

Risk and Return Trade-off in the U.S. Treasury Markets

Eric Ghysels* Anh Le[†] Sunjin Park[‡] Haoxiang Zhu[§]

December 14, 2013

Preliminary. Comments welcome

Abstract

This paper characterizes the risk-return trade-off in the U.S. Treasury markets. We propose a discrete-time no-arbitrage term structure model, in which bond prices are solved in closed form and the conditional variances of bond yields are decomposed into a short-run component and a long-run component, each of which follows a GARCH-type process. Estimated using Treasury yields data from January 1962 to August 2007, our model simultaneously matches the conditional volatility dynamics and the deviation from the expectations hypothesis in the data. We find that a higher short-run volatility component of bond yields significantly predicts a higher future excess return, above and beyond the predictive power of the yields. The long-run volatility component does not predict bond excess returns.

Keywords: bond risk premium, stochastic volatility, term structure models

JEL classification: G12, C58

*Kenan–Flagler Business School and Department of Economics, University of North Carolina at Chapel Hill, eghysels@unc.edu

[†]Kenan–Flagler Business School, University of North Carolina at Chapel Hill, anh.le@unc.edu

[‡]Kenan–Flagler Business School, University of North Carolina at Chapel Hill, Sunjin_Park@kenan-flagler.unc.edu

[§]MIT Sloan School of Management, zhuh@mit.edu

1 Introduction

Does a higher risk lead to a higher expected excess return in the U.S. Treasuries markets? To answer this fundamental question, we need an accurate measure of risk as well as an accurate account of returns predictability in the data. Most term structure models to date have difficulties providing both simultaneously. For example, although Gaussian affine models can accurately replicate [Campbell and Shiller \(1987\)](#)'s characterization of return predictability, these models assume that yields are homoskedastic, therefore missing entirely the time-varying volatility in the data. Other affine term structure models with stochastic volatility typically fail to generate the patterns of return predictability documented by [Campbell and Shiller \(1987\)](#).¹ Reasonable volatility processes, such as the GARCH model of [Bollerslev \(1986\)](#) or the EGARCH model of [Nelson \(1991\)](#), rarely make their way into the no-arbitrage term structure literature because, if those stochastic volatilities are imposed on the risk-neutral dynamics, one loses the analytical solutions to bond prices.

In this paper, we propose a discrete-time model that integrates the advantages of both the affine term structure models and the GARCH models of volatility. Not only do we retain the tractability of the affine models, we also inherit the ability of GARCH models to accurately capture time-varying volatilities of yields. The key to our approach is an asymmetric treatment of conditional volatility under the physical (\mathbb{P}) and risk-neutral (\mathbb{Q}) measures. Specifically, since conditional variances under \mathbb{P} and \mathbb{Q} need not be the same in a discrete-time setup (see e.g. [Le, Singleton, and Dai \(2010\)](#) for a discrete-time stochastic volatility model in which \mathbb{P} and \mathbb{Q} conditional variances are distinct), we choose to model the \mathbb{Q} conditional variances in a way that guarantees analytical affine pricing, while letting the \mathbb{P} conditional variances follow a GARCH-type process.

Following the ARCH-in-mean literature, pioneered by [Engle, Lilien, and Robins \(1987\)](#), we allow the \mathbb{P} conditional variances to affect the physical drifts of the state variables, the principal components yields. This provides a direct channel through which volatility, a measure of risk, can potentially forecast future yields and bond excess returns, hence the risk-return trade-off. Importantly, because the feedback of volatility into the physical drifts happens entirely under \mathbb{P} , it does not interfere with tractable bond pricing under \mathbb{Q} .

More specifically, we allow the \mathbb{P} conditional variances to affect the physical drifts of the state variables in four ways. First, in the spirit of [Engle and Lee \(1999\)](#), we let the \mathbb{P} conditional variances of yields to be driven by a long-run component and a short-run component, each of which follows its own GARCH-like process with different degrees of persistence. From these components, we construct the short-run and long-run volatility components of two equal-weighted yield portfolios: a near-maturity portfolio consisting of five yields with one to five years to maturity and a far-maturity portfolio consisting of five yields with six to ten years to maturity. (The cutoff at five-year maturity is motivated by a diagnostic analysis, but our results are not sensitive to this choice.) These four volatility components allow us to differentiate the contributions to risk premiums of long-run and

¹See, for example, [Dai and Singleton \(2002\)](#) and [Joslin and Le \(2012\)](#). In the context of no-arbitrage affine term structure models, [Joslin and Le \(2012\)](#) discuss in depth why there must be trade-offs between fitting volatility and the predictability of bond returns.

short-run volatility components of both the long end and short end of the yield curve.

Using weekly Treasury yields data from January 1962 to August 2007, we find a significantly positive relation between risk and return in U.S. Treasury markets. A higher conditional volatility this week predicts a lower yield level next week, thus a higher bond excess return. Notably, it is the short-run component of volatility, not the long-run component, that matters for return predictability. According to our estimates, the long-run component has a half-life of more than 60 years, whereas the short-run component has a half life of about two years.² Therefore, the risk-return relation is unlikely the results of transitory shocks or trading frictions, which we expect to move at much higher frequencies; nor is it driven by the extremely persistent time trend in volatility. The return-predicting short-run volatility component moves at roughly the business-cycle frequency, which we find reasonable and intuitive.

Moreover, the return-predicting power of the short-run volatility component predominantly comes from the short-end of the yield curve. Volatilities of far-maturity yields do not have additional predictive power for future yields once we control for volatility of the short end. Further, the slope and the curvature of the yield curve are not predicted by any volatility components that we study.

Putting all together, our main evidence is that a higher short-run volatility component of near-maturity yields predicts a lower yield level next week and thus a higher excess return. The economic magnitudes are large. For example, the volatility factor accounts for 14%, 42%, and 40% of the predictable components of weekly excess returns on one-year, five-year, and ten-year zero coupon bonds, respectively. (The other predictive factors are the principal components of yields.)

In addition to the empirical findings, our paper also contributes to the literature on term structure modeling. Most affine term structure models with stochastic volatilities imply that the conditional variances of yields are perfectly explained, or “spanned,” by yields themselves. By imposing this spanning condition, these affine stochastic volatility models are potentially restrictive in at least two aspects. First, there is considerable evidence that volatility is not fully spanned by yields (see [Collin-Dufresne and Goldstein, 2002](#); [Collin-Dufresne, Goldstein, and Jones, 2009](#); and [Andersen and Benzoni, 2010](#)). Second, imposing a spanning condition can induce a “cross-measures” tension, in the sense of [Joslin and Le \(2012\)](#), that could prevent a no-arbitrage model from fully capturing the predictability patterns of bond returns in the data.³ Our model addresses both issues by decoupling the risk-neutral conditional variances from their physical counterparts. As a result, not only do we match the volatility

²The long-run component may appear extremely persistent, but such persistence is not uncommon. For example, in a related context, [Stock and Watson \(2007\)](#) propose to include a random walk component to the (log) volatility of the inflation process. They find support for this model using U.S. inflation data over a similar sample period to ours. Given the intimate relation between nominal yields and inflation, it is not surprising that we detect a similar degree of persistence in the long run volatility component of yields.

³Specifically, in an affine setup, volatility factors must be autonomous to remain strictly positive under both the \mathbb{P} and \mathbb{Q} measures. This autonomy requirement when applied to spanned volatilities requires that the \mathbb{P} and \mathbb{Q} feedback matrices share some common left-eigenvectors. Understandably, the resulting closeness between the \mathbb{P} and \mathbb{Q} feedback matrices limits the ability of the model to explain returns predictability.

dynamics of yields very well, our model also closely replicates the return predictability patterns documented by [Campbell and Shiller \(1987\)](#).

Our modeling approach using a GARCH-like volatility process complements a few existing approaches. For example, compared with the unspanned stochastic volatility models by [Collin-Dufresne and Goldstein \(2002\)](#) and [Collin-Dufresne, Goldstein, and Jones \(2009\)](#), our model does not restrict model parameters to eliminate the spanning of conditional volatilities by yields.⁴ Our approach also differs from the “general stochastic volatility model” proposed by [Trolle and Schwartz \(2009\)](#) and the high-frequency approach proposed by [Cieslak and Povala \(2013\)](#). A GARCH-like model allows us to identify yield volatility dynamics with considerable precision over a long sample period, not restricted by the availability of high-frequency data.

Finally, we add to the ARCH-in-mean literature in two important aspects. First, because both the yield volatility and the principal components of yields can forecast future yields, our model allows two determinants of the dynamics of risk premiums: the quantity of risk and the market price of risk. (Evidence supporting the simultaneous presence of both channels can be found in [Dai and Singleton \(2000\)](#) and [Dai and Singleton \(2002\)](#) as well as our motivating exercises in the next section.) By contrast, ARCH-in-mean models typically only allow the quantity of risk, but not the market price of risk, to explain risk premiums (see, for example, [Engle, Lilien, and Robins, 1987](#); and [Adrian and Rosenberg, 2008](#)).⁵ Second, whereas the ARCH-in-mean models are typically applied to bond excess returns of individual maturities, our model is designed to match bond prices/yields as well as bond excess returns of all maturities, tied together by the no-arbitrage condition.⁶ Our model, therefore, offers a more coherent characterization of the risk-return trade-off in Treasury markets.

2 Motivating Exercises

In this section, we conduct several simple exercises that shed light on the potential relations between risks and returns in the bond markets. These will serve as guidance for us in designing our models for subsequent analysis.

Our first exercise is to estimate an ARCH-in-mean model in the spirit of [Engle, Lilien, and Robins \(1987\)](#) (ELR):

$$xr_{n,t+1} = \alpha + \delta h_t + e_{t+1} \tag{1}$$

⁴These restrictions require that the volatility factors of a model have certain mean reversion rates in order to result in an exact cancellation of the convexity effects. See [Joslin \(2007\)](#) for an in-depth discussion.

⁵[Haubrich, Pennacchi, and Ritchken \(2012\)](#) also use a GARCH process to model yields volatility in a no-arbitrage setup. Their model is nested within the completely affine class of [Dai and Singleton \(2000\)](#). As a result, they only allow for the quantity of risk channel to determine bond risk premiums.

⁶To see how matching prices/yields is a much stronger requirement than matching excess returns, note that for a model with a stochastic discount factor M_{t+1} to match excess returns R_{t+1}^e , the requirement is that $E_t[M_{t+1}R_{t+1}^e] = 0$. However, any scaled version of such a discount factor will also match excess returns equally well. As a result, the requirement to match excess returns cannot pin down the conditional mean of the stochastic discount factor $E_t[M_{t+1}]$, the one period bond price.

where $xr_{n,t+1}$ denotes the weekly excess return on an n -period zero coupon bond, starting at time t and realized one week later, at time $t + 1$. h_t denotes the conditional volatility of the shocks e_{t+1} and h_t is assumed to follow an ARCH process with thirteen weekly lags (one quarter):

$$h_t^2 = \alpha_A + \sigma_A \sum_{i=1}^L w_i e_{t-i}^2, \quad (2)$$

where $L = 13$. In ELR's implementation of the ARCH process, the loadings on the lag squared residuals (w_i 's) are set to fixed constants to avoid estimation uncertainty. ARCH models may indeed involve quite a few parameters, which led to the popularity of GARCH models which we will consider below. We adopt first an ARCH specification without parameter proliferation, opting for a flexible, but parsimonious lag structure. This is reminiscent of MIDAS polynomials and therefore follow a setup where the w_i 's are hyper-parameterized via a normalized beta probability density function, $w_i = (L + 1 - i)^{\theta-1} / \sum_{i=1}^L (L + 1 - i)^{\theta-1}$. This weighting scheme (through one parameter θ) is parsimonious but nonetheless is known to reasonably capture volatility dynamics in the data.⁷ By keeping the lag length short we have implicitly a short-run volatility effect in mind, a specification we will contrast with the GARCH discussed next, which will feature a more persistent volatility process.

The expected component of equation (1) says that the expected excess returns, $E_t[xr_{n,t+1}]$, should be linearly related to bonds' volatility as given by h_t . If the estimate of δ is positive and statistically significant, it suggests that there is a positive risk-return relationship in bonds markets.

We use zero yields data from [Gurkanyak, Sack, and Wright \(2007\)](#) (GSW).⁸ The data is sampled weekly, starting in January 1962 and ending in August 2007.⁹ The GSW dataset is useful for our purposes because it is considerably smoothed.¹⁰ Thus, volatility measures constructed from this data are less susceptible to the influence of outliers. Nonetheless, the short-maturity yields from the GSW data involve a high degree of extrapolation. As such, for maturities shorter than six months, we bootstrap zero yields from CRSP's raw bond prices using the standard Fama-Bliss algorithm. We implement the estimations of equations (1) and (2) using QMLE.

[Table 1](#) reports the results for twenty different maturities, from 6- to 120-month. Each row corresponds to a different maturity n in (1). The second column, labelled ARCH-M, reports the estimates of δ . Standard errors are calculated from the Newey-West covariance matrix, constructed using thirteen lags. Consistent with ELR, we find that the estimates of δ are

⁷For more discussion on MIDAS polynomial specifications, see e.g. [Ghysels, Sinko, and Valkanov \(2007\)](#) and [Ghysels \(2013\)](#).

⁸The data can be downloaded from <http://www.federalreserve.gov/pubs/feds/2006>.

⁹Our sample ends in August 2007 for at least two reasons. First, as is currently typical for the term structure literature, we avoid the era with near-zero interest rate in the wake of the global financial crisis. Second, recent work has shown that for equities the risk-return relation breaks down during financial crisis as flight-to-quality concerns dominate (see e.g. [Ghysels, Plazzi, and Valkanov \(2013\)](#) and references therein).

¹⁰It is possible that the smoothing algorithm used by GSW also filters out information from the yield curves and thus weakens potential predictive relationships in the data.

Table 1: Predictability of weekly excess returns from ARCH volatilities, GARCH volatilities and yields PCs. *, **, *** denote the conventional significance levels of 1%, 5%, and 10%, respectively.

Maturity	ARCH-M	GARCH-M	OLS: PC			GLS: ARCH and PC	
	δ	δ	$\delta_{PC1} \times 10^4$	$\delta_{PC2} \times 10^4$	$\delta_{PC3} \times 10^4$	δ_{ARCH}	$\delta_{PC2} \times 10^4$
6m	0.21***	0.18***	0.20	0.10	-0.34	0.11	-0.02
12m	0.24***	0.21***	0.27	0.81*	0.94	0.14*	0.68**
18m	0.19***	0.16***	0.31	1.23*	1.47	0.18**	1.06*
24m	0.17***	0.14***	0.35	1.63*	1.87	0.22***	1.36*
30m	0.15***	0.12***	0.39	2.02**	2.14	0.23***	1.60*
36m	0.13***	0.10**	0.41	2.40**	2.31	0.23***	1.82*
42m	0.11***	0.08*	0.44	2.79**	2.39	0.23***	2.04*
48m	0.10**	0.07	0.46	3.17**	2.40	0.22***	2.27*
54m	0.09**	0.06	0.48	3.55**	2.34	0.20***	2.52*
60m	0.08**	0.05	0.50	3.92**	2.24	0.18***	2.80*
66m	0.08***	0.05	0.52	4.30**	2.08	0.17***	3.14*
72m	0.08***	0.05	0.54	4.67**	1.90	0.15**	3.55**
78m	0.08***	0.05	0.56	5.03***	1.67	0.14**	4.03**
84m	0.09***	0.04	0.58	5.39***	1.42	0.12**	4.54**
90m	0.09***	0.04	0.61	5.75***	1.14	0.11*	5.08**
96m	0.09***	0.04	0.63	6.10***	0.84	0.10*	5.62***
102m	0.10***	0.04	0.66	6.45***	0.50	0.10	6.17***
108m	0.10***	0.04	0.69	6.79***	0.15	0.09	6.69***
114m	0.11***	0.04	0.73	7.12***	-0.23	0.09	7.16***
120m	0.11***	0.05	0.76	7.45***	-0.63	0.09	7.57***

significantly positive across the entire maturity spectrum, suggesting a positive risk-return relationship.

In the second exercise, we replace the ARCH process in (2) by a GARCH(1,1) process:

$$h_t^2 = \alpha_G + \beta h_{t-1}^2 + \sigma_G e_t^2. \quad (3)$$

All else is kept identical to the first exercise.

The estimates of δ are reported in the third column, labelled GARCH-M, of Table 1. Notably, although the estimates of δ are all positive, they are only significant for the short range of maturities up to 42 months (three and a half years). Note that the GARCH volatility can be written as an infinite sum of lagged squared residuals, $h_t^2 = \text{constant} + \sigma_G \sum_{i=0}^{\infty} \beta^i e_{t-i}^2$. The contrast between the ARCH-M and GARCH-M columns of Table 1 suggests that more recent volatility may be more important in determining the risk premia in bond markets.

The δ estimates for both the ARCH-M and GARCH-M columns have a similar pattern: they are highest (most positive) at the 12-month maturity, then decreasing as maturity increases to about five years, and finally flattening out for longer maturities. The weaker risk-return relation along the maturity spectrum suggests that (a) excess returns become less predictable as maturity lengthens, (b) volatility becomes less predictive for longer-dated bonds,

or both (a) and (b). Either way, this evidence suggests that distinguishing short-maturity volatility and long-maturity volatility may prove fruitful for a coherent understanding of the risk-return relationship across all maturities.

In the two exercises implemented so far, we only allow the time-varying quantity of risk, h_t , to predict excess returns. Absent from this setup is an independent role for time-varying market prices of risks in determining bond risk premia. This role is provided by the large literature on Gaussian term structure models, in which the quantity of risk (yields volatility) is assumed constant and thus all returns predictability is generated by the time variation in market prices of risks.

In the third exercise, we run a simple OLS regression that predicts weekly excess returns using three principal components (PCs) of yields:

$$xr_{n,t+1} = \alpha_{PC} + \delta_{PC1}PC1_t + \delta_{PC2}PC2_t + \delta_{PC3}PC3_t + e_{t+1}. \quad (4)$$

In standard affine Gaussian term structure models, these three PCs (PC1-3) reasonably capture the time variation in the state variables, which also govern the market prices of risks implied by these models. We construct the PCs from yields with maturities of 6 months, 1 year, 2 years, 3 years, 5 years, 7 years, and 10 years.¹¹ As is standard, the first three PCs are the level, slope, and curvature of the yield curve, respectively.

Estimated coefficients for the OLS regression in (4) are reported in the columns under the heading “OLS:PC” of [Table 1](#). Consistent with the established results in the literature (for example, [Campbell and Shiller \(1991\)](#) and [Fama and Bliss \(1987\)](#)), we find that the slope factor is strongly predictive of future excess returns, particularly so for the longer dated bonds.

In the last exercise, we combine the two sets of predictive variables, volatilities and yield PCs, in the following regression:

$$xr_{n,t+1} = \alpha_{PC} + \delta_{ARCH}h_t + \delta_{PC1}PC1_t + \delta_{PC2}PC2_t + \delta_{PC3}PC3_t + e_{t+1}, \quad (5)$$

where we use the ARCH volatility h_t implied from the first exercise. To save space, we only report estimates of δ_{ARCH} and δ_{PC2} in the last two columns of [Table 1](#). We see that volatility is a significant predictor beyond the three PCs, and it is significantly if it is constructed from 1-year to 8-year yields.

Taken together, the exercises in this section suggest that (1) yields volatilities can forecast excess returns above and beyond the information embedded in the current yield curve, (2) information in more recent volatility seems more important in determining bond risk premia, and (3) volatilities constructed from short-maturity yields and long-maturity yields have different information content for excess returns. These three observations directly drive our modeling approach in the next section.

¹¹We choose these maturities since they represent the most liquid segments of the yield curve. These maturities are also covered by the Federal Reserve Board’s H.15 releases.

3 Model

In this section, we formally develop a model of the term structure of interest rates in discrete time.

3.1 The risk-neutral dynamics and bond pricing

Our model specification for the risk-neutral dynamics is standard. The vector of state variables, X_t , follows a Gaussian VAR(1) under the risk-neutral (\mathbb{Q}) measures, and the short rate, r_t , is a linear function of X_t :

$$X_{t+1} = K_0^{\mathbb{Q}} + K_1^{\mathbb{Q}}X_t + \epsilon_{t+1}^{\mathbb{Q}} \text{ with } \epsilon_{t+1}^{\mathbb{Q}} \sim N(0, \Sigma_X), \quad (6)$$

$$r_t = \delta_0 + \delta_1'X_t. \quad (7)$$

It immediately follows from the affine structure of the setup that zero-coupon yields at all maturities are affine in X_t :

$$y_{n,t} = A_{n,X} + B_{n,X}X_t \quad (8)$$

with the loadings $(A_{n,X}, B_{n,X})$ obtained from standard yield pricing recursions.

3.2 The time-series dynamics

We assume that X_t follows affine dynamics with conditionally Gaussian innovations under the physical (\mathbb{P}) measure:

$$X_{t+1} = K_0 + K_1X_t + K_VV_t + \epsilon_{t+1} \text{ with } \epsilon_{t+1} \sim N(0, \Sigma_t), \quad (9)$$

where matrices K_0 , K_1 , and K_V are $N \times 1$, $N \times N$, and $N \times M$, respectively. We now turn to the specifications of the conditional variance, Σ_t , and the GARCH-in-mean term, V_t .

Dynamics of conditional variance Σ_t

Given the evidence provided in [Section 2](#) that recent and distant volatilities might have different information for excess returns, we explicitly model the long-run and short-run components of Σ_t in the spirit of [Engle and Lee \(1999\)](#) (EL):

$$\Sigma_t = S_t + L_t, \quad (10)$$

$$S_t = \rho_S S_{t-1} + \alpha(\epsilon_t \epsilon_t' - \Sigma_{t-1}), \quad (11)$$

$$L_t = \Sigma_X(1 - \rho_L) + \rho_L L_{t-1} + \phi(\epsilon_t \epsilon_t' - \Sigma_{t-1}), \quad (12)$$

where ρ_S , ρ_L , α , and ϕ are all positive scalars. Σ_X is the same variance matrix in [\(6\)](#) and thus is positive semi-definite (psd). Clearly, the Gaussian models are obtained as a special case when $\rho_S = \rho_L = \alpha = \phi = 0$.

The interpretation of this model is straightforward. The total variance matrix Σ_t is decomposed into a short-run component, S_t , and a long-run component, L_t . This decomposition is a simple way to differentiate the impact of recent volatilities on returns dynamics from that of distant volatility information. Each component follows its own autoregressive process with different persistence, captured by ρ_S and ρ_L . Without loss of generality, we impose the restriction that $\rho_S < \rho_L$. In both equations (11) and (12), the last term, $(\epsilon_t \epsilon_t' - \Sigma_{t-1})$, represents news about volatility. A piece of volatility news dissipates at a faster rate for S_t than for L_t .

In addition, the lack of the intercept term in the AR(1) process of S_t implies that the population mean of S_t is zero. In this sense, L_t is a low-frequency trend component of Σ_t , whereas S_t is a high-frequency, transitory component around zero.

Finally, to guarantee that Σ_t is strictly positive definite, we impose the restriction:

$$1 > \rho_L > \rho_S > \alpha + \phi, \quad (13)$$

in addition to the positivity requirement for ρ_S , ρ_L , α , and ϕ . Condition (13) is imposed by EL in their univariate setting. Along the lines of proofs in EL, we can show that condition (13) implies that Σ_t and L_t are positive definite in our multivariate model.¹²

We observe that our model naturally gives rise to unspanned stochastic volatility (USV) under the \mathbb{P} measure. By construction, the conditional variance Σ_t , the sum of lagged “squared residuals,” is not spanned by X_t . In this regard, our model presents a significant departure from the traditional affine models with spanned volatilities. Collin-Dufresne and Goldstein (2002), Collin-Dufresne, Goldstein, and Jones (2009), Andersen and Benzoni (2010), among others, show the importance of allowing for USV in bond markets.

The GARCH-in-mean term V_t

As in the (G)ARCH-in-mean literature, the role of V_t is to summarize volatility information relevant for forecasting excess returns. In our multivariate setting, the challenge is dimensionality. For an N -factor model (N being the dimension of X_t), there are $N(N + 1)$ unique entries in the matrices S_t and L_t . A typical three-factor model has 12 conditional variance and covariances. Clearly, including all of these elements in V_t would make the model over-parameterized.

To keep the model parsimonious, we focus on the conditional volatilities of two particular portfolios of bond yields. The first portfolio is an equal-weighted portfolio of yields with maturities 1, 2, 3, 4, and 5 years. The second is an equal-weighted portfolio of yields with

¹²In particular, the long-term component, L_t , can be expressed as an infinite-order polynomial of the product $\epsilon_t \epsilon_t'$. Under condition (13), all coefficients of this polynomial are non-negative, which guarantees that L_t is positive definite. To see how the positive definiteness of Σ_t follows from this, we add equations (11) and (12) side by side to arrive at:

$$\Sigma_t = \Sigma_X(1 - \rho_L) + (1 - \rho_L)L_{t-1} + (\rho_S - \alpha - \phi)\Sigma_{t-1} + (\alpha + \phi)e_t e_t'$$

Conditions (13) and the positivity of α and ϕ means that all the scalar loadings are positive. Hence, as long as L_{t-1} is positive definite, so is Σ_t , by induction.

maturities 6, 7, 8, 9, and 10 years. This choice is motivated from the exercise in [Section 2](#) that yield volatilities inferred from short to medium range of the yield curve contain information for excess returns beyond the yield curve factors; yield volatilities inferred from longer maturities of yields, less so. This separation is nondegenerate as long as $N \geq 2$. By separating the two yield portfolios by maturities, we will uncover the impacts of volatilities in the short and long ends of the yield curve. We emphasize that the choice of the cutoff point, 5 years, is not critical to our results. We also estimate model using a cutoff point of 3 years, and the results are similar.

With two yield portfolios and two horizons, V_t contains four elements. These four entries can be explicitly derived from the conditional variance Σ_t in the following way. We denote by $B_{1-5,X}$ and $B_{6-10,X}$ the weighting vectors of the two portfolios on the factor X_t . Thus, by (8), the conditional variance of the first yield portfolio is

$$\text{Var}_t \left[\frac{1}{5} \sum_{n=1}^5 y_{n,t+1} \right] = B_{1-5,X} \Sigma_t B'_{1-5,X}. \quad (14)$$

A similar calculation gives the conditional variance of the second yield portfolio, with maturities 6–10 years. The long-term variance of these two portfolios can be computed in a similar manner by replacing Σ_t with L_t . Putting them together, we write

$$V_t = \begin{pmatrix} \sqrt{B_{1-5,X} \Sigma_t B'_{1-5,X}} - \sqrt{B_{1-5,X} L_t B'_{1-5,X}} \\ \sqrt{B_{1-5,X} L_t B'_{1-5,X}} \\ \sqrt{B_{6-10,X} \Sigma_t B'_{6-10,X}} - \sqrt{B_{6-10,X} L_t B'_{6-10,X}} \\ \sqrt{B_{6-10,X} L_t B'_{6-10,X}} \end{pmatrix}. \quad (15)$$

The first (second) element of V_t is the short-run (long-run) volatility component of the yield portfolio with maturities 1–5 years. The third (fourth) element of V_t is the short-run (long-run) volatility component of the yield portfolio with maturities 6–10 years. Note that the first element of V_t is written this way, instead of “ $\sqrt{B_{1-5,X} S_t B'_{1-5,X}}$,” because S_t needs not be positive definite. But L_t is.

Implications on bond excess returns

Combining (8) and (9), we can write the one-period expected excess return on the n -period bond as:

$$\begin{aligned} E_t[xr_{n,t+1}] &= \text{constant} + (nB_{n,X} - (n-1)B_{n-1,X}K_1 - B_{1,X})X_t \\ &\quad - (n-1)B_{n-1,X}K_V V_t. \end{aligned} \quad (16)$$

Clearly, we capture a volatility component as well as a pure yield curve component of bond risk premia. As n varies, equation (16) gives us a term structure of risk-return relations across the maturity spectrum.

We note that the derivation of expected excess returns in (16) can also be obtained through the stochastic discount factor implied by our model. Specifically, given our specifications of the \mathbb{P} and \mathbb{Q} dynamics, the implied stochastic discount factor can be written as:

$$M_{t,t+1} = e^{-r_t} \frac{f_t^{\mathbb{Q}}(X_{t+1})}{f_t^{\mathbb{P}}(X_{t+1})} \quad (17)$$

where $f_t^{\mathbb{Q}}(X_{t+1})$ and $f_t^{\mathbb{P}}(X_{t+1})$ denote the conditional densities of X_{t+1} given the time- t information set under the \mathbb{Q} and \mathbb{P} measures, respectively. Since the state variables X_t share a common support under both measures, $M_{t,t+1}$ defines a valid and strictly positive pricing kernel (which rules out arbitrage).

To summarize, our model is fully characterized by the risk neutral dynamics in (6) and (7), the time series dynamics in (9), and the associated volatility dynamics given by (10), (11), and (12). The full parameter set is given by: $\Theta_X = (\delta_0, \delta_1, K_0^{\mathbb{Q}}, K_1^{\mathbb{Q}}, \Sigma_X, K_0, K_1, K_V, \rho_S, \alpha, \rho_L, \phi)$.

3.3 Econometric identification

Canonical setup

To obtain econometric identification, we apply the standard rotations of the affine term structure literature (see, for example, Dai and Singleton, 2000). Since the state variables in our setup are not bounded, any rotation from X to $Z = U_0 + U_1 X$ for any (U_0, U_1) is admissible, as we show in Appendix A. In other words, for any affine transformation of X to Z , we will obtain another observationally equivalent instance of our model, characterized by the same set of equations (6)-(13), with a different parameter set Θ_Z . Explicit mappings between Θ_X and Θ_Z (as functions of (U_0, U_1)) are provided in Appendix A.

By rotating the state variables freely, we obtain econometric identification using the canonical setup of Joslin, Singleton, and Zhu (2011). Specifically, under this canonical setup, δ_1 is a vector of ones, $K_0^{\mathbb{Q}}$ is a vector of zeros, and $K_1^{\mathbb{Q}}$ is of Jordan form. Thus, the risk-neutral means of the state variables are zeros, and the intercept term in the short rate equation, δ_0 , becomes the risk-neutral long-run mean of the short rate r_t . Adopting the notation of Joslin, Singleton, and Zhu (2011), we replace δ_0 by $r_{\infty}^{\mathbb{Q}}$ in the canonical setup. Thus, the econometrically identified parameter set becomes $\Theta_X = (r_{\infty}^{\mathbb{Q}}, K_1^{\mathbb{Q}}, \Sigma_X, K_0, K_1, K_V, \rho_S, \alpha, \rho_L, \phi)$.

Rotation to observable yield portfolios

Consider a $J \times 1$ vector of yields y_t . The affine structure of bond yields,

$$y_t = A_X + B_X X_t, \quad (18)$$

means that any non-degenerate yields portfolios \mathcal{P}_t characterized by a $N \times J$ loadings matrix W must be affine in the states: $\mathcal{P}_t = WA + WBX_t$.

Our ability to freely rotate once again means that we can rotate X to \mathcal{P} and thus we can simply replace our canonical model by one in which \mathcal{P}_t serves as state variables. As shown by

Joslin, Singleton, and Zhu (2011) and Joslin, Le, and Singleton (2012), using yields portfolios, which are observable, as state variables greatly enhances the (numerical) identification of model parameters. With \mathcal{P}_t as the state variables, we denote the parameter set by:

$$\Theta_{\mathcal{P}} = (r_{\infty}^{\mathbb{Q}}, K_1^{\mathbb{Q}}, \Sigma_{\mathcal{P}}, K_0, K_1, K_V, \rho_S, \alpha, \rho_L, \phi).$$

The parameters $(r_{\infty}^{\mathbb{Q}}, K_1^{\mathbb{Q}}, \rho_S, \alpha, \rho_L, \phi)$ are invariant to rotations and thus are identical across Θ_X and $\Theta_{\mathcal{P}}$. The remaining parameters are rotation-specific. For example, Σ_X refers to the \mathbb{Q} conditional variance of the latent state variables under the canonical setup, whereas $\Sigma_{\mathcal{P}}$ refers to the \mathbb{Q} conditional variance of \mathcal{P}_t . As in Joslin, Singleton, and Zhu (2011), the first three parameters $(r_{\infty}^{\mathbb{Q}}, K_1^{\mathbb{Q}}, \Sigma_{\mathcal{P}})$ determine the loadings of bond yields with \mathcal{P}_t as states. The next three parameters (K_0, K_1, K_V) determine the \mathbb{P} conditional mean of \mathcal{P}_t (given by equation (9) replacing X_t by \mathcal{P}_t). The last four parameters $(\rho_S, \alpha, \rho_L, \phi)$ together with $\Sigma_{\mathcal{P}}$ determine the dynamics of the \mathbb{P} conditional variances of \mathcal{P}_t (through equations (10), (11), (12) with Σ_X replaced by $\Sigma_{\mathcal{P}}$). Because the volatility factors, V_t , are volatility components of observable yield portfolios, they remain invariant to factor rotations. (See Appendix A for more details.)

Joslin, Le, and Singleton (2012) show that by letting \mathcal{P}_t be the lower order principal components (PCs) of bond yields, estimation of the model is least sensitive to assumptions regarding the observational errors of bond yields. Using this observation, we will use a representation of our model with \mathcal{P}_t being the first N PCs of bond yields in our empirical implementation. As a result, subsequent mentions of state variables should be understood as references to the first N PCs of bond yields.

3.4 Discussion of modeling choices

Our model can be viewed as a generalization of Gaussian no-arbitrage models, which can be recovered from our setup by setting Σ_t to a constant matrix ($\Sigma_{\mathcal{P}}$), and K_V to zeros. As explained earlier, we choose an asymmetric approach in which only the \mathbb{P} conditional variances are stochastic, whereas the conditional variances under \mathbb{Q} remain constant. This asymmetric treatment is possible because diffusion invariance need not hold in a discrete-time model.

Keeping the \mathbb{Q} -volatility constant has the benefit of parsimony. If generality were the objective, one would set the \mathbb{Q} conditional variances stochastic too, as long as analytical pricing remains feasible. One such model under \mathbb{Q} is provided by Le, Singleton, and Dai (2010) (a discrete-time counterpart to the stochastic volatility models in Dai and Singleton (2000)), in which the conditional variances are time-varying but affine in states in a way that affine pricing of yields is preserved.

We argue, however, that as long as yields are affine in states, the *direct* effects of alternative \mathbb{Q} dynamics on risk premiums are insensitive to the volatility structure under \mathbb{Q} . That is, as far as risk premium is concerned, setting a constant variance under \mathbb{Q} is almost without loss of generality. The remaining of this subsection goes through the logic of this argument, based on analysis by Joslin and Le (2012).

Consider the expression for expected excess returns in (16). The direct channel that alternative \mathbb{Q} dynamics have on risk premiums is through the yields loadings B . The other parameters (K_1, K_V) come from the \mathbb{P} dynamics. Joslin and Le (2012) show that estimates of yields loadings B are very similar across different affine models with distinct volatility structures. Intuitively, since the \mathbb{Q} dynamics is typically strongly identified in the data, minimizing cross-sectional pricing errors has priority in maximum-likelihood estimations. Thus, regardless of the volatility structure, estimates of yield loadings B in affine models are typically very close to the unconstrained estimates obtained by regressing yields onto yields PCs.

Different \mathbb{Q} dynamics can also have *indirect* effects on the dynamics of risk premiums if they can somehow influence the estimates of the time series parameters K_1 and/or K_V in (16). Joslin and Le (2012) provide one example in the context of affine stochastic volatility models, in which the \mathbb{Q} conditional variances are at least partially priced in the cross-section of yields and thus are partially spanned by yields. For these models, Joslin and Le (2012) show that the positivity requirement under \mathbb{Q} for volatility can impose strong constraints on the time series parameters. Specifically, in an affine setup, volatility factors must be autonomous to remain strictly positive under both the \mathbb{P} and \mathbb{Q} measures. When applied to spanned volatilities, this autonomy requires that the \mathbb{P} and \mathbb{Q} feedback matrices share some common left-eigenvectors. The resulting closeness between the \mathbb{P} and \mathbb{Q} feedback matrices brings the model closer to the expectations hypothesis (in which the \mathbb{P} and \mathbb{Q} conditional means are the same). As a result, these constraints can prevent a model from fully explaining the risk-return relation in the data—the very focal point of our study.

There are at least two ways to address the tension documented by Joslin and Le (2012). One is our approach, in which the time series parameters (K_1, K_V) are unambiguously unconstrained since they are free parameters. The other is to adopt the risk-neutral setup of an affine stochastic volatility model but then impose constraints on the risk-neutral parameters so that our model exhibits (completely) unspanned stochastic volatility (USV, see Collin-Dufresne and Goldstein, 2002; and Collin-Dufresne, Goldstein, and Jones, 2009). But with volatility not priced in the cross-section, the yields pricing equation induced by USV and the yields pricing equation (8) of our setup are observationally equivalent. In both cases, what appear on the right hand side of (8) are pure yield curve factors and not stochastic volatility. It follows that either way the models’ implications for bond risk premia—the decomposition of expected excess returns into a volatility component and a pure yield curve component in (16)—will be similar.¹³ Thus, for simplicity, we maintain the assumption that the conditional variance matrix is constant under \mathbb{Q} .

4 Results

In this section we present the estimation results of the model of Section 3. As is standard, we use $N = 3$ factors. Again, the three factors are chosen to be the first three PCs of bond

¹³The distinction between our model and a USV model is more pronounced for nonlinear securities such as calls or puts. This is not our focus, however.

yields, denoted by \mathcal{P}_t .

4.1 Estimation

We use the same dataset over the same sample period as described in [Section 2](#). We also adopt the same procedure in constructing the PCs of bond yields. Following [Joslin, Le, and Singleton \(2012\)](#), we assume that the first three PCs of bond yields are observed perfectly.¹⁴ As is standard, we assume that the remaining higher-order PCs, denoted by \mathcal{P}_e , are observed with i.i.d. uncorrelated Gaussian errors with one common variance. That is,

$$\mathcal{P}_{e,t}^o = \mathcal{P}_{e,t} + e_t \text{ and } e_t \sim N(0, I\sigma_e^2), \quad (19)$$

where σ_e is a scalar; and the superscript o indicates an observed quantity as opposed to a theoretical construct. Let W_e denote the loading matrix corresponding to the higher-order yield PCs. Then, $\mathcal{P}_{e,t}^o = W_e y_t^o$, whereas the theoretical counterpart of $\mathcal{P}_{e,t}^o$ can be computed by $\mathcal{P}_{e,t} = W_e(A_{\mathcal{P}} + B_{\mathcal{P}}\mathcal{P}_t)$. Recall that the yields loadings $A_{\mathcal{P}}$ and $B_{\mathcal{P}}$ can be obtained from the loadings A_X and B_X in [\(18\)](#) with necessary adjustments to account for the rotation from X to \mathcal{P} (see [Appendix A](#) for details).

The likelihood function of the observed data, \mathcal{L} , is given by:

$$\mathcal{L} = \sum_t f(\mathcal{P}_{t+1}|\mathcal{P}_t) + f(\mathcal{P}_{e,t+1}|\mathcal{P}_{t+1}), \quad (20)$$

where f denotes log conditional density. The first term captures the density of the time-series dynamics and can be written as:

$$\sum_t f(\mathcal{P}_{t+1}|\mathcal{P}_t) = \text{constant} - \frac{T}{2} \log(|\Sigma_t|) - \frac{1}{2} \sum_t \|\Sigma_t^{-1/2}(\mathcal{P}_{t+1} - K_0 - K_1\mathcal{P}_t - K_V V_t)\|_2^2, \quad (21)$$

where T denotes the sample length and $\|\cdot\|$ denotes the L^2 norm. The second term of the likelihood function captures the density of the cross-sectional fit and can be expressed as:

$$\sum_t f(\mathcal{P}_{e,t+1}|\mathcal{P}_{t+1}) = \text{constant} - \frac{T}{2} \log(\sigma_e^{2(J-N)}) - \frac{1}{2\sigma_e^2} \sum_t \|\mathcal{P}_{e,t+1} - W_e(A_{\mathcal{P}} + B_{\mathcal{P}}\mathcal{P}_{t+1})\|_2^2, \quad (22)$$

where J denotes the number of yields used in estimation.

Estimates of the model parameter set, $\Theta_{\mathcal{P}}$, and the standard deviation of pricing errors, σ_e , are obtained by maximizing the likelihood of observed data \mathcal{L} . For inferences, we do not use classical MLE standard errors. Rather, we use standard procedures to convert our ML

¹⁴[Joslin, Le, and Singleton \(2012\)](#) show that this assumption, as opposed to assuming individual yields are observed perfectly, guarantees that our estimates are close to those obtained from the Kalman filter with more general distributions of observational errors.

estimation problem into a set of GMM moment conditions.¹⁵ Standard errors that are robust to serial autocorrelation and heteroskedasticity in the residual errors are then obtained using the [Newey and West \(1987\)](#) matrix.

4.2 Model diagnosis

Our first step in the estimation is to diagnose the individual effect of the four entries of V_t in our model. Recall that V_t affects the conditional means of the state variables through K_V , a 3×4 matrix. The three rows of K_V correspond to the three PCs used as state variables. The first two columns of K_V correspond to the first two elements of V_t , which capture the effects of short-run and long-run components of the short end of the yield curve. The last two columns of K_V correspond to the last two elements of V_t , which capture the effects of short-run and long-run components of the long end of the yield curve.

To understand the individual effect of each element of V_t for each element of the pricing state variables \mathcal{P}_t , we estimate 12 specifications of the model by allowing one entry of K_V to be nonzero at a time. For example, the first specification only allows the (1,1) entry of K_V to be a free parameter, and sets all other entries of K_V to be zero. This specification therefore solely examines the effect of the first entry of V_t —the short-run volatility component in near-maturity bond yields—on the level of the term structure of yields. Similarly, the second specification only allows the (1,2) entry of K_V to be a free parameter, and examines the effect of the first entry of V_t on the slope of yields.

[Table 2](#) reports the results. We see that among specifications 1–12, only specifications 1 and 3 lead to a statistically significant estimate of K_V . A higher first entry of V_t , i.e. a higher short-run volatility component in near-maturity bond yields, predicts a lower PC1 next week. So does a higher third entry of V_t , i.e., a higher short-run volatility component in far-maturity bond yield. Because a lower PC1 is associated with a higher bond excess returns, this evidence suggests that there is a positive risk-return relation in bond markets, but the risk must be measured as the short-run component.

More concretely, the effect of V_t on bond risk premiums is spelled out exactly in equation (16). For each bond with n periods to maturity, the effect of V_t on its risk premium is the (negative of the) product of the exposure of the bond to each risk factor, as captured by the yield loading B , and the effect V_t has on the forecast of future PCs, as captured by K_V . For specifications 1 and 3, for which the last two rows of K_V are set to zeros, the last two elements of B are inconsequential, i.e., V_t only affects bond risk premium through exposure to level risk. Moreover, all yield loadings on the level factor are positive and precisely estimated (since the \mathbb{Q} dynamics is very strongly identified). As a result, for specifications 1 and 3, a negative and significantly estimated estimate for the non-zero entry of K_V translates into a positive trade-off between risk and expected excess returns.

In specification 13, we allow both the (1,1) and (1,3) entries of K_V to be free parameters, and find that only the (1,1) entry is significant. This further indicates that near-maturity

¹⁵Specifically, we take the first-order derivative of \mathcal{L} with respect to each parameter being estimated. Let \mathcal{L}_t be the time- t element of \mathcal{L} ($\mathcal{L} = \sum_t \mathcal{L}_t$). This first-order condition gives rise to a moment condition of the form $E \left[\frac{\partial \mathcal{L}_t}{\partial \theta} \right] = 0$, where θ denotes any given parameter being estimated.

Table 2: Validating model choices.

	Non-zero entries	Estimate of non-zero entries	P-val
Spec 1:	$K_V(1,1)$	-27.75***	0.003
Spec 2:	$K_V(1,2)$	7.42	0.319
Spec 3:	$K_V(1,3)$	-20.99**	0.032
Spec 4:	$K_V(1,4)$	4.01	0.623
Spec 5:	$K_V(2,1)$	13.41	0.502
Spec 6:	$K_V(2,2)$	8.49	0.254
Spec 7:	$K_V(2,3)$	10.95	0.370
Spec 8:	$K_V(2,4)$	1.16	0.856
Spec 9:	$K_V(3,1)$	-8.62	0.612
Spec 10:	$K_V(3,2)$	-25.84	0.180
Spec 11:	$K_V(3,3)$	-31.98	0.159
Spec 12:	$K_V(3,4)$	-24.16	0.240
Spec 13:	$K_V(1,1)$	-36.84**	0.018
	$K_V(1,3)$	12.67	0.412

bonds contain more volatility information about expected excess returns than far-maturity bonds do.

Moving beyond the above initial exercise, to find a statistically supported specification, we conduct a more comprehensive search by roaming over different variants of our model, each with a different specification for K_V . Ideally, we would try all combinations of zero constraints on all 12 entries of K_V , giving us $2^{12} = 4096$ different specifications. However, given the computational burden, we consider combinations of rows and columns of K_V and set the corresponding rows and columns to zero. With three rows and four columns, we end up with $2^3 \times 2^4 = 128$ different specifications of K_V . We use the BIC scores to compare across different specifications. Interestingly, specification 1, in which only the (1,1) entry of K_V is set free, emerges as the most preferred candidate.

Based on the evidence provided in this subsection, our main focus for the remaining of the paper will be on specification 1.¹⁶

4.3 Parameter estimates

Table 3 reports the full model estimation for specification 1, together with the p-values. To facilitate a comparison between feedback matrices under \mathbb{P} and \mathbb{Q} , we report the \mathbb{Q} feedback

¹⁶We also check whether our findings regarding the relative contributions of the short-maturity and far-maturity yield portfolios are sensitive to the five year cutoff. Specifically, we repeat all the exercises with a different construction of the yield portfolios. We let the short-maturity (long-maturity) portfolio be an equal-weighted portfolio of three (seven) yields with maturities ranging from one year to three (four to ten) years. All of our results in this subsection remain essentially the same.

Table 3: Parameter estimates of the model, only allowing the (1,1) entry of K_V to be a free parameter

	Estimates			P-val		
K_0	0.02***			0.00		
	-0.02***			0.00		
	-0.08***			0.00		
K_1	1.00***	0.00	0.00*	0.00	0.75	0.05
	0.00***	0.99***	-0.01***	0.00	0.00	0.00
	0.00	-0.00	0.95***	0.14	0.67	0.00
K_V	-27.75***	0.00	0.00	0.00	0.01	
	0.00	0.00	0.00	0.00		
	0.00	0.00	0.00	0.00		
$K_{1\mathcal{P}}^Q$	1.00***	0.01***	0.01***	0.00	0.00	0.00
	-0.00***	0.99***	-0.02***	0.00	0.00	0.00
	-0.00***	-0.01***	0.97***	0.00	0.00	0.00
$\Sigma_{\mathcal{P}}$	0.54			0.77		
	-0.16	0.73		0.78	0.77	
	0.97	-0.06	1.74	0.77	0.83	0.77
ρ_L^{52}	0.99***			0.00		
ρ_S^{52}	0.73***			0.00		
α	0.04***			0.00		
ϕ	0.04***			0.00		
r_∞^Q	0.10***			0.00		
σ_e (bps)	7.22***			0.00		

matrix, $K_{1\mathcal{P}}^Q$, with the PCs of yields \mathcal{P}_t used as state variables.¹⁷ As expected, the diagonal values of the K_1 and $K_{1\mathcal{P}}^Q$ matrices are close to one, suggesting a high persistence of the PCs at the weekly frequency. Notably, all elements of $K_{1\mathcal{P}}^Q$ are statistically significant. The differences between the two feedback matrices reveal the contributions of the PCs to bond risk premiums, which we will examine more closely in the next subsection. For now, we note that the differences between K_1 and $K_{1\mathcal{P}}^Q$ and the statistical significance of the K_V matrix means that both the PCs of bond yields and the short-run volatility component seem to have independent contributions to the dynamics of bond risk premiums.

For ease of interpretation, we report the annualized degrees of persistence, ρ_L^{52} and ρ_S^{52} ,

¹⁷That is, $K_{1\mathcal{P}}^Q$ satisfies: $E_t^Q[\mathcal{P}_{t+1}] = \text{constant} + K_{1\mathcal{P}}^Q \mathcal{P}_t$. $K_{1\mathcal{P}}^Q$ is obtained from the K_1^Q matrix and the yields loadings B_X of the canonical model, using the rotations described in [Appendix A](#).

instead of ρ_L and ρ_S . The long-run volatility component L_t is extremely persistent, with ρ_L^{52} estimated to be 0.99, suggesting a half life of more than 60 years. The short-run volatility component is less persistent, and the estimated ρ_S^{52} suggests that it has a half life of about two years. Given the highly persistent L_t , it is perhaps not surprising that the estimated $\Sigma_{\mathcal{P}}$ is not statistically significant: by (12), the point estimation of $\Sigma_{\mathcal{P}}$ involves dividing by $(1 - \rho_L)$, a very small scalar, which leads to large standard errors. The estimated α and ϕ are both 0.04, so about 8% of the variance shock in each week, $\epsilon_t \epsilon_t' - \Sigma_{t-1}$, enters next week's conditional variance Σ_t .

Further, as is typically the case for three-factor models, our model prices bonds well, with standard deviation of pricing errors estimated to be about seven basis points.

As a robustness check we also implement specifications 3 and 13, with estimates reported in Table 4. We see that the estimates for those two specifications are essentially the same as those for specification 1 (Table 3). For specification 3 (top half of Table 4), a visible difference is that $\rho_S^{52} = 0.79$, suggesting slightly higher persistence of S_t than in specification 1. This is probably because the higher persistence of volatility in the far-maturity yields makes the estimated S_t more persistent through the GARCH-in-mean term $K_V V_t$. For specification 13 (bottom half of Table 4), the estimates are essentially identical to those in Table 3, except for a larger (1,1) entry of K_V , suggesting that the short-run volatility component of far-maturity bond yields does not provide incremental information relative to the short-run volatility component of near-maturity yields. Both specifications imply a similar magnitude of bond pricing errors.

The statistical significance of the short-run volatility component and the insignificance of long-run volatility component map well back to our motivating exercises in Section 2. The volatility process in the ARCH-in-mean equation (2) only collects recent histories of squared residuals; therefore, it is short-run by construction. The GARCH(1,1) volatility process in (3) collects a much longer history of squared residuals; therefore, it is a mixture of short-run and long-run volatility information. By mixing the return-relevant short-run volatility component and the return-irrelevant long-run component, a single-component GARCH volatility process suppresses the predictability of the former. Indeed, recall from the first two columns of Table 1 that ARCH volatilities show strong predictive power for excess returns across all maturities, but GARCH volatilities do so only for relatively short maturities. More generally, the same argument can apply to other term structure models with a single volatility factor: unless this single volatility factor is predominantly short-run, the risk-return trade-off can be contaminated by the long-run component and becomes hard to detect in the data.

4.4 Economic significance

The evidence reveals that there is a significantly positive risk-return relation in bond markets. A natural next question is the economic magnitude of the effect of volatility for bond excess returns. From (16), we can decompose the predictive component of one-week excess return of an n -week bond, $xr_{n,t+1}$, into a \mathcal{P}_t -related component and a V_t -related component. For each

Table 4: Two robustness checks of the model. The top table shows the estimation for specification 3. The bottom table shows the estimation for specification 13.

	Estimates				P-val		
K_0	0.02**				0.01		
	-0.02***				0.00		
	-0.08***				0.00		
K_1	1.00***	0.00	0.00*		0.00	0.44	0.10
	0.00***	0.99***	-0.01***		0.00	0.00	0.00
	0.00	-0.00	0.95***		0.14	0.81	0.00
K_V	0.00	0.00	-20.99**	0.00			0.02
	0.00	0.00	0.00	0.00			
	0.00	0.00	0.00	0.00			
$K_{1\mathcal{P}}^Q$	1.00***	0.01***	0.01***		0.00	0.00	0.00
	-0.00***	0.99***	-0.02***		0.00	0.00	0.00
	-0.00***	-0.01***	0.97***		0.00	0.00	0.00
$\Sigma_{\mathcal{P}}$	0.49				0.66		
	-0.06	0.67			0.74	0.66	
	0.83	-0.12	1.50		0.67	0.72	0.67
ρ_L^{52}	0.99***				0.00		
ρ_S^{52}	0.79***				0.00		
α	0.04***				0.00		
ϕ	0.04**				0.01		
r_∞^Q	0.10***				0.00		
σ_e (bps)	7.05***				0.00		

	Estimates				Pvals		
K_0	0.02***				0.00		
	-0.02***				0.00		
	-0.08***				0.00		
K_1	1.00***	0.00	0.00**		0.00	0.85	0.04
	0.00***	0.99***	-0.01***		0.00	0.00	0.00
	0.00	-0.00	0.95***		0.15	0.72	0.00
K_V	-36.84**	0.00	12.67	0.00	0.03		0.47
	0.00	0.00	0.00	0.00			
	0.00	0.00	0.00	0.00			
$K_{1\mathcal{P}}^Q$	1.00***	0.01***	0.01***		0.00	0.00	0.00
	-0.00***	0.99***	-0.02***		0.00	0.00	0.00
	-0.00***	-0.01***	0.97***		0.00	0.00	0.00
$\Sigma_{\mathcal{P}}$	0.52				0.75		
	-0.18	0.69			0.77	0.75	
	0.95	-0.02	1.67		0.75	0.93	0.75
ρ_L^{52}	0.99***				0.00		
ρ_S^{52}	0.72***				0.00		
α	0.03***				0.00		
ϕ	0.04***				0.00		
r_∞^Q	0.08***				0.00		
σ_e (bps)	7.04***				0.00		

Table 5: Risk premium decomposition.

	1-yr	2-yr	3-yr	4-yr	5-yr	6-yr	7-yr	8-yr	9-yr	10-yr
PC1	0.41	0.27	0.21	0.18	0.15	0.14	0.12	0.11	0.10	0.09
PC2	0.35	0.41	0.46	0.50	0.53	0.53	0.52	0.51	0.49	0.47
PC3	0.09	0.14	0.11	0.06	0.01	0.00	0.01	0.03	0.06	0.10
PC1-3	0.86	0.82	0.78	0.73	0.69	0.67	0.65	0.65	0.65	0.65
V_t -component	0.14	0.24	0.32	0.38	0.42	0.44	0.44	0.43	0.41	0.40

maturity n , we calculate the fraction

$$\frac{Var [(n-1)B_{n-1,p}K_V V_t]}{Var [(nB_{n,p} - (n-1)B_{n-1,p}K_1 - B_{1,p})\mathcal{P}_t - (n-1)B_{n-1,p}K_V V_t]} \quad (23)$$

in sample as a proxy for the contribution of the volatility component V_t for bond excess returns. We calculate the equivalent fraction for the contribution of the PC component. In this calculation, we use the estimates from specification 1, that is, with only the (1,1) entry of K_V is set free and other entries are set to zero.

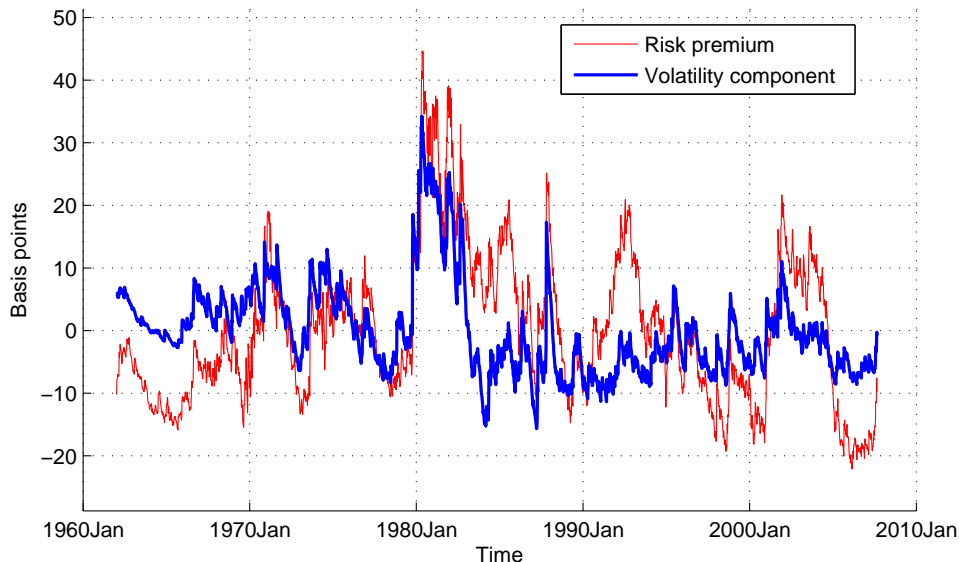
The contributions of the PC-related (V_t -related) components are reported in the first four rows (last row) of [Table 5](#). Each column corresponds to a given bond maturity. The first three rows report the individual contributions of each of the three PCs. It is perhaps not surprising that the slope factor represents the most important contribution to risk premiums. This is consistent with established results in the literature (see, for example, [Fama and Bliss, 1987](#); and [Campbell and Shiller, 1991](#)). The level factor seems most important for relatively short-dated bonds, whereas the curvature factor seems inconsequential across the entire maturity spectrum. Summing up the first three rows gives the overall contributions to risk premiums of all three PCs (because the PCs are uncorrelated).

Because of the correlation between V_t and \mathcal{P}_t , the sum of the last two rows can exceed one, but it is no higher than 1.11 for the 10 maturities, suggesting that the correlation is not a severe concern. In fact, the in-sample correlations between the \mathcal{P}_t -related components and the V_t -related component is very close to zero for each bond maturity. This means that these components represent essentially independent channels through which risk premiums are determined.

[Table 5](#) reveals that the volatility measure, V_t , is an important contributor to expected bond excess returns, with a magnitude comparable to that of yield PCs. About 14% of the predictive component of the excess returns of one-year zero-coupon bond can be attributed to V_t . This fraction increases with maturity, reaching its peak of 44% for the six-year and seven-year bond, and then slowly declines to 40% for the 10-year bond. The contribution by the PC components, by contrast, declines from 86% at one-year maturity to 65% at seven-year maturity, and then stabilizes at 65% for the remaining far end of the yield curve.

[Figure 1](#) plots the demeaned time-series of the total expected excess returns and the volatility component for the 10-year zero-coupon bond. By construction, this volatility

Figure 1: Model-implied (weekly) risk premium (demeaned) on 10-year zero coupon bond. In the model only the (1,1) entry of K_V is estimated, and all other entries of K_V are set to zero.



component is a linear function of the first entry of the V_t vector. It has high time variation and captures a large fraction of expected bond excess returns. [Figure 1](#) also shows a sizeable increase in risk premiums during the Fed experiment regime of the early 1980s. According to our model, much of this increase is attributable to an increase in the short-run volatility component—a result we find reasonable and intuitive.

Our results in this subsection imply that to fully characterize the dynamics of bond risk premiums, a model should allow for two channels: a time-varying market price of risk and a time-varying quantity of risk. Moreover, the market prices of risks must represent an independent source of time variation and cannot be subsumed by the quantity of risk. For example, the habit-based term structure model of [Le, Singleton, and Dai \(2010\)](#) allows for both channels but the time-varying market prices of risks depend exclusively on yield volatility. As a result, risk premiums implied by their model are solely determined by yield volatility. Based on our findings, their model is likely to miss a sizeable portion of time variation in bond risk premiums. Our results corroborate the findings of [Le and Singleton \(2013\)](#) in their analysis of structural term structure models.

Although it can be tempting to interpret the last row of [Table 5](#) as the contribution by the quantity of risk, and the second last row as the contribution by the market price of risk, such an interpretation may not be entirely accurate. The reason is that, in principle, the market prices of risk can also depend on yield volatility.

4.5 Matching return predictability and conditional volatilities in the data

Having examined the statistical and economic significance of the conditional volatility V_t for bond excess returns, we now take a step back and examine how well the model fits salient empirical patterns of return predictability and time-varying volatility in bond markets. We emphasize that a key contribution of our model is to well match both aspects of the data.

Figure 2 shows the Campbell and Shiller (1987) regression coefficients implied by the model, together with those implied by the data. As is well known, if the expectations hypothesis holds and thus risk premiums are not time-varying, the coefficients of this regression should be uniformly ones across all maturities. Instead, the coefficients obtained from the data are significantly negative and increasingly so as maturities increase. Our model does a relatively decent job in capturing this pattern of the data, arguably as well as the Gaussian affine term structure models do (see Dai and Singleton, 2002). A weakness of the Gaussian models is the constant-volatility assumption; thus, those models cannot match the conditional volatilities of yields.

Figure 3 plots the model-implied one-week ahead conditional volatilities of the first three PCs of the yield term structure, as well as those of the 1-year, 5-year, and 10-year yields. As a comparison, we also plot realized volatilities and conditional volatilities estimated from a univariate EGARCH model. At each point in time, the realized volatilities are computed using daily changes in yields over the preceding three months. The EGARCH model is implemented using weekly yields over the entire sample. Volatilities implied by our model are very close to those two commonly used volatility measures.

Figure 2: Campbell-Shiller regression coefficients in data and implied by model. In the model only the (1,1) entry of K_V is estimated, and all other entries of K_V are set to zero.

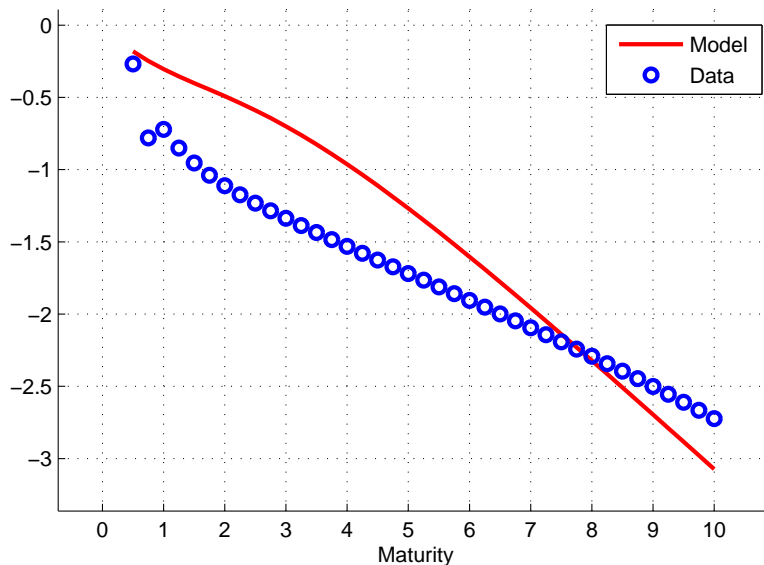
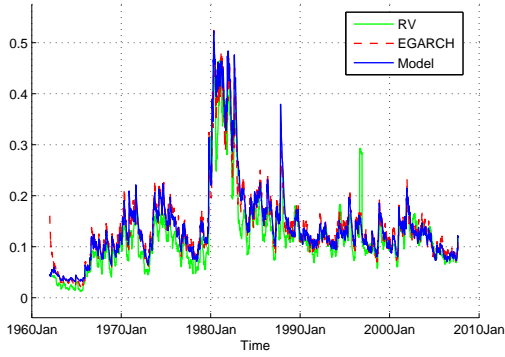
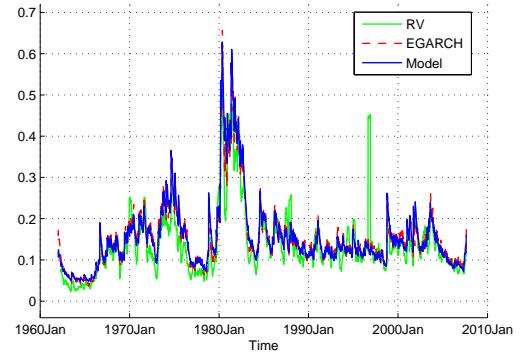


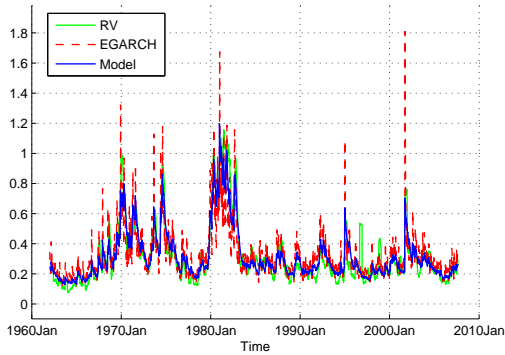
Figure 3: Model-implied volatilities and realized volatilities. In the model only the (1,1) entry of K_V is estimated, and all other entries of K_V are set to zero.



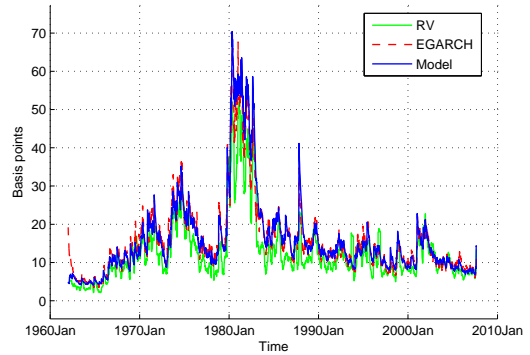
(a) PC1



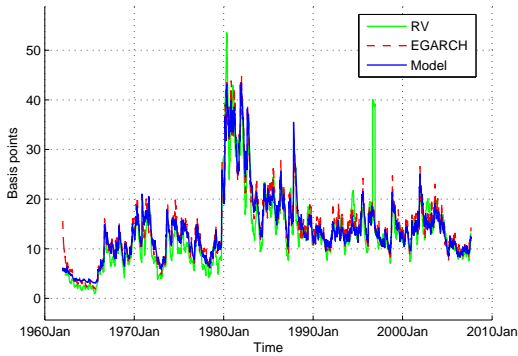
(b) PC2



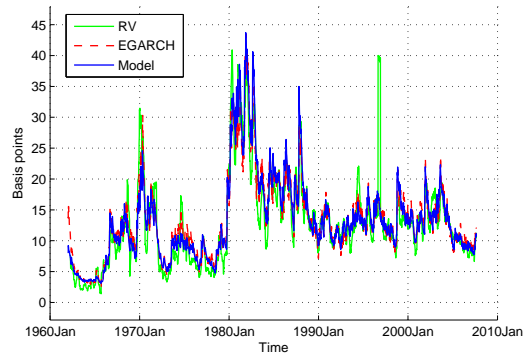
(c) PC3



(d) 1-year yield



(e) 5-year yield



(f) 10-year yield

5 Conclusion

In this paper, we study the risk-return tradeoff in the U.S. Treasury markets through the lens of a new discrete-time no-arbitrage term structure model. The model combines the tractability of affine term structure models with the ability of GARCH models to deliver an accurate measure of yield volatility. Not only does this model fit yields and yield volatilities well across all maturities, it also closely replicates the returns predictability characterized by the [Campbell and Shiller \(1987\)](#) regressions. Moreover, this model also allows us to differentiate the contributions to risk premiums of long-run and short-run volatility components of both the short end and long end of the yield curve.

Using yields data from 1962 to 2007, we find a significantly positive relation between risk and return in U.S. Treasury markets. A higher conditional volatility this week predicts a higher expected excess return next week. Notably, it is the short-run component of volatility, not the long-run component, that matters for return predictability. Moreover, the return-predicting power of the short-run volatility component predominantly comes from the short-end of the yield curve. Volatilities of far-maturity yields do not have additional predictive power for future yields once we control for volatility of the short end. Volatilities have economically important effects on bond risk premiums. For example, the volatility factor accounts for 14%, 42%, and 40% of the predictable components of weekly excess returns on one-year, five-year, and ten-year zero coupon bonds, respectively. The other source of predictability comes from the principal components of yields.

Our results have important implications. We show that the volatility factor, a proxy for quantity of risk, and the principal components of yields, a proxy for the market price of risk, have comparable weights in determining bond risk premiums. Therefore, models that rule out either channel only provide an incomplete characterization of the risk premium dynamics. Furthermore, because the short-run volatility component, not the long-run component, is responsible for predicting returns, the risk-return trade-off may be missed by models with a single volatility factor.

A Appendix: Rotation of Factors

This appendix describes the detailed steps of rotating one set of factors into another.

A.1 Summary of the model

As explained in the text, our model is characterized by the following risk-neutral (\mathbb{Q}) and time-series (\mathbb{P}) dynamics of the latent factors X_t :

$$r_t = \delta_0 + \delta_1' X_t, \quad (24)$$

$$X_{t+1} = K_0^{\mathbb{Q}} + K_1^{\mathbb{Q}} X_t + \epsilon_{t+1}^{\mathbb{Q}}, \quad (25)$$

$$X_{t+1} = K_0 + K_1 X_t + K_V V_t + \epsilon_{t+1}, \quad (26)$$

where $\epsilon_{t+1}^{\mathbb{Q}} \sim N(0, \Sigma_X)$ and $\epsilon_{t+1} \sim N(0, \Sigma_t)$, with Σ_t following the component model of [Engle and Lee \(1999\)](#):

$$\Sigma_t = S_t + L_t, \quad (27)$$

$$S_t = \rho_S S_{t-1} + \alpha(\epsilon_t \epsilon_t' - \Sigma_{t-1}), \quad (28)$$

$$L_t = (1 - \rho_L) \Sigma_X + \rho_L L_{t-1} + \phi(\epsilon_t \epsilon_t' - \Sigma_{t-1}). \quad (29)$$

Combining (24) and (25), we see that yields at all maturities are affine in the state variables:

$$y_{n,t} = A_{n,X} + B_{n,X} X_t, \quad (30)$$

with $A_{n,X}$ and $B_{n,X}$ computed from the standard recursive equations for bond pricing.

The volatility vector V_t is given by:

$$V_t = \begin{pmatrix} \sqrt{B_{1-5,X} \Sigma_t B_{1-5,X}' - B_{1-5,X} L_t B_{1-5,X}'} \\ \sqrt{B_{1-5,X} L_t B_{1-5,X}'} \\ \sqrt{B_{6-10,X} \Sigma_t B_{6-10,X}' - B_{6-10,X} L_t B_{6-10,X}'} \\ \sqrt{B_{6-10,X} L_t B_{6-10,X}'} \end{pmatrix}, \quad (31)$$

where $B_{1-5,X}$ and $B_{6-10,X}$ denote the weighting vectors for the equal-weighted yield portfolios with near maturities (1–5 years) and far maturities (6–10 years).

Our model is fully characterized by the following parameter set

$$\Theta_X = (\delta_0, \delta_1, K_0^{\mathbb{Q}}, K_1^{\mathbb{Q}}, \Sigma_X, K_0, K_1, K_V, \rho_S, \alpha, \rho_L, \phi).$$

A.2 Rotation of factors

Consider a rotation from factors X_t to $Z_t = U_0 + U_1 X_t$ for any given pair (U_0, U_1) . For example, the new factors Z_t could be the yield PCs. It is straightforward to see that both the risk-neutral and time-series dynamics will be of the same affine form:

$$r_t = \tilde{\delta}_0 + \tilde{\delta}'_1 Z_t, \quad (32)$$

$$Z_{t+1} = \tilde{K}_0^{\mathbb{Q}} + \tilde{K}_1^{\mathbb{Q}} Z_t + \tilde{\epsilon}_{t+1}^{\mathbb{Q}}, \quad (33)$$

$$Z_{t+1} = \tilde{K}_0 + \tilde{K}_1 Z_t + \tilde{K}_V \tilde{V}_t + \tilde{\epsilon}_{t+1}, \quad (34)$$

where $\tilde{\epsilon}_{t+1}^{\mathbb{Q}} \sim N(0, \tilde{\Sigma}_X)$ and $\tilde{\epsilon}_{t+1} \sim N(0, \tilde{\Sigma}_t)$. The mappings from the original model to the rotated model are as follows:

$$\tilde{\delta}_0 = \delta_0 - \delta'_1 U_1^{-1} U_0, \quad (35)$$

$$\tilde{\delta}'_1 = \delta'_1 U_1^{-1}, \quad (36)$$

$$\tilde{K}_0^{\mathbb{Q}} = U_0 + U_1 K_0^{\mathbb{Q}} - U_1 K_1^{\mathbb{Q}} U_1^{-1} U_0, \quad (37)$$

$$\tilde{K}_1^{\mathbb{Q}} = U_1 K_1^{\mathbb{Q}} U_1^{-1}, \quad (38)$$

$$\tilde{\Sigma}_X = U_1 \Sigma_X U_1', \quad (39)$$

$$\tilde{K}_0 = U_0 + U_1 K_0 - U_1 K_1 U_1^{-1} U_0, \quad (40)$$

$$\tilde{K}_1 = U_1 K_1 U_1^{-1}, \quad (41)$$

$$\tilde{K}_V = U_1 K_V. \quad (42)$$

Additionally, the conditional covariance matrix under \mathbb{P} , as well as its long-run and short-run counterparts, are simply given by:

$$\tilde{\Sigma}_t = U_1 \Sigma_t U_1', \quad \tilde{L}_t = U_1 L_t U_1', \quad \text{and} \quad \tilde{S}_t = U_1 S_t U_1'. \quad (43)$$

Combining these with the dynamics in (27), (28), and (29), we see that the parameters $(\rho_S, \rho_L, \alpha, \phi)$ governing the volatility dynamics are invariant to rotations.

To calculate the yield loadings, we observe that $y_{n,t} = A_{n,X} + B_{n,X} X_t = A_{n,X} + B_{n,X} U_1^{-1} (Z_t - U_0)$. Thus, the loadings with respect to the new state variable Z are given by:

$$A_{n,Z} = A_{n,X} - B_{n,X} U_1^{-1} U_0 \quad \text{and} \quad B_{n,Z} = B_{n,X} U_1^{-1}. \quad (44)$$

Finally, note that V_t is invariant to rotations in that $\tilde{V}_t \equiv V_t$. Intuitively, this is due to the fact that V_t is measured by the volatilities of observable yields portfolios. More concretely, take one of the terms in the construction of \tilde{V}_t , say $B_{1-5,Z} \tilde{\Sigma}_t B'_{1-5,Z}$, we see that:

$$B_{1-5,Z} \tilde{\Sigma}_t B'_{1-5,Z} = B_{1-5,X} U_1^{-1} U_1 \Sigma_t U_1' U_1^{-1} B'_{1-5,X} = B_{1-5,X} \Sigma_t B'_{1-5,X}.$$

References

- Adrian, T., and J. Rosenberg, 2008, "Stock Returns and Volatility: Pricing the Short-Run and Long-Run Components of Market Risk," *The Journal of Finance*, 63, 2997–3030.
- Andersen, T., and L. Benzoni, 2010, "Do Bonds Span Volatility Risk in the U.S. Treasury Market? A Specification Test for Affine Term Structure Models," *Journal of Finance*, 65, 603–653.
- Bollerslev, T., 1986, "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.
- Campbell, J., and R. Shiller, 1987, "Cointegration and Tests of Present Value Models," *Journal of Political Economy*, 95, 1062–1087.
- Campbell, J., and R. Shiller, 1991, "Yield Spreads and Interest Rate Movements: A Bird's Eye View," *Review of Economic Studies*, 58, 495–514.
- Cieslak, A., and P. Povala, 2013, "Information in the term structure of yield curve volatility," Discussion paper, Northwestern University and University of London, Birkbeck.
- Collin-Dufresne, P., R. Goldstein, and C. Jones, 2009, "Can Interest Rate Volatility Be Extracted From the Cross Section of Bond Yields?," *Journal of Financial Economics*, 94, 47–66.
- 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.
- Engle, R., and G. Lee, 1999, "A Long-Run and Short-Run Component Model of Stock Return Volatility," in R. Engle, and H. White (ed.), *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W. J. Granger* . pp. 475 – 497, Oxford University Press.
- Engle, R., D. Lilien, and R. Robins, 1987, "Estimating Time Varying Risk Premia in the Term Structure: the ARCH-M Model," *Econometrica*, 55, 391–407.
- Fama, E. F., and R. R. Bliss, 1987, "The Information in Long-Maturity Forward Rates," *American Economic Review*, 77, 680–692.
- Ghysels, E., 2013, "Matlab Toolbox for Mixed Sampling Frequency Data Analysis using MIDAS Regression Models," Discussion paper, UNC.

- Ghysels, E., A. Plazzi, and R. Valkanov, 2013, “The Risk-Return Relationship and Financial Crises,” Discussion Paper UCSD, University of Lugano and UNC.
- Ghysels, E., A. Sinko, and R. Valkanov, 2007, “MIDAS regressions: Further results and new directions,” *Econometric Reviews*, 26, 53–90.
- Gurkanyak, R., B. Sack, and J. Wright, 2007, “The U.S. Treasury Yield Curve: 1961 to the Present,” *Journal of Monetary Economics*, 54, 2291–2304.
- Haubrich, J., G. Pennacchi, and P. Ritchken, 2012, “Inflation Expectations, Real Rates, and Risk Premia: Evidence from Inflation Swaps,” *Review of Financial Studies*, 25, 1588–1629.
- Joslin, S., 2007, “Pricing and Hedging Volatility in Fixed Income Markets,” Discussion paper, Working Paper, MIT.
- Joslin, S., and A. Le, 2012, “Interest Rate Volatility and No-Arbitrage Term Structure Models,” Discussion paper, USC and UNC.
- 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., K. Singleton, and H. Zhu, 2011, “A New Perspective on Gaussian Dynamic Term Structure Models,” *Review of Financial Studies*, 24, 926–970.
- Le, A., and K. Singleton, 2013, “The Structure of Risks in Equilibrium Affine Model of Bond Yields,” Discussion paper, University of North Carolina.
- Le, A., K. Singleton, and J. Dai, 2010, “Discrete-Time Affine^Q Term Structure Models with Generalized Market Prices of Risk,” *Review of Financial Studies*, 23, 2184–2227.
- Nelson, D., 1991, “Conditional Heteroskedasticity in Asset Returns: A New Approach,” *Econometrica*, 59, 347–370.
- Newey, W., and K. D. West, 1987, “A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 55, 703–708.
- Stock, J., and M. Watson, 2007, “Why Has U.S. Inflation Become Harder to Forecast?,” *Journal of Money Credit and Banking*, 39, 3–33.
- Trolle, A., and E. Schwartz, 2009, “A General Stochastic Volatility Model for the Pricing of Interest Rate Derivatives,” *Review of Financial Studies*, 22, 2007–2057.