model's short rate forecasts with respect to slope-based predictive signals. It is important to emphasize that this interpretation is conceptually orthogonal to a model's LPY(i) performance.

Turning to the results, the three-factor almost affine models ($AA_1(3)$, $AA_2(3)$) again outperform their affine counterparts. Their LPY(ii) coefficients lie close to the unit line, performing on par with the pure Gaussian model $A_0(3)$. In contrast, the $A_2(3)$ model—with two volatility factors—deviates substantially, indicating weaker short rate forecast efficiency.

The four-factor models reveal a particularly compelling picture. All models – except possibly $A_1(4)$ – deliver LPY(ii) coefficients that closely track the unit line. Even the $A_1(4)$ model performs reasonably well, certainly better than its three-factor counterpart, $A_1(3)$.

Dai and Singleton (2002) interpret strong LPY(ii) performance by four-factor affine models as evidence that they better capture risk-neutral yield dynamics, whereas three-factor models do not. However, this interpretation seems at odds with the fact that yield pricing errors are uniformly small – less than 10 basis points –across all models.

The J-L representation provides a more plausible interpretation. Specifically, the $A_1(3)$ and $A_2(3)$ models appear to generate short rate forecasts that underutilize slope information. In contrast, adding a fourth yield factor appears to equip the affine models with additional flexibility, enabling them to incorporate slope-based predictive signals more effectively.

This raises an interesting question: how can the four-factor affine models, which fail to capture conditional mean dynamics in LPY(i) regressions (as is seen in Figure 4), nonetheless produce efficient short rate forecasts across a wide range of horizons? The answer lies in the nature of the LPY(ii) regression, which focuses exclusively on information embedded in yield slopes. This naturally leads to a broader question: what about information in other components of the yield curve, such as level, curvature, or higher-order factors?

To address this, we regress forecast revisions of the short rate onto each principal component of the yield curve as in Equation (7), and report the p-values of joint significance tests over 40 forecast horizons ranging from 3 months, 6 months, ... to 10 years. If a model fully incorporates the information from a given factor, the corresponding p-value should be large (insignificant).

Table 7 reports these p-values. Panel A shows that all three-factor models efficiently incorporate information from the level factor. However, the affine models $A_1(3)$ and $A_2(3)$ exhibit statistically significant inefficiencies with respect to slope, and in the case of $A_1(3)$, also curvature. In contrast, the almost affine models $AA_1(3)$ and $AA_2(3)$ appear to incorporate all three factors efficiently in forming their short rate forecasts. The evidence here corroborates what we have seen in the top row of Figure 5.

Panel B reveals that all four-factor models efficiently process slope information, consistent with the bottom row of Figure 5. However, all three affine models – including the Gaussian model $A_0(4)$ – struggle to fully utilize higher-order factors such as PC4. The $A_2(4)$ model in particular performs poorly, with significant inefficiencies in level, curvature, and PC4. Interestingly, even the almost affine model $AA_2(4)$, with two volatility factors, fails to synthesize information from curvature and PC4 fully. By contrast, the $AA_1(4)$ model emerges as the most efficient, being the only four-factor model that incorporates all four principal components without significant forecast inefficiencies. This result aligns with our model selection findings in Section B.5, where all statistical criteria favor the $AA_1(4)$ model as the best-performing four-factor specification.

Overall, this subsection has examined an alternative LPY(ii) representation that offers a sharper interpretation: the regression serves as a test of short rate forecast efficiency. However, we also caution that such regressions are much more meaningful when they account for the full informational content of the yield curve.

6.3 Term premium analysis

An n-period bond yield is often decomposed into two components: an expectations hypothesis (EH) term, representing the expected average of future short rates over the bond's life, and a term premium component:

$$y_{n,t} = E_t^{\mathbb{P}} \left[\frac{1}{n} \sum_{i=0}^{n-1} r_{t+i} \right] + TP_{n,t}. \tag{28}$$

Up to Jensen effects, the term premium can be written as the difference between the risk-neutral and physical expectations of the average future short rate:

$$TP_{n,t} \approx (E_t^{\mathbb{Q}} - E_t^{\mathbb{P}}) \left[\frac{1}{n} \sum_{i=0}^{n-1} r_{t+i} \right]. \tag{29}$$

This expression makes it clear that the term premium reflects the compensation investors require for bearing interest rate risks. Under risk neutrality, where $\mathbb{P} \equiv \mathbb{Q}$, the term premium would be negligible.

Understanding and accurately estimating the dynamics of the term premium is essential for both policymakers and market participants. For investors, this decomposition informs fixed-income pricing, risk management, and the interpretation of monetary policy signals. For central banks, particularly in implementing monetary policies targeting inflation and output stabilization, misattributing shifts in long-term yields to changing expectations rather than term premiums may result in misguided policy responses. Importantly, term premiums are sensitive to broader financial conditions, including risk sentiment, market volatility, and the global demand for safe assets, making them a key barometer of market stress and uncertainty. Notably, both the Federal Reserve Bank of New York and the Federal Reserve Bank of San Francisco regularly publish model-based estimates of term premiums on their respective websites.

In practice, term premiums are often estimated using variants of the pure Gaussian affine term structure model, $A_0(N)$. The main advantage of this model is its demonstrated ability to match the conditional mean dynamics of bond yields. However, a critical limitation is that volatility is assumed constant over time. As a result, any observed variation in term premiums – say, a 20 basis point increase in the 10-year term premium – is attributed entirely to changes in investors' risk preferences rather than changes in the quantity of risk. This assumed constancy in yield volatility is clearly counter-factual.

While affine models with stochastic volatility, such as $A_1(N)$ or $A_2(N)$, allow for time-varying second moments, they tend to fit the conditional means of yields poorly. Consequently, they likely produce unreliable decompositions of yields into EH and TP components, limiting their usefulness in policy contexts.

As demonstrated in the previous subsections, our proposed almost affine models overcome these limitations. They match the conditional mean dynamics of yields as effectively as the pure Gaussian models, while also offering substantially greater flexibility in capturing

time-varying volatility of yields. In this section, we use our estimated almost affine models to generate term premium estimates consistent with Equation (28).

We then ask: To what extent can variations in these estimated term premiums be attributed to changes in the quantity of risk? To address this, we examine the model-implied conditional volatility of the average short rate over the bond's horizon:

$$\sigma_t \left[\frac{1}{n} \sum_{i=0}^{n-1} r_{t+i} \right]. \tag{30}$$

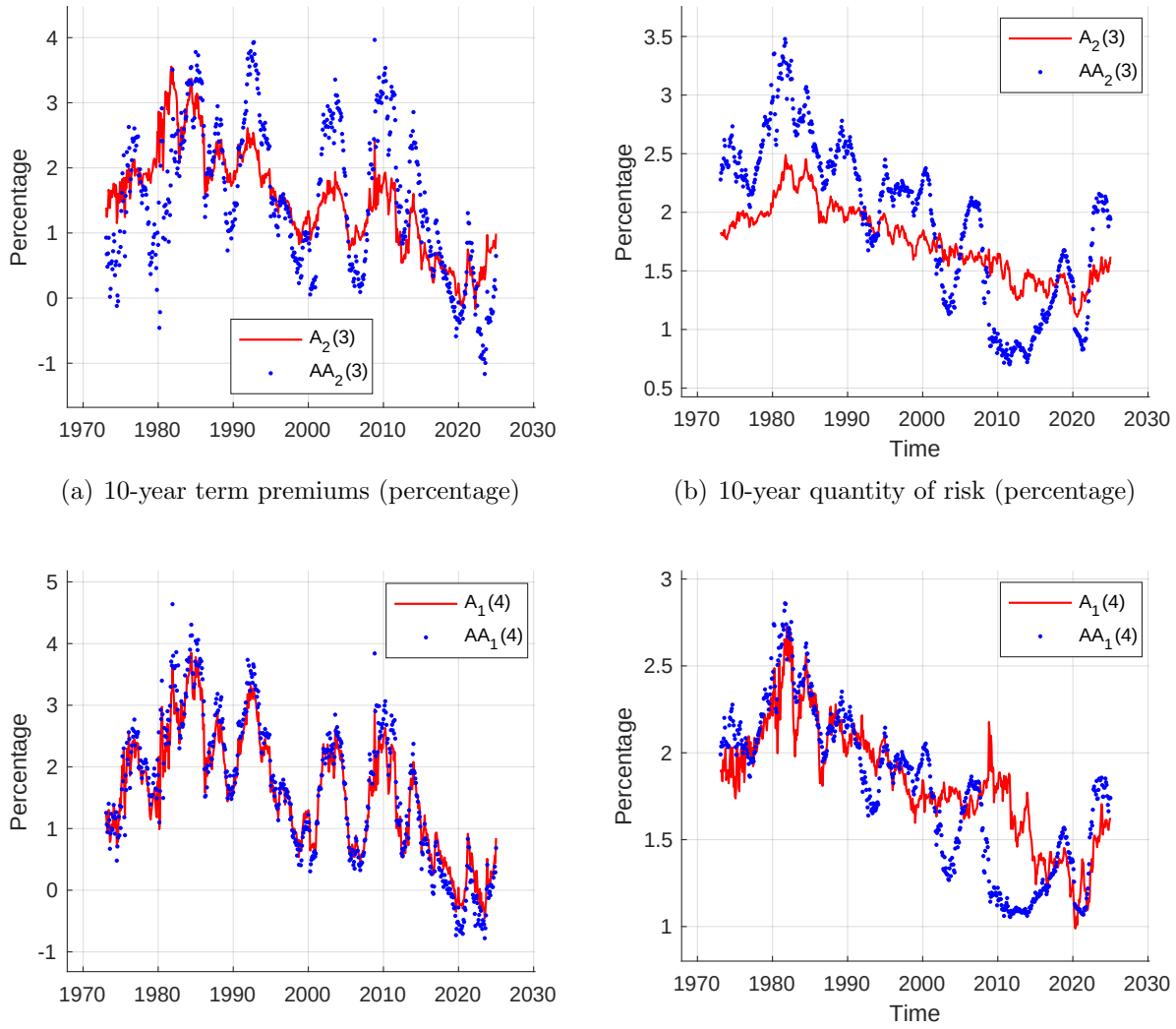
While alternative measures of risk quantity exist, this one aligns naturally with the term premium decomposition in Equation (29) and provides a direct link between risk compensation and model-implied uncertainty.

Figure 6 plots the 10-year term premium estimates and the corresponding measures of risk quantity produced by the two preferred almost affine specifications – $AA_2(3)$ and $AA_1(4)$ – alongside those from their respective pure affine counterparts, $A_2(3)$ and $A_1(4)$. The panels in the top row focus on the 3-factor specifications. As shown in panel (a), the almost affine model $AA_2(3)$ generates 10-year term premium estimates that differ markedly from those of the pure affine model $A_2(3)$, especially during periods of heightened market stress. These differences are most evident during the early 1980s, the early 2000s, and the COVID-19 period, where the $AA_2(3)$ model tends to imply both higher and more volatile term premiums. Turning to panel (b), the divergence is even more striking in the estimated quantity of risk: the $AA_2(3)$ model exhibits a more pronounced response to changing market conditions, especially during the 1970s inflation shock, the Global Financial Crisis, and COVID-19, reflecting its ability to capture time-varying volatility.

In contrast, the 4-factor models (panels c and d) seem to show more alignment in term premium estimates. As seen in panel (c), the term premiums implied by $AA_1(4)$ and $A_1(4)$ are broadly similar in level and dynamics over most of the sample period. However, notable deviations occur during volatile episodes such as the Paul Volcker disinflation period, the Great Recession, and the COVID-19 shock. During these times, differences of up to 100 basis points (Volcker and Great Recession) and around 50 basis points (COVID) emerge. Panel (d) reveals that the quantity of risk implied by the pure affine model $A_1(4)$ is consistently higher and less responsive to economic conditions compared to $AA_1(4)$. For example, during the Great Recession, the $A_1(4)$ model's risk quantity peaks at nearly twice the level suggested by $AA_1(4)$, suggesting that the pure affine model may overstate the level of perceived risk due to its constraint in selecting volatility instruments.

Next, we regress model-implied term premiums on their corresponding estimates of the quantity of risk, measured by σ_t, and report the adjusted R^2 statistics in Table 8. We interpret these R^2 values as the share of term premium variation that can be attributed to the quantity of risk channel. To account for potential nonlinearity, we also consider regressions on σ_t^2 as well as regressions on both σ_t and σ_t^2.

Table 8 reveals a striking contrast between the almost affine and pure affine models. Across all dimensions – whether one considers 3- or 4-factor models, or focuses on the 5-year or 10-year horizons – the almost affine specifications attribute only a small portion of the



(a) 10-year term premiums (percentage)

(b) 10-year quantity of risk (percentage)

(c) 10-year term premiums (percentage)

(d) 10-year quantity of risk (percentage)

Figure 6: 10-year term premiums and the corresponding quantity of risks implied by preferred almost affine models.

variation in term premiums to changes in the quantity of risk.

The most dramatic example is the $AA_2(3)$ model, which produces adjusted R^2 values that are essentially zero (or even slightly negative), suggesting that the quantity of risk plays virtually no explanatory role in this specification. In sharp contrast, the corresponding pure affine model, $A_2(3)$, assigns over 80% of the term premium variation to the volatility channel for both maturities. This difference underscores a key implication of our modeling approach: by relaxing the rigid structure of affine models and allowing for more flexible volatility dynamics, term premiums are less prone to be mechanically linked to fluctuations in the quantity of risk.

	5-year yield			10-year yield		
	σ_t	σ_t^2	$\sigma_t + \sigma_t^2$	σ_t	σ_t^2	$\sigma_t + \sigma_t^2$
Panel A: 3-factor models:						
$A_0(3)$	-0.16	-0.16	-0.32	-0.16	-0.16	-0.32
$A_1(3)$	58.54	57.08	57.02	67.29	65.71	65.66
$A_2(3)$	83.20	83.45	83.43	83.18	83.42	83.39
$AA_1(3)$	2.52	3.24	3.08	1.14	2.13	1.97
$AA_2(3)$	0.08	0.32	0.16	-0.14	-0.11	-0.27
Panel B: 4-factor models:						
$A_0(4)$	-0.16	-0.16	-0.32	-0.16	-0.16	-0.32
$A_1(4)$	51.31	49.78	49.70	64.99	63.26	63.21
$A_2(4)$	59.21	57.59	57.52	71.65	70.20	70.15
$AA_1(4)$	17.07	18.47	18.34	17.07	19.53	19.40
$AA_2(4)$	8.04	8.62	8.48	7.79	9.42	9.28

Table 8: In-sample explantory power of quantity of risk (adjusted R-squared) for term premiums implied by almost affine models.

Turning to the 4-factor case, we observe similar, though somewhat less extreme, patterns. For example, at the 10-year horizon, the pure affine model $A_1(4)$ attributes approximately 63–64% of term premium variation to the volatility channel, while the almost affine model $AA_1(4)$ yields a substantially lower R^2 of about 17–19%. These findings suggest that while introducing an additional factor in the almost affine setting improves explanatory power, a large share of the time variation in term premiums remains orthogonal to the quantity of risk measures used here.

Table 9 reports the population analogs of the adjusted R^2 statistics presented earlier in Table 8. Rather than relying on in-sample estimates, we use each fitted model to simulate a long sample – comprising one million months of yields – and then estimate the same regressions of term premiums on risk quantity measures using the simulated data. This procedure provides a slightly cleaner view of the models' implications, uncontaminated by in-sample sampling noise.

The results in Table 9 broadly reinforce the findings from the in-sample analysis. If anything, the differences between the models become even more pronounced. For instance, the proportion of term premium variation explained by the quantity of risk in the pure affine model $A_1(4)$ rises substantially in population, reaching approximately 90% for the 10-year maturity. By contrast, the corresponding statistic for the almost affine model $AA_1(4)$ falls to just 12%.

Overall, this analysis highlights that pure affine models may overstate the role of volatility in driving term premiums due to their restrictive structure. The almost affine models, by decoupling volatility dynamics from the cross-section of yields, offer a more nuanced and

	5-year yield			10-year yield		
	σ_t	σ_t^2	$\sigma_t + \sigma_t^2$	σ_t	σ_t^2	$\sigma_t + \sigma_t^2$
Panel A: 3-factor models:						
$A_0(3)$	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
$A_1(3)$	79.57	84.37	84.37	82.27	86.49	86.49
$A_2(3)$	78.36	79.89	79.89	78.23	79.35	79.34
$AA_1(3)$	10.84	9.40	9.40	6.77	5.77	5.77
$AA_2(3)$	4.03	3.99	3.99	1.43	1.50	1.50
Panel B: 4-factor models:						
$A_0(4)$	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
$A_1(4)$	81.85	87.42	87.42	87.27	92.14	92.14
$A_2(4)$	53.98	54.39	54.39	73.62	73.92	73.92
$AA_1(4)$	13.95	13.31	13.31	12.50	12.12	12.12
$AA_2(4)$	13.79	16.98	16.98	14.09	17.01	17.01

Table 9: In population explantory power of quantity of risk (adjusted R-squared) for term premiums implied by 4-factor models.

arguably more realistic view of the sources of term premium variation.

A natural question that arises is: what is the economic intuition behind the muted relationship between the term premium and the quantity of risk? After all, the basic "high risk, high return" principle would seem to support the high R^2 values implied by standard affine models. However, it's important to recognize that Treasury bonds differ from typical risky assets in at least one key respect: in bad times, they are widely regarded as safe-haven assets.

In adverse economic states, Treasury bonds provide insurance-like payoffs, delivering values precisely when other assets perform poorly. This insurance characteristic could imply negative risk premiums – higher volatility can be associated with more negative compensation. In contrast, during good times, bonds must compete with other risky investments and tend to offer positive risk premiums. That is, in good states of the world, the usual positive risk-return relationship reemerges.

As a result, the relationship between bond risk premiums and volatility, when estimated over long historical samples, reflects a mix of these opposing dynamics: negative in bad times, positive in good times. Depending on the relative frequency and magnitude of these regimes, the average relationship may appear quite weak – or even statistically insignificant. This averaging effect helps explain why a strong link between risk and return, as suggested by standard affine models, may not be observed in practice.

7 Concluding remarks

This paper has introduced and empirically evaluated a new class of almost affine term structure models that relax the restrictive volatility assumptions inherent in traditional affine frameworks. By decoupling the volatility dynamics from the cross-section of yields, our models achieve a more realistic representation of bond yield behavior, particularly during periods of heightened market volatility.

Our findings demonstrate that the almost affine models retain the desirable features of pure Gaussian models in capturing the conditional mean dynamics of yields, while significantly improving the ability to accommodate time-varying second moments. Importantly, we show that the resulting term premium estimates are less mechanically tied to fluctuations in volatility, providing a more nuanced understanding of the underlying drivers of risk compensation in bond markets.

The decomposition of term premiums reveals substantial differences between the almost affine and traditional affine models, especially during turbulent episodes such as the Great Recession and the COVID-19 crisis. Moreover, the quantity of risk explains a markedly smaller share of term premium variation in our proposed framework, highlighting the importance of separating volatility dynamics from the pricing kernel.

Overall, the almost affine approach offers a flexible and empirically sound alternative for modeling the term structure of interest rates. It opens the door to richer economic interpretations and more robust empirical analysis, particularly in environments characterized by volatile and nonlinear dynamics. Future work may explore extensions to macro-finance settings, multi-country models, or applications in monetary policy and risk management.

References

Campbell, J., and R. Shiller, 1991, "Yield Spreads and Interest Rate Movements: A Bird's Eye View," Review of Economic Studies, 58, 495–514.

Cheridito, R., D. Filipovic, and R. Kimmel, 2007, "Market Price of Risk Specifications for Affine Models: Theory and Evidence," Journal of Financial Economics, 83, 123 – 170.

Collin-Dufresne, P., R. Goldstein, and C. Jones, 2008, "Identification of Maximal Affine Term Structure Models," Journal of Finance, LXIII, 743–795.

Collin-Dufresne, P., and R. S. Goldstein, 2002, "Do Bonds Span the Fixed Income Markets? Theory and Evidence for 'Unspanned' Stochastic Volatility," Journal of Finance, 57, 1685–1730.

Dai, Q., and K. Singleton, 2000, "Specification Analysis of Affine Term Structure Models," Journal of Finance, 55, 1943–1978.

Dai, Q., and K. Singleton, 2002, "Expectations Puzzles, Time-Varying Risk Premia, and Affine Models of the Term Structure," Journal of Financial Economics, 63, 415–441.

Doshi, H., K. Jacobs, and R. Liu, 2024, "Modeling volatility in dynamic term structure models," Journal of Financial Economics, 161.

Duffee, G., 2002, "Term Premia and Interest Rates Forecasts in Affine Models," Journal of Finance, 57, 405–443.

Duffee, G., 2011, "Forecasting with the Term Structure: the Role of No-Arbitrage," Discussion paper, Johns Hopkins University.

Duffie, D., and R. Kan, 1996, "A Yield-Factor Model of Interest Rates," Mathematical Finance, 6, 379–406.

Duffie, D., J. Pan, and K. Singleton, 2000, "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," Econometrica, 68, 1343–1376.

Joslin, S., and A. Le, 2021, "Interest Rate Volatility and No-Arbitrage Affine Term Structure Models," Management Science, 67, 7391–7416.

Joslin, S., A. Le, and K. Singleton, 2012, "Why Gaussian Macro-Finance Term Structure Models Are (Nearly) Unconstrained Factor-VARs," Journal of Financial Economics, forthcoming.

Joslin, S., A. Le, and K. Singleton, 2013, "Gaussian Macro-Finance Term Structure Models with Lags," Journal of Financial Econometrics, 11, 581–609.

Joslin, S., K. Singleton, and H. Zhu, 2011, "A New Perspective on Gaussian DTSMs," Review of Financial Studies.

Le, A., and K. Singleton, 2024, "The Structure of Risks in Equilibrium Affine Models of Bond Yields," Discussion paper, Pennsylvania State University.

A Further estimation details

A.1 Parameterization

In terms of parameterization, the insight rorm JSZ is that we use parameters that are tied to observable quantities. That is, $\Sigma_{i,\mathcal{P}}$ is much more easily identified in estimation than $\Sigma_{i,X}$. Nevertheless, we cannot use $\Sigma_{i,\mathcal{P}}$ directly as primitive parameters because we cannot go from $\Sigma_{i,\mathcal{P}}$ back to $\Sigma_{i,X}$ in a straightforward manner. Unlike the pure Gaussian case, the loadings B_X are dependent on both $K_{1X}^{\mathbb{Q}}$ as well as the parameters that govern the time-variation of the conditional variances, including $\Sigma_{i,X}$ and β_X. Therefore, there is a nonlinear mapping between $\Sigma_{i,\mathcal{P}}$ and $\Sigma_{i,X}$ that is hard to invert.

To circumvent the above issue, we consider a corresponding set of convexity-adjusted yields:

$$Y_t^c = A_X^c + B_X^c X_t \tag{31}$$

where the loadings are obtained from the same risk-neutral dynamics but ignoring convexity effects. This is essentially the same pricing equation obtained by setting $\Sigma_{i,X}$ to zeros. In this case, B_X^c is only dependent on $K_{1X}^{\mathbb{Q}}$.

We let:

$$\Sigma_{i,\mathcal{P}}^c = (WB_X^c)\Sigma_{i,X}(WB_X^c)', \tag{32}$$

$$\beta_{\mathcal{P}}^c = \beta_X(WB_X^c)^{-1} \tag{33}$$

being the convexity-adjusted counterparts to $\Sigma_{i,\mathcal{P}}$ and $\beta_{\mathcal{P}}$ and use these as primitive parameters. These parameters correspond to observable portfolios of yields (that are convexity-adjusted) and thus should be easily identified.

Finally, to fix the scaling of Z, we normalize $\beta_{\mathcal{P}}^c$ such that its first column as an absolute value of ones. If $\mathcal{P}$ is the vector of yield PCs, then the loading of each volatility instrument on the level factor is either 1 or -1.

A.2 Recap

The parameter set includes: $r_\infty^{\mathbb{Q}}, K_{1X}^{\mathbb{Q}}, \alpha_{\mathcal{P}}, \beta_{\mathcal{P}}^c, \Sigma_{i,\mathcal{P}}^c, K_{0,\mathcal{P}}^{\mathbb{P}}, K_{1,\mathcal{P}}^{\mathbb{P}}$. From these parameters, the model can be constructed in the following steps:

- From $K_{1X}^{\mathbb{Q}}$, one can compute the convexity-adjusted yields loadings B_X^c.
- With B_X^c and $\Sigma_{i,\mathcal{P}}^c$, we can compute $\Sigma_{i,X}$: $\Sigma_{i,X} = (WB_X^c)^{-1}\Sigma_{i,\mathcal{P}}^c(WB_X^c)'^{-1}$
- Similarly, with B_X^c and $\beta_{\mathcal{P}}^c$, we can compute β_X: $\beta_X = \beta_{\mathcal{P}}^c(WB_X^c)$
- With $r_\infty^{\mathbb{Q}}, K_{1X}^{\mathbb{Q}}, \beta_X, \Sigma_{i,X}$, we can compute the yield loadings B_X and the intercept terms A_X as a linear function of α_X. Having $\alpha_{\mathcal{P}}$, we can back out the corresponding value for α_X from the equation: $\alpha_{\mathcal{P}} = \alpha_X - \beta_{\mathcal{P}}WA_X$.
- Equipped with the yield loadings B_X, we can compute $\Sigma_{i,\mathcal{P}} = (WB_X)\Sigma_{i,X}(WB_X)'$ and $\beta_{\mathcal{P}} = \beta_X(WB_X)^{-1}$ and then, using $K_{0,\mathcal{P}}^{\mathbb{P}}, K_{1,\mathcal{P}}^{\mathbb{P}}$, evaluate the time series density.

A.3 Concentrate out $r_\infty^{\mathbb{Q}}$ and σ_e

Assume that we can write $A_{\mathcal{P}}$ as linear in $r_\infty^{\mathbb{Q}}$:

$$A_{\mathcal{P}} = H_0 + H_r r_\infty^{\mathbb{Q}} \tag{34}$$

then the pricing errors will be linear in $r_\infty^{\mathbb{Q}}$

$$p_{e,t+1} = W_e(Y_{t+1} - (H_0 + H_r r_\infty^{\mathbb{Q}}) - B_{\mathcal{P}}\mathcal{P}_{t+1}). \tag{35}$$

Assuming these pricing errors are iid gaussian with one common variance σ_e^2, the optimal estimate for $r_\infty^{\mathbb{Q}}$ can be obtained from:

$$(W_e H_r)' E_T[W_e(Y_{t+1} - (H_0 + H_r r_\infty^{\mathbb{Q}}) - B_{\mathcal{P}}\mathcal{P}_{t+1})] = 0 \tag{36}$$

which implies:

$$r_\infty^{\mathbb{Q}} = \frac{(W_e H_r)' W_e E_T[Y_{t+1} - H_0 - B_{\mathcal{P}}\mathcal{P}_{t+1}]}{(W_e H_r)' W_e H_r}. \tag{37}$$

Given the optimal estimate for $r_\infty^{\mathbb{Q}}$, we can compute p_e and then the optimal estimate for σ_e as:

$$\sigma_e^2 = E_T[p_{e,t+1}(:)^2]. \tag{38}$$

Now, how do we compute H_0 and H_r? Note that A_X is linear in the intercept parameters. In particular, we can write:

$$A_X = A_0 + A_r r_\infty^{\mathbb{Q}} + A_a \alpha_X. \tag{39}$$

Plus,

$$\begin{aligned} \alpha_{\mathcal{P}} =& \alpha_X - \beta_{\mathcal{P}} W A_X, & (40)\\ =& \alpha_X - \beta_{\mathcal{P}} W (A_0 + A_r r_\infty^{\mathbb{Q}} + A_a \alpha_X) & (41) \end{aligned}$$

which implies:

$$\alpha_X = (I - \beta_{\mathcal{P}} W A_a)^{-1}(\alpha_{\mathcal{P}} + \beta_{\mathcal{P}} W (A_0 + A_r r_\infty^{\mathbb{Q}})). \tag{42}$$

Substitute this back into A_X we have:

$$\begin{aligned} A_X =& A_0 + A_r r_\infty^{\mathbb{Q}} + A_a (I - \beta_{\mathcal{P}} W A_a)^{-1}(\alpha_{\mathcal{P}} + \beta_{\mathcal{P}} W (A_0 + A_r r_\infty^{\mathbb{Q}})), & (43)\\ =& \underbrace{A_0 + A_a (I - \beta_{\mathcal{P}} W A_a)^{-1}(\alpha_{\mathcal{P}} + \beta_{\mathcal{P}} W A_0)}_{H_{0,X}} + \underbrace{(A_r + A_a (I - \beta_{\mathcal{P}} W A_a)^{-1} \beta_{\mathcal{P}} W A_r)}_{H_{r,X}} r_\infty^{\mathbb{Q}}. & (44) \end{aligned}$$

Since $A_{\mathcal{P}} = A_X - B_{\mathcal{P}} W A_X$, it is straightforward to see that:

$$\begin{aligned} H_0 &= H_{0,X} - B_{\mathcal{P}} W H_{0,X}, & (45)\\ H_r &= H_{r,X} - B_{\mathcal{P}} W H_{r,X}. & (46) \end{aligned}$$

B Parameter estimates of 4-factor models

B.1 Conditional mean parameters

	Risk-neutral parameters					Time-series parameters				
	$K_0^{\mathbb{Q}}$	$K_1^{\mathbb{Q}}$				$K_0^{\mathbb{P}}$	$K_1^{\mathbb{P}}$			
$A_0(4)$	0.042	0.999	0.082	0.236	-0.444	0.169	0.994	-0.019	0.027	0.405
	-0.033	-0.002	0.972	-0.234	0.411	0.001	0.002	0.961	-0.138	0.182
	-0.022	-0.001	0.005	0.917	0.292	-0.066	-0.000	0.002	0.866	0.008
	0.017	0.000	-0.004	0.026	0.892	-0.024	0.000	-0.003	0.008	0.701
$A_1(4)$	-0.018	1.001	0.090	0.197	-0.457	0.016	0.995	0.005	0.048	-0.547
	0.020	-0.003	0.965	-0.197	0.424	-0.019	0.000	0.963	-0.181	0.053
	0.011	-0.001	0.000	0.940	0.299	-0.004	-0.001	-0.005	0.909	0.231
	-0.009	0.000	0.000	0.009	0.883	-0.005	0.000	-0.003	0.019	0.854
$AA_1(4)$	0.038	1.000	0.082	0.235	-0.443	0.158	0.995	-0.003	0.180	0.065
	-0.036	-0.002	0.972	-0.234	0.407	0.025	0.001	0.954	-0.143	0.234
	-0.023	-0.001	0.006	0.918	0.290	-0.018	-0.001	-0.007	0.903	0.082
	0.018	0.000	-0.004	0.026	0.893	-0.022	-0.000	-0.001	-0.008	0.810
$A_2(4)$	-0.016	1.001	0.094	0.214	-0.426	0.126	0.993	-0.012	0.077	-0.292
	0.019	-0.003	0.961	-0.214	0.392	0.007	0.000	0.960	-0.161	0.181
	0.010	-0.002	-0.002	0.930	0.278	-0.003	-0.000	-0.005	0.928	0.186
	-0.009	0.001	0.002	0.016	0.898	-0.019	0.001	0.000	0.021	0.858
$AA_2(4)$	0.037	1.000	0.080	0.236	-0.450	0.126	0.996	-0.013	0.191	-0.370
	-0.039	-0.002	0.973	-0.237	0.410	0.054	0.001	0.954	-0.082	0.176
	-0.023	-0.001	0.006	0.916	0.294	-0.022	-0.001	-0.007	0.908	0.026
	0.018	0.000	-0.004	0.028	0.891	-0.019	-0.000	0.000	0.000	0.812

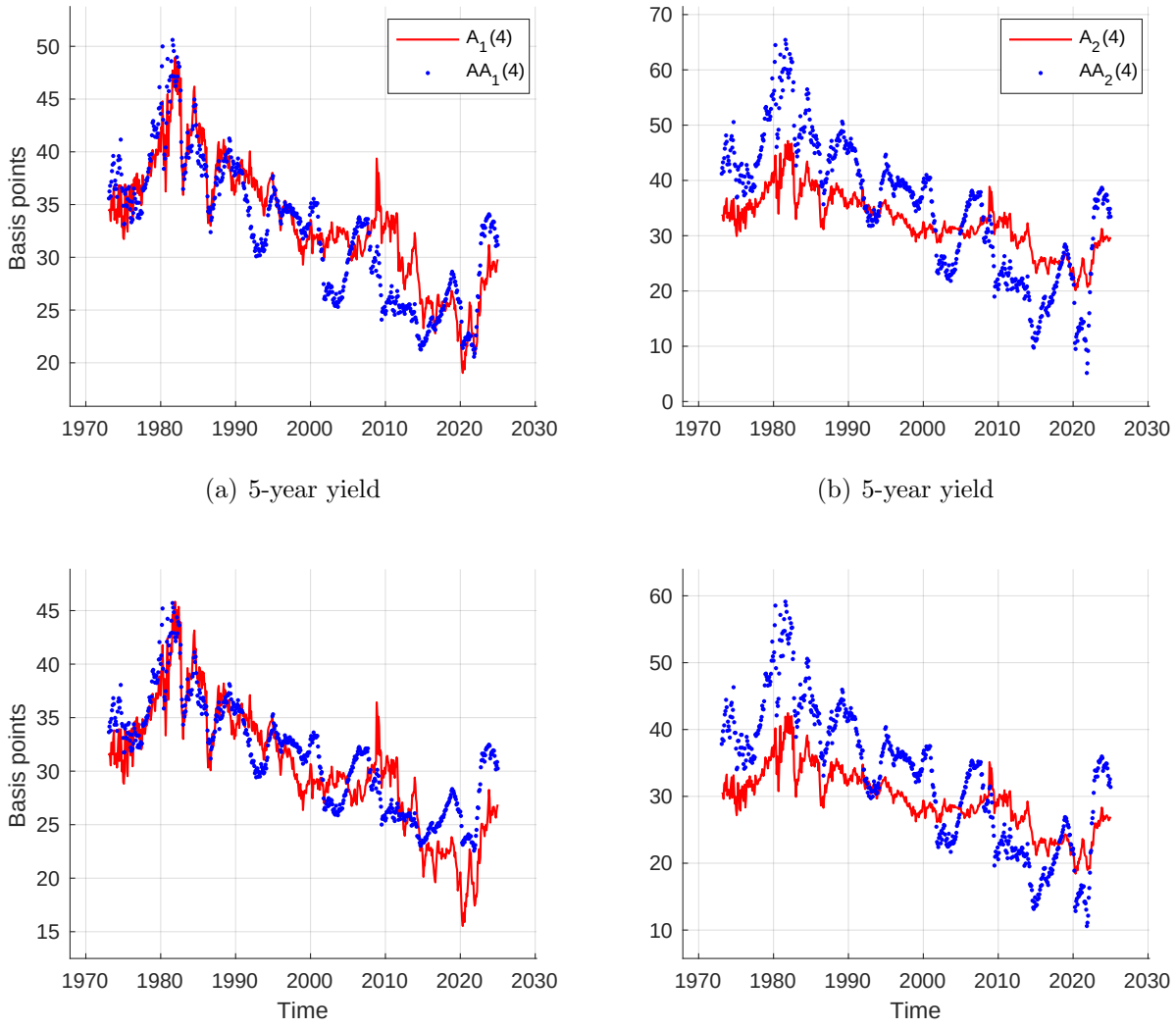
Table 10: 4-factor models: conditional mean parameter estimates.

B.2 Conditional variance parameters

Panel A: 4-factor models with 1 volatility factor								
	$A_1(4)$				$AA_1(4)$			
α	-1.60				-0.55			
β	1.06	2.83	-9.12	-17.75	1.02	-0.91	-2.94	-9.98
$chol(\Sigma_0) \times 100$	29.75				38.18			
	11.42	0.17			19.84	6.70		
	7.24	0.59	0.02		2.35	-6.30	0.00	
	-0.14	0.20	-0.01	0.00	0.48	0.20	0.00	0.00
$chol(\Sigma_1) \times 100$	16.54				19.70			
	-1.57	5.71			-2.79	4.92		
	0.18	0.02	2.78		0.63	0.91	3.54	
	-0.16	0.11	-0.17	1.58	-0.33	-0.25	0.15	2.22
Panel B: 4-factor models with 2 volatility factors								
	$A_2(4)$				$AA_2(4)$			
α	-2.05				-0.69			
	20.74				11.91			
β	1.05	2.15	-10.29	-25.96	1.05	-0.86	-3.10	-10.38
	1.04	2.18	-27.25	30.09	1.08	-2.60	-15.03	-1.00
$chol(\Sigma_0) \times 100$	29.17				0.06			
	14.56	1.00			6.06	0.30		
	3.52	-1.63	1.57		-5.27	-0.03	0.00	
	0.96	0.99	-0.20	0.07	0.03	0.05	0.00	0.00
$chol(\Sigma_1) \times 100$	15.93				23.87			
	-2.54	4.34			0.55	5.20		
	0.19	0.14	1.47		1.37	2.57	1.66	
	-0.21	-0.19	1.05	0.66	-0.37	-1.04	1.58	0.01
$chol(\Sigma_2) \times 100$	2.76				0.02			
	1.37	0.87			0.51	4.31		
	0.89	-0.92	1.25		-1.46	-0.42	0.00	
	-0.27	0.63	-0.17	0.01	-0.19	0.66	0.00	0.00

Table 11: 4-factor models: conditional variance parameter estimates.

B.3 Model-implied one-month ahead yields volatility



(a) 5-year yield

(b) 5-year yield

(c) 10-year yield

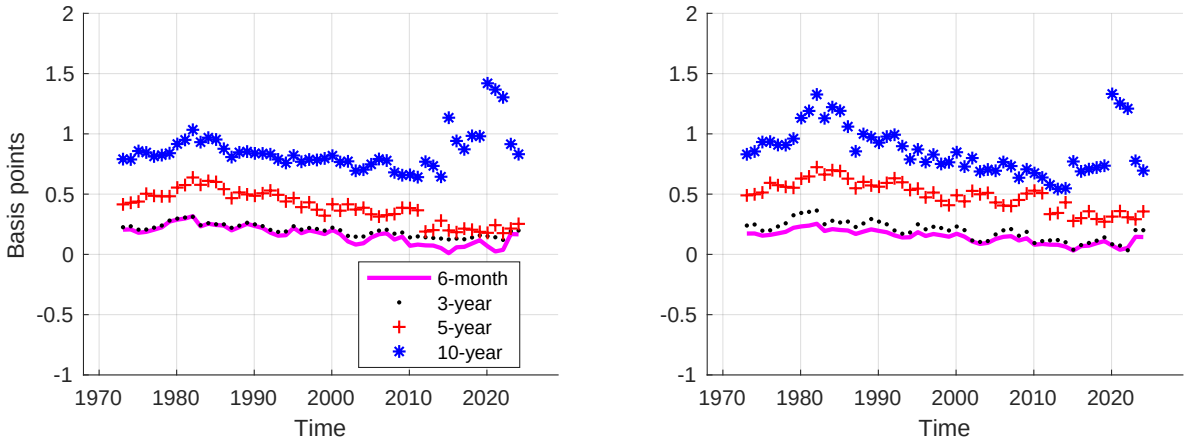
(d) 10-year yield

Figure 7: One month ahead yield volatility in basis points implied by 4-factor models with volatility

B.4 Approximation quality

	$AA_1(4)$				$AA_2(4)$			
	6-month	3-year	5-year	10-year	6-month	3-year	5-year	10-year
mean	0.16	0.20	0.39	0.86	0.14	0.19	0.49	0.87
std	0.08	0.05	0.13	0.16	0.05	0.08	0.12	0.20
min	0.01	0.12	0.18	0.64	0.03	0.03	0.27	0.54
10th percentile	0.06	0.13	0.20	0.69	0.07	0.09	0.31	0.66
median	0.17	0.20	0.40	0.82	0.15	0.20	0.51	0.83
90th percentile	0.25	0.26	0.56	1.00	0.21	0.28	0.64	1.20
max	0.32	0.31	0.64	1.42	0.25	0.36	0.72	1.33

Table 12: 4-factor Almost Affine models: summary statistics of linear approximation absolute pricing errors in basis points.



(a) $AA_1(4)$ (b) $AA_2(4)$

Figure 8: Time series of linear approximation absolute pricing errors in basis points implied by 4-factor Almost Affine models

B.5 Statistical model selection criteria

	$A_0(4)$	$A_1(4)$	$AA_1(4)$	$A_2(4)$	$AA_2(4)$
llk	22130	187	429	171	389
# of parameters	36	3	14	3	28
AIC	-44189	-369	-829^*	-336	-721^*
AICc	-44184	-368	-825^*	-335	-711^*
BIC	-44029	-356	-767^*	-322	-597^*

Table 13: 4-factor models: log likelihood (ll), number of parameters, and fitness scores relative to the constant-volatility model $A_0(4)$.