Information in (and not in) Treasury Options

Hoyong Choi
Rotterdam School of Management

Contact

Address:	Department of Finance
	Rotterdam School of Management
	Burgemeester Oudlaan 50, Room T8-21
	P.O. Box 1738
	3000 DR Rotterdam, The Netherlands
Phone:	+31 (0)10 408 1495
Mail:	choi@rsm.nl

Information in (and not in) Treasury Options

Abstract

This paper studies the impact of variance risk in the Treasury market on both term premia and the shape of the yield curve. Under minimal assumptions shared by standard structural and reduced-form asset pricing models, I show that an observable proxy of variance risk in the Treasury market can be constructed via a portfolio of Treasury options. The observable variance risk has the ability to explain the time variation in term premia, but is largely unrelated to the shape of the yield curve. Using the observable variance risk, I also propose a new representation of no-arbitrage term structure models. All the pricing factors in the model are observable, tradable, and hence economically interpretable. The representation can also accommodate both unspanned macro risks and unspanned stochastic volatility in the term structure literature.

1 Introduction

What is the role of variance risk in the Treasury market? How big is its impact on the risk-return trade-off? How does it affect the shape of the yield curve? What kind of macroeconomic uncertainty drives it? The first step to addressing these questions is to *identify* the variance risk in the Treasury market. In this paper, I suggest a novel approach to identifying variance risk, by utilizing information in Treasury bond options to answer the above questions.

I first show that variance risk can be proxied by implied variance measures from bond option markets and that this is true under a set of mild assumptions which are shared by many well-known structural and reduced form asset pricing models. Specifically, I prove that a bond VIX^2 (a portfolio of Treasury options constructed akin to the VIX^2 in the equity market[1]) represents the variance risk in the Treasury market under the assumptions that (i) the short-term interest rate is a linear function of the state variables and (ii) the state follows an affine diffusion process under the risk-neutral measure. In other words, the bond VIX^{2}s span time-varying variances in Treasury yields under the two assumptions. As a consequence, the impact of variance risk on both term premia and the shape of the yield curve is directly measurable via the *observable* variance risk: the bond VIX^{2}s. Using this theoretical framework, I obtain the following three novel results.

First, I propose a novel return-forecasting factor that *jointly* exploits the bond VIX^{2}s and the implication of leading macro-finance asset pricing models. The bond VIX^{2}s identify economic fundamentals that determine the conditional variances of bond yields in many well-known consumption-based asset pricing models: for example, the time-varying variance of consumption growth in Bansal and Yaron (2004), the probability of a rare disaster in Wachter (2013), and the external habit in Le, Singleton, and Dai (2010). Interestingly, the unobservable fundamentals are both the drivers of the variances of Treasury yields and the sole sources of time variation in term premia under these frameworks (see e.g. Le and Singleton (2013)). Hence, one common implication of the models is that excess returns on bonds should be completely explained by the bond VIX^{2}s. In particular, in the long-

[1]The VIX is a measure of volatility, and hence the VIX^2 is a measure of variance.

run risk framework of Bansal and Yaron (2004), the risk premium is time-varying solely due to the time variation in the quantity of risk. Moreover, changes in the variance of yields are the manifestation of time-varying macroeconomic uncertainties in the long-run risk framework. Because the short-term interest rate is postulated to be linear in affine diffusion states in the economy, the bond VIX^2s span the time-varying variance in yields. Hence, the time variation in expected excess returns should be captured by the bond VIX^2s. The same implication can also be obtained from the rare disaster framework of Wachter (2013). Time-varying probability of a rare disaster is assumed to follow an affine diffusion process in the framework, and it is also a sole driver of time variation in both risk premia and interest rate variance. Hence, the bond VIX^2s are the manifestation of time-varying disaster probabilities, and should have the ability to predict future excess returns. The affine$^{\mathbb{Q}}$ habit model in Le, Singleton, and Dai (2010) is another class of models in which the bond VIX^2s should be driven by the factor underlying the time variation in risk premia. In this framework, the external habit of the representative agent - the source of time-varying price of risks - is the only factor driving both the time variation in the variance of yields and the time-varying risk premia. Moreover, the drift in the pricing kernel is assumed to be linear in the state variables following an affine-diffusion process under the risk-neutral measure, and hence the bond VIX^2s reflect the time-varying price of risk. In sum, the space of time-varying risk premia and the space of the bond VIX^2s are identical under the standard macro-finance asset pricing models, the long-run risk, rare disaster, and affine$^{\mathbb{Q}}$ habit formation frameworks.

In the above three frameworks, the noise in realized excess returns on bonds can be completely removed by projecting realized excess returns onto the bond VIX^2s space. Hence, using implied variance measures from options on Treasury futures with different tenors, I apply a projection in line with Cochrane and Piazzesi (2005) to obtain a single return-forecasting factor. Similar to their regressions where they project bond excess returns of different maturity onto forward rates, I find that the linear combination of bond VIX^2s also produces a tent-shape factor forecasting excess returns. Interestingly, the predictive ability of this return-forecasting factor mainly stems from the excess returns of relatively short-term bonds, while the linear combination of forward rates in Cochrane and Piazzesi

(2005) is superior in predicting excess returns on long-term bonds. Moreover, the single factor from the bond VIX^2s and the Cochrane-Piazzesi factor are complementary, and the predictability for bond returns increases significantly in joint regressions.

Second, I analyze the observable variance risk's impact on the shape of the yield curve. Its marginal impact is assessed by projecting yields onto the bond VIX^2s as well as the first three yield principal components: level, slope and curvature. After controlling for these factors, I find that variance risk is largely unrelated to the shape of the yield curve. This result corroborates earlier evidence of unspanned stochastic volatility (USV) whereby yield variance can only be very weakly identified from the cross-section of yields (see e.g., Collin-Dufresne and Goldstein (2002) among many others). However, the strict condition for the USV effect is rejected by a newly devised statistical test exploiting the observational variance risk. In sum, it is hard to identify the volatility of interest rates from the yield curve movements, but the knife-edge conditions for the USV effect do not seem to hold in the data.

Third, to assess the variance risk's impact on term premia and the cross-section of yields within a fully-fledged framework, I suggest a new representation of affine no-arbitrage term structure models that incorporate the observable variance risk. The representation follows in the spirit of Joslin, Singleton, and Zhu (2011), and extends their work to affine models with stochastic volatility. The risk factors are represented as a portfolio of yields and options. Hence, all the pricing factors are observable, tradable, and economically interpretable. In addition, due to the observable variance risk, the factor dynamics under the physical measure can easily be estimated by generalized least squares. Furthermore, the observable proxy of variance incorporates information in volatility-sensitive instruments, namely the Treasury options. As a result, the variance risk in interest rates is well identified, in contrast to the conventional latent factor approaches. Finally, the model can be easily extended to reflect the unspanned macro risks in Joslin, Priebsch, and Singleton (2014) (henceforth JPS). Given that the observable variance factor can be unspanned by yields, the model can accommodate the two distinct types of unspanned risks in the term structure literature: unspanned macro risk factor (hidden factor) and unspanned stochastic volatility. The estimates of the model indicate that both unspanned macro risk and stochastic volatility drive expected returns.

The stochastic volatility factor in the estimated model is not literally unspanned by yields, but its impact on the shape of the yield curve is noticeably small and can be effectively treated as an unspanned factor.

This paper also contributes to the recent discussion on unspanned macro risks in the macro-finance term structure literature. The unspanned macro risks are macroeconomic factors that are informative about macroeconomic fluctuations and term premia, but largely unrelated to the term structure movements. One open question with this strand of studies[2] is, among the hundreds of macroeconomic variables, which one *should* or *could* be treated as an unspanned macro risk? For example, Bauer and Rudebusch (2016) show that estimates of risk premia can differ significantly depending on whether a measure of the *level* or the *growth* in economic activity is used as unspanned risk. I show that the LPY ("linear projection of yields") criteria in Dai and Singleton (2002) provide informative guidance on this issue. The LPY criteria are descriptive statistics that measure whether a term structure model can match the pattern of violation of the expectations hypothesis as in Fama and Bliss (1987) or Cambpell and Shiller (1991). For the issue of choosing level or growth indicators of economic activity as an unspanned macro risk, the LPY criteria indicate that level variable is more relevant measure of economic activity in term structure modeling perspective. In other words, the models with level of economic activity as an unspanned macro risk are better at re-producing the pattern for the failure of expectations hypothesis in the data than the models with growth indicator as unspanned macro risk. Furthermore, in the LPY dimension, the stochastic volatility models with/without unspanned macro risks outperform the corresponding Gaussian models. This shows that the observational variance risk is (i) properly identified and (ii) beneficial in explaining the time variation in risk premia.

This paper is related to several different strands of the literature. First, the construction of an observable proxy of variance risk in interest rates is based on the methodology of Mele and Obayashi (2013) and Choi, Mueller, and Vedolin (2016). However, while these papers utilize bond VIX^{2}s to study the price of variance risk or variance risk premium in a model-free manner, this paper (i) initially identifies the classes of asset pricing models under which the bond VIX^2 is equivalent to the interest rate variance risk of the models, (ii) and then

[2]See e.g., Duffee (2011b), Chernov and Mueller (2012) and Joslin, Priebsch, and Singleton (2014).

jointly utilizes both the bond VIX^2s and the implication of the asset pricing models for a better understanding of expected excess returns on long-term bonds (rather than variance trading). In other words, given an asset pricing model within the class characterized by (i) affine short rate and (ii) affine state under the risk-neutral measure, variance risk takes the form of bond VIX^2, and this observable portfolio of options inherits all the properties and implications of the variance risk in the model. In this paper, the bond VIX^2s are utilized as instruments to identify such variance risks *within* the models. For structural asset pricing models with the two assumptions, the bond VIX^2s identify economic fundamentals that drive variance risks in the Treasury market. Hence, the bond VIX^2s should inherit all the asset pricing implications of the fundamentals.

This idea implies that within the long-run risk, rare disaster, and affine$^{\mathbb{Q}}$ habit formation frameworks, the bond VIX^2s should predict excess returns on bonds because the set of risk factors underlying variation in risk premia is the sole source of time-varying variances in bond yields. Hence, the return-forecasting factor in this paper is based on the theoretical prediction of those specific models, contrary to the return-forecasting factors from the yield curve as in Fama and Bliss (1987), Cambpell and Shiller (1991), and Cochrane and Piazzesi (2005). Furthermore, the bond VIX^2s can be measured in real time and contain forward-looking information, in contrast with infrequently-updated macro data as in Bansal and Shaliastovich (2013) or Ludvigson and Ng (2009).

The benefit of observable variance is also highlighted in the connection of the bond VIX^2 to the no-arbitrage affine dynamic term structure models (henceforth, ADTSM). Under the assumption of ADTSM, the bond VIX^2 directly identifies variance risk in ADTSM, which has been considered one of the most challenging tasks in the term structure literature. While previous term structure models also incorporate information from volatility-sensitive instruments into their estimation procedure for better identification of variance risk[3], the approach of this paper circumvents their computational difficulties. Specifically, previous studies match *individual* derivative prices from the models to actual derivative prices in their estimation procedures, but the calculations of the derivative prices are extremely cum-

[3]See e.g. Jagannathan, Kaplin, and Sun (2003), Bibkov and Chernov (2009), Trolle and Schwartz (2009), Bibkov and Chernov (2011), Almeida, Graveline, and Joslin (2011) and Joslin (2014) among many others.

bersome computationally. By formulating a specific option *portfolio* that directly reflects the changes in the underlying variance factor, the approach I propose simplifies the incorporation of information in volatility-sensitive instruments into ADTSM.

Furthermore, the observable variance risk enables ADTSM to be represented by observable and tradable factors, contrary to all the previous dynamic term structure models with stochastic volatility. Hence, the new representation of ADTSM that I posit here is based on the observable variance risk, and extends the representation for both spanned Gaussian ADTSM in Joslin, Singleton, and Zhu (2011) and Gaussian ADTSM with unspanned macro risk in Joslin, Priebsch, and Singleton (2014) into more general setting. Joslin and Le (2014) also utilize a parameterization scheme for ADTSM with stochastic volatility, in which the time-varying variance factor is approximated by observable portfolio of yields. Their volatility instrument can only be identified after the estimation of the model, while the bond VIX^2 identifies variance risk even before the estimation of ADTSM. Furthermore, the approach of this paper is robust to unspanned stochastic volatility (USV), because option prices are utilized to detect variance risk. On the other hand, yields do not span variance risk in the presence of USV, and hence one cannot construct a yield portfolio that captures time-varying variance as in Joslin and Le (2014).

Finally, while all the other USV models in the literature should be estimated with hardwired constraints to generate USV effects[4], the approach here does not impose a priori constraints for the USV effect and lets the data speak about the presence of USV. With the bond VIX^2 at hand, the estimation of ADTSM reveals the relative importance of the already-identified variance factor in determining the shape of the yield curve. Once the bond VIX^2 turns out to play little role in explaining the cross-section of yields, then one can effectively treat it as an unspanned stochastic volatility factor.

The paper proceeds as follows. Section 2 theoretically shows how one can construct an observable proxy of the variance risk in the Treasury market by utilizing information in option markets. Section 3 argues why the observable measure of variance could capture the time-variation in risk premia, and investigates its predictive ability for excess returns.

[4]See e.g. Bibkov and Chernov (2009), Collin-Dufresne, Goldstein, and Jones (2009), Trolle and Schwartz (2009), Joslin (2015) and Creal and Wu (2015) among many others.

Section 4 analyzes the relation between the variance risk and the shape of the yield curves. Section 5 introduces a new representation for no-arbitrage term structure models in which the variance risk is identified as a portfolio of Treasury options. The representation is extended to accommodate unspanned macro risks in Section 6. In Section 7, the models in Section 6 but with different types of unspanned macro risks are evaluated based on the LPY criteria. Section 8 explores the properties of risk premia in more depth. Finally, Section 9 concludes. All proofs are deferred to the Appendices.

2 Observable Volatility

To start, let us assume the state variable $Z_t = (X_t', V_t')' \in \mathbb{R}^{N-m} \times \mathbb{R}_+^m$ follows the Ito diffusion under the risk-neutral measure $\mathbb{Q}$

$$d \begin{bmatrix} X_t \\ V_t \end{bmatrix} = \mu_{Z,t} dt + \Sigma_{Z,t} dB_t^{\mathbb{Q}} \tag{1}$$

where

$$\mu_{Z,t} = \begin{bmatrix} \mu_{X,t} \\ \mu_{V,t} \end{bmatrix} = \begin{bmatrix} K_{0X} \\ K_{0V} \end{bmatrix} + \begin{bmatrix} K_{1X} & K_{1XV} \\ K_{1VX} & K_{1V} \end{bmatrix} \begin{bmatrix} X_t \\ V_t \end{bmatrix}, \text{ and } \Sigma_{Z,t}\Sigma_{Z,t}' = \Sigma_{Z0} + \sum_{i=1}^{m} \Sigma_{Zi} V_{it}$$

with a set of restrictions on the parameters to ensure the non-negativity of the volatility factor V_t as in Duffie, Filipović, and Schachermayer (2003). $B_t^{\mathbb{Q}}$ is a N-dimensional Brownian motion under $\mathbb{Q}$. The short rate (the negative of the drift in a pricing kernel) is assumed to be linear in the state Z_t

$$r_t = \delta_0 + \delta_1 Z_t \tag{2}$$

In addition, denote the following portfolios of options as $\mathcal{V}_t$ which is a measure of model-free implied variance akin to the Chicago Board Options Exchange (CBOE) VIX^2 in equity markets:

$$\mathcal{V}_t(T, \mathbb{T}) = \frac{2}{P_{t,T}} \left[\int_0^{F_t(T,\mathbb{T})} \frac{\text{Put}_t(K, T, \mathbb{T})}{K^2} dK + \int_{F_t(T,\mathbb{T})}^{\infty} \frac{\text{Call}_t(K, T, \mathbb{T})}{K^2} dK \right] \tag{3}$$

where $P_{t,T}$ is the price of a zero-coupons bond expiring at T, and $F_t(T, \mathbb{T})$ is the forward price at t, for delivery at T, of the bond maturing at $\mathbb{T}$. $\text{Put}_t(K, T, \mathbb{T})$ and $\text{Call}_t(K, T, \mathbb{T})$

are European options with strike price K and tenor T written on $P_{t,\mathbb{T}}$. It is well-known in the equity literature that cross-sectional information from options enables us to recover the risk-neutral probability density of underlying asset (Breeden and Litzenberger (1978)). The CBOE VIX is a specific application of this theory, to proxy the forward-looking risk-neutral volatility of the one-month return on S&P 500 index. Similarly, with T being equal to one-month, $\sqrt{\mathcal{V}_t\left(T,\mathbb{T}\right)}$ can be considered as a forward-looking measure of one-month volatility in $P_{t,\mathbb{T}}$ under the risk-neutral measure.

Under the two assumptions that the state is an affine process as in (1) and that the short rate is affine in the state Z_t as like (2), it can be shown that $F_t\left(T,\mathbb{T}\right)$ follows a diffusion process of which instantaneous variance is a linear function of the latent factor V_t. When $F_t\left(T,\mathbb{T}\right)$ follows a diffusion process, it is well-known that equation (3) represents the expected quadratic variation of the forward under $\mathbb{Q}_T$ measure of which numéraire is the bond $P_{t,T}$ (see, e.g., Carr and Madan (1998)). Furthermore, the change of measure between the forward measure $\mathbb{Q}_T$ and the risk-neutral measure $\mathbb{Q}$ is determined by the volatility of $F_t\left(T,\mathbb{T}\right)$ in a linear fashion: see for example Björk (2009). As a consequence, $\mathcal{V}_t$ can be expressed as a linear function of V_t, which means that one can observe the latent variance factor up to its linear transformation and its shocks via the option portfolio $\mathcal{V}_t$.

Proposition 1. *Suppose that the short rate is an affine function of the $\mathbb{Q}$ affine process in (1). Then,*

$$\mathcal{V}_t\left(T,\mathbb{T}\right)=\alpha\left(\Theta^{\mathbb{Q}};t,T,\mathbb{T}\right)+\beta\left(\Theta^{\mathbb{Q}};t,T,\mathbb{T}\right)\cdot V_t \tag{4}$$

where $\Theta^{\mathbb{Q}}$ is the set of parameters for (1) and (2).

Proof: See Appendix A.

One of the key features of Proposition 1 is that it does not require any specification of the market price of risk (or the dynamics of Z_t under $\mathbb{P}$) to completely characterize a pricing kernel. In other words, the proposition can be utilized even though Z_t follows a non-linear process under $\mathbb{P}$. In sum, for large classes of asset pricing models, one can capture the innovations in variance factor through the portfolio of options, $\mathcal{V}_t$.

As can be seen from equation (3), the option portfolio $\sqrt{\mathcal{V}_t}$ is a Treasury market version of the CBOE VIX in the equity market. While the VIX has been intensively studied and utilized in the literature,[5] studies about its analogue for US Treasuries (henceforth, the bond VIX) started relatively recently. Mele and Obayashi (2013) develop theories on pricing Treasury volatility (i.e. expected value of Treasury volatility under a forward measure), and suggest a practical way of representing the price as a portfolio of Treasury futures options. Based on their methodology, CBOE launched the 10-year U.S. Treasury Note Volatility Index (TYVIX) in May 2013. Choi, Mueller, and Vedolin (2016) show how investors can make use of the bond VIX to get pure exposure to variance risk in the fixed income market and document the empirical properties of the trading strategy. They construct the bond VIX named as Treasury Implied Volatility index (TIV) for a 10-year T-note, plus TIV for a 5-year Treasury bill and a 30-year Treasury bond.

This paper utilizes their TIVs since the three measures of volatility with different underlying bonds enable us to identify multiple latent volatility factors via Proposition 1. For a detailed description of how to construct TIV, I refer the reader to Choi, Mueller, and Vedolin (2016). Figure 1 provides a plot of the CBOE VIX, the CBOE TYVIX, and the 10-year TIV; following the custom in practice, they are the square root of the annualized variances expressed in percent. The 10-year TIV is virtually identical to the TYVIX, and they are largely correlated with the VIX. The bond VIXs are driven by the variance factor in the discount rates (or the pricing kernel), while the VIX reflect the variance factor in both the discount rates and cash flow dynamics. The figure shows that the impact of the variance factor in the cash flow dynamics became less important from late 90s.

[Insert Figure 1 here.]

To summarize, Proposition 1 gives the implication of the model-free measure of implied volatility in the Treasury market, the bond VIX, once it is combined with additional structures embedded in many economic models. Once the information in bond VIX is incorporated with an economic model in which the short rate is linear in $\mathbb{Q}$ affine diffusion state variables, the bond VIXs can completely identify the volatility factors. This result

[5]See, e.g., Carr and Wu (2009), Drechsler and Yaron (2011), Bollerslev, Tauchen, and Zhou (2009), Ang, Hodrick, Xing, and Zhang (2006), Adrian and Shin (2010), Nagel (2012), Brunnermeier, Nagel, and Pedersen (2009), Bao, Pan, and Wang (2011), Amengual and Xiu (2014),Bekaert, Hoerova, and Duca (2013), Kelly, Pástor, and Veronesi (2016) among many others.

also implies that some economic fundamentals in macro-finance asset pricing models can be identified via the bond VIX^2s if the fundamentals determine the conditional variances of bond yields.

3 Predictability

The assumptions for Proposition 1 are that (i) the drift of a pricing kernel is affine in the state variable and (ii) the state variable follows affine diffusion under $\mathbb{Q}$. Three classes of well-known consumption-based asset pricing models incorporate this feature. They are the long-run risk framework of Bansal and Yaron (2004), the rare disasters framework of Wachter (2013), and the affine$^{\mathbb{Q}}$ habit formation model of Le, Singleton, and Dai (2010). Importantly, in all of these models, the source of time variation in risk premia is entirely spanned by volatilities in yields only (see Le and Singleton (2013) for detailed explanations). Once this salient feature of those models is incorporated with Proposition 1, it means that the bond VIX^2s should predict future excess returns and they are the sole source of time variation in risk premia.

Specifically, the assumptions in Proposition 1 are canonical in most long-run risks models without jumps (or rare disasters) - see e.g. Bansal and Yaron (2004), Bollerslev, Tauchen, and Zhou (2009), Bansal and Shaliastovich (2013), Zhou and Zhu (2015). For example, in Bansal and Shaliastovich (2013), (i) the drift of the pricing kernel is a linear function of the subset of affine diffusion state variables - expected consumption growth, expected inflation and their variance factors, and (ii) the $\mathbb{P}$ affine state variable, in conjunction with their market price of risk, imply $\mathbb{Q}$ affine state variable. The time-varying volatilities in expected consumption growth and expected inflation are the two economic fundamentals that induce time variation in volatilities in yields. In this economy, Proposition 1 implies that the bond VIX^2s should be the manifestation of uncertainty about the two macroeconomic fundamentals: expected consumption growth and expected inflation (see Appendix B for a formal derivation). Note that in long-run risk economies, the time-varying quantity of macroeconomic risk is the only source of time variation in risk premia - the price of risk is pinned down by Epstein-Zin preference. As a consequence, the bond VIX^2s should capture the entire innovations in risk premia though the channel of time-varying quantity of risk.

The rare disaster framework with time-varying disaster probabilities is another example that fits the assumptions of Proposition 1 - see e.g. Wachter (2013), and Tsai (2016). In this framework, the short rate is linearly dependant on time-varying risk of disasters (intensity of a disaster more precisely). The intensity process follows affine diffusion under both the physical and the risk-neutral measures, and it also determines volatilities in yields. Then, the bond VIX^2s disclose the time-varying probability of a disaster because of Proposition 1. Moreover, in this economy, time variation in risk premia solely stems from the time-varying probability of a disaster. Hence, the bond VIX^2s should have the ability to predict future excess returns.

The habit formation model in Le, Singleton, and Dai (2010), henceforth LSD, is another class of asset pricing models in which the bond VIX^2s should explain the entire time variation in risk premia. The model, based on Campbell and Cochrane (1999) and Wachter (2006), uses the two assumptions in Proposition 1 to obtain affine pricing. By doing so, they specify the market price of risk as a non-linear function of the states as in Duarte (2004) and, as a result, the state variable follows a non-linear process under $\mathbb{P}$. LSD shows that their model approximately nests Wachter's model and closely resembles its prominent features. In this type of affine$^{\mathbb{Q}}$ habit formation models with external habit level H_t, the consumption surplus ratio $s_t = \log\left[\left(C_t - H_t\right)/C_t\right]$ is the sole source of time-varying risk premia since the shocks on consumption growth (that drives the quantity of risk in the economy) are assumed to be homoscedastic. Furthermore, the volatilities in yields are driven by the non-negative process $\varphi_t = s_{\max} - s_t$ where $s_{\max}$ is the upper bound of s_t.[6] Hence, the bond VIX^2 is linear in φ_t, the inverse consumption surplus ratio, and contains the full information on risk premia through the reflection of the time-varying price of risk.

Motivated by the implication of Proposition 1 for the three classes of asset pricing models, I examine whether the bond VIX^2s explain time variation in expected bond excess returns. To assess their predictive ability, I initially apply MA2 filters for the one-month bond VIX^2s (with 5yr, 10yr and 30yr bonds as underlying assets) constructed in Choi, Mueller, and Vedolin (2016) with the aim of removing transitory shocks potentially due to measurement errors and institutional effects (see, for example, Kim (2007)). Then, I

[6]Note that s_t is always negative.

regress one-year holding period excess returns of bonds with different maturities onto the space of the three (filtered) one-month bond VIX^2, henceforth denoted as TIV^{2}s following Choi, Mueller, and Vedolin (2016). The projections indicate that, across all maturities, the excess returns' loadings on the TIV^{2}s exhibit tent-shape pattern akin to the pattern in Cochrane and Piazzesi (2005). Hence, in the spirit of Cochrane and Piazzesi (2005), I construct a single factor by projecting the average (across maturity) excess returns onto the three TIV^{2}s:

$$\overline{rx}_{t+12} = \gamma_0 + \gamma_1 TIV^2_{t,5yr} + \gamma_2 TIV^2_{t,10yr} + \gamma_3 TIV^2_{t,30yr} + e_t$$

The time-series of the fitted values (henceforth, CIV) is the return-forecasting factor, and is utilized to predict realized excess returns on each bond with maturity of n.

$$rx^{(n)}_{t+12} = b^{(n)}_0 + b^{(n)}_1 \left(\hat{\gamma}_0 + \hat{\gamma}_1 TIV^2_{t,5yr} + \hat{\gamma}_2 TIV^2_{t,10yr} + \hat{\gamma}_3 TIV^2_{t,30yr}\right) + e^{(n)}_t$$

For comparison purposes, the Cochrane-Piazzesi return-forecasting factor (henceforth, CP) is constructed by regressing mean excess returns onto the *spreads* of five Fama-Bliss forward rates with maturities of 1 through 5 years as in Cochrane and Piazzesi (2008). Excess returns from the Gürkaynak, Sack, and Wright (GSW) data set (with maturities of one through 10 years) are also utilized to assess excess returns on long-term bonds since the longest time-to-maturity of yields in the Fama-Bliss (FB) data set is five years.

[Insert Table 1 and Figure 2 here.]

Table 1 present adjusted R^2s and coefficients from predictive regressions of twelve-month bond excess returns on CP, CIV, and both CP and CIV jointly. Panel B shows that, for both sets (FB and GSW) of excess returns, each estimated coefficient on CIV is statistically significant during the sample period, and the variation in CIV explains more than 20% of the variation in realized mean excess returns. Once CIV and CP are jointly utilized, R^2s increase more than 10 percentage points in addition to the statistical significance of both coefficients. Panel A reports adjusted R^2 from regressing each excess return on the predictors, and it reveals that the predictability of the CIV stems mainly from excess returns of bonds with short-term maturities while the predictability of the CP comes from relatively long-term maturities. As a result, the adjusted R^2s are improved significantly once CIV and CP are utilized jointly to predict excess returns. Their joint significance can also be observed in Figure 2 where the time-series of CIV, CP and the mean realized excess returns from the GSW data set are plotted together. Between 1998 and 2002, for example, CIV and CP exhibit heterogeneous movements and can assist each other to explain the time variation in the excess returns.

4 Variance Risk and the Shape of the Yield Curve

How does volatility risk affect the shape of the yield curve? Do yields strongly/weakly load on volatility risk? Can we extract a reliable measure of interest rate volatility from the cross-section of bonds? The first step to addressing these questions is to identify volatility risk in a framework where volatility has a systematic impact on the yields across different maturities.

The class of affine dynamic term structure models (henceforth, ADTSM) is a typical example of such a framework, and has also served as the workhorse in the literature to assess the impact of volatility risk on the cross-section of bond yields; ADTSMs are fully characterized by (i) the two assumptions in Proposition 1, (ii) the specification of the market price of risk, and (iii) a set of parametric restrictions needed to identify the model. It is well established that the affine models successfully capture the cross-sectional properties of yields; see for example Dai and Singleton (2000). However, the ADTSMs' ability to capture variation in the volatility of interest rates is questionable and controversial, especially once volatility-sensitive derivatives are not incorporated into the estimation procedure of the model. For instance, using U.S. swap data only, Collin-Dufresne, Goldstein, and Jones (2009) show that the model implied volatilities from affine models seem unrelated to their non-parametric or semi-parametric counterparts (i.e. realized volatility estimates and GARCH estimates).

Because ADTSMs are built up on the two assumptions in Proposition 1, the bond VIX^2s directly represent variance risk in the model. In other words, the variance risk in ADTSM is readily identifiable via the bond VIX^2s as a consequence of Proposition 1. This identification strategy is beneficial in several ways. First, it is based directly on option prices that tend to be more sensitive to the changes in volatility than nominal bond prices. This is in line with previous studies pointing out that the introduction of volatility-sensitive instruments into the estimation procedure can significantly mitigate the difficulty in identifying volatility risk of affine models; see, for example, Bibkov and Chernov (2009), Almeida, Graveline, and Joslin (2011), Jagannathan, Kaplin, and Sun (2003), and Joslin (2014). In addition, the approach doesn't require us to estimate a specific model, and allows variance risk to be measured in real time. Furthermore, since the variance measure is constructed in a model-free manner, the approach can be easily incorporated into the class of Gaussian quadratic term structure models, as in Ahn, Dittmar, and Gallant (2002). In this case, $\mathcal{V}_t$ is a quadratic function of the Gaussian state factors in the model [7] (See Appendix C for a detailed explanation).

Before estimating a fully-fledged model to analyze the impact of variance risk on the cross-section of yields, I conduct two simple regression-based tests on the relationship between variance risk and the cross-section of yield. First, I examine the marginal impact of the variance risk on the shape of the yield curve beyond the traditional term structure factors: level, slope and curvature of the yield curve. The results suggest that variance risk is largely unrelated to the shape of the yield curve and that at least three non-volatility factors are required to adequately explain the cross-section of yields. The second test investigates whether variance risk can be identified from the cross-section of yields: the unspanned stochastic volatility (USV) effect in Collin-Dufresne and Goldstein (2002). The USV effect can be or cannot be rejected, depending on the number of variance factors. The empirical evidence will be utilized as guidance for designing highly parameterized term structure models in later sections.

4.1 The Shape of the Yield Curve and $\mathcal{V}_t$

Provided that (i) the state follows affine diffusion and (ii) the short rate is an affine function of the state, the yield on a zero-coupon bond of maturity n is affine in the state variable Z_t:

$$y_{n,t} = A_n\left(\Theta^{\mathbb{Q}}\right) + B_n\left(\Theta^{\mathbb{Q}}\right) Z_t \tag{5}$$

[7] For Gaussian quadratic term structure models, the short rate equation is a quadratic function of Gaussian state vector. The conditional variance of yields is linear in the square of a subset of the state.

where A_n and B_n are obtained from standard recursions as in Duffie and Kan (1996). The linear relationship between yields and factors in equation (5) implies that yields can be treated as state variables; given a set of maturities equal in number to the number of latent factors, one can rotate the underlying factor into the yields (see for example Pearson and Sun (1994), Chen and Scott (1993) and Duffie and Kan (1996) among many others). One can further rotate the risk factors into portfolios of yields, especially the principal component of yields, $\mathcal{P}_t$, as in Joslin, Singleton, and Zhu (2011) for example. As will be shown thoroughly in Section 5, Proposition 1 enables us to rotate the latent risk factors into portfolio of yields $\mathcal{P}_t$ and portfolio of options $\mathcal{V}_t$

$$y_t^o = \mathcal{A} + \mathcal{B}_{\mathcal{Z}}\mathcal{Z}_t = \mathcal{A} + \mathcal{B}_{\mathcal{P}}\mathcal{P}_t + \mathcal{B}_{\mathcal{V}}\mathcal{V}_t + e_t, \quad e_t \sim N\left(0, \sigma_e^2 I\right) \tag{6}$$

where y_t^o denotes a vector of stacked observed yields and e_t represents measurement error assumed to be an independent and homoscedastic Gaussian random variable (as commonly assumed in the literature). Because all the variables in equation (5) can be observable, the yields' loading on the factors $\mathcal{A}$, $\mathcal{B}_{\mathcal{P}}$ and $\mathcal{B}_{\mathcal{V}}$ can be estimated by linear regressions. The estimated model, then, can be treated as a standard linear factor model nesting the no-arbitrage affine models since $\mathcal{A}$, $\mathcal{B}_{\mathcal{P}}$ and $\mathcal{B}_{\mathcal{V}}$ are non-linear functions of $\Theta^{\mathbb{Q}}$ under the affine bond pricing models: see for example, Duffee (2011a), Hamilton and Wu (2012), Joslin and Le (2014) and Joslin, Le, and Singleton (2013).

The marginal impact of the variance risk beyond traditional yield factors like level, slope and curvature factors can be examined by comparing the likelihood of (6) with the following restricted version of it:

$$y_t^o = \mathcal{A}^* + \mathcal{B}_{\mathcal{P}}^*\mathcal{P}_t + e_t^*, \qquad e_t^* \sim N\left(0, \sigma_{e^*}^2 I\right) \tag{7}$$

Table 2 reports the test statistics of the likelihood ratio test for the hypothesis of the zero coefficients on the additional variable in the unrestricted version. The first column presents the right hand side variables in the restricted models where PC1-PC3 denotes the first, second and third principal components of yields on U.S. Treasury nominal zero-coupon bonds with maturities of six months and 1 through 10 years[8]. The remaining columns present an additional variable in each version of the unrestricted model and its corresponding LR statistics. VPC1 and VPC2 denotes the first and second principal components of the MA2 filtered 5, 10 and 30-year TIV^2s as in Section 3. Each of VPC1 and VPC2 capture respectively 94% and 5.6% of the variation in the three TIV^2s. The last column shows the 5% critical value of the test statistics which follows $\chi^2(11)$ distribution. The table indicates that for each version of the restricted model, its likelihood ratio is greatest when a yield factor (PC3 or PC4) is the additional variable in the unrestricted model. In other words, adding a PC factor to the restricted models is the best extension for the purpose of a better cross-sectional fit. Moreover, for the unrestricted models with the variance factors as the additional variables, the null can be rejected or not, but the magnitude of test statistics is not very large, regardless of their statistical significance. Similar results are obtained once two or three representative yields, instead of the yield PCs, are utilized as the right hand side variables of the restricted models (the results are omitted in the paper for the sake of brevity). In sum, the exercise indicates that the marginal benefit of adding variance factors is fairly limited and it is hard to identify variance risk from the cross-section of yields.

[Insert Table 2 here.]

The exercise also implies that it is empirically difficult to extend the estimation approach of Hamilton and Wu (2012) into affine bond pricing models with stochastic volatilities. They propose a minimum-chi-square estimation procedure of Gaussian ADTSM in which the risk-neutral parameters of the model are inferred by minimizing the differences between the ordinary least square (OLS) estimates of the cross-sectional equation (7) and the corresponding yields' loadings from Gaussian ADTSM. In theory, their approach can be applied to equation (6) for the estimation of ADTSM with stochastic volatilities. However, the limited impact of the variance risk on the shape of the yield curve causes difficulties in its empirical implementation. The OLS estimates of $\mathcal{B}_{\mathcal{V}}$ in equation (6) are not informative enough to precisely pin down the risk-neutral parameters related to the volatility factors.

[8]The yields with maturities of two to ten years are from Gürkaynak, Sack, and Wright (2007). The six-month and one-year yields are bootstrapped from observed bond prices using the Fama-Bliss methodology. My thanks to Anh Le for allowing me to use this data set.

4.2 Unspanned Stochastic Volatility Effect

The difficulty of identifying volatility risk under ADTSM stems from the multiple roles of volatility risk in the class of affine models. The volatility risk affects (i) the second moments of yields, (ii) the expectation of future interest rates under both physical and risk-neutral measures, and (iii) the so-called convexity effect introduced by the non-linear relationship between bond prices and the latent factors. The various roles of volatility enables us to infer it through multiple channels, but this feature causes tension rather than a complimentary effect in identifying it (see Joslin and Le (2014) for a detailed explanation).

One potential resolution for the issue is to impose a set of model-based restrictions to remove the dependence of the cross-section of yields on volatility, a set of restrictions coined as an "unspanned stochastic volatility" (USV) restriction by Collin-Dufresne and Goldstein (2002). More broadly, the USV effects mean that the yields curve itself fails to span the volatilities in the changes in yields. In their seminar paper, Collin-Dufresne and Goldstein (2002) define the USV effect as the existence of a set of parameters $\{\phi_1, ..., \phi_N\}$ that are not all zero such that

$$\sum_{i=1}^{N} \phi_i B_{n,i} = 0 \quad \forall n > 0 \tag{8}$$

where N is the number of pricing factors and $B_{n,i}$ the i-th element of B_n in equation (5). The authors further show that, under the existence of such a set of parameters with $N \geq 3$, one can find a rotation such that the variance factor V_t has no effect on the price of bonds. As a consequence, the variance factor cannot be extracted from the cross-section of observed yields (see Collin-Dufresne and Goldstein (2002), and Joslin (2015) for further details).

The following studies, however, have accumulated conflicting evidence on the USV effect. Decoupling the dual role of volatility through the USV restrictions helps the model to produce more realistic model-implied volatility, even though the model's cross-sectional fit is slightly impeded (see for example Creal and Wu (2015) and Collin-Dufresne, Goldstein, and Jones (2009) among others). Andersen and Benzoni (2010) also show that their measure of intraday volatility in yields is largely unexplained by term structure factors, which is in line with the USV effect. On the other hand, the USV effect is rejected once the model-specific restrictions are directly tested by the likelihood-ratio or the Wald test (Bibkov and

Chernov (2009), Joslin (2015)). Utilizing the observable volatility proxy, I devise a new test for the USV effects, which can shed new light on the debate.

The condition for the USV effect in equation (8) can be translated into the statement that the matrix $\mathcal{B}_{\mathcal{Z}} \equiv [\mathcal{B}_{\mathcal{P}}, \mathcal{B}_{\mathcal{V}}]$ in equation (6) is not full rank, regardless the maturities of the yields on the left hand side of the equation (see Appendix E for a formal derivation). As a result, a statistical test for the rank of the estimated matrix $\hat{\mathcal{B}}_{\mathcal{Z}}$ is a test of the USV effect. The null hypothesis is

$$H_0 : \text{rank}\left(\mathcal{B}_{\mathcal{Z}}\right) \leq N - 1 \tag{9}$$

where N is the total number of factors. I use the Kleibergen-Paap rank test, among many other rank tests. The test statistic follows χ^2 distribution: for details, see Kleibergen and Paap (2006).

The approach has several benefits not shared by other tests for the USV effect in the literature. First, it is a formal statistical test - many of others in the literature are not formal statistical tests as pointed out by Bibkov and Chernov (2009). Second, while Bibkov and Chernov (2009) and Joslin (2015) conduct formal tests for the set of restrictions generating the USV effect, the USV restrictions are not unique as pointed out by Joslin (2015). For example, two different sets of restriction on the $A_1(4)$ specification can induce the USV effect while the two models fit volatilities in significantly different manners; see for example Creal and Wu (2015). The rank test that I posit here is free from this issue. Finally, the test can be implemented even in the presence of hidden factors as in Duffee (2011b) or Joslin, Priebsch, and Singleton (2014) - a detailed explanation of the hidden factors can also be found in Section 6. The test only exploits the cross-sectional relationship between the yields and variance factors, so the test results should be identical even after taking into account hidden factors.

[Insert Table 3 here.]

Table 3 reports the test statistics for specifications with one through two volatility factors in conjunction with two through three additional non-volatility factors. Following Dai and Singleton (2000), $A_m(N)$ denotes an N factor model with m factor driving volatility. The data set is the same as the one in Section 4.1, and the first m PCs of the MA2 filtered TIV^{2}s are used as the variance factors for $A_m(N)$ models. Specifications with up to two volatility factors are considered for the exercise because the first two PCs of TIV^{2}s explain 99% variation of the three TIV^{2}s as pointed out in Section 4.1. The table shows that, for all the specifications, the null (the presence of USV effects) is rejected at the 10% significance level. Hence, the conditions for the affine models to generate USV effect do not hold in the data.

In sum, it is true that variance risks are hard to identify from the cross-section of yields as shown in Section 4.1, however, the knife-edge conditions for the USV effect are rejected in the data. In other words, the variance risk is effectively unspanned by yields not because of the USV restrictions but because of its limited impact on the shape of the yield curve, and it can be hardly identified without help of option prices.

5 A New Representation of ADTSM

In this section, I suggest a new representation of ADTSM in which all factors are represented as portfolios of bonds and options. The representation inherits the spirit of Joslin, Singleton, and Zhu (2011), and the advantages of their representation. Since all the term structure factors (including volatility) are observable, the estimation procedure becomes greatly simplified and economic interpretation of the model is more straightforward compared to conventional latent factor approaches.

For econometric identification, I initially assume that the risk-neutral dynamics of the latent factor in equation (1) is drift normalized as in Joslin (2015) or Creal and Wu (2015). The yield on a zero-coupon bond of maturity n is affine in the states Z_t:

$$y_{n,t} = A_n\left(\Theta^{\mathbb{Q}}\right) + B_n\left(\Theta^{\mathbb{Q}}\right) Z_t$$

where A_n and B_n are obtained from standard recursions as in Duffie and Kan (1996). I let $(n_1, n_2, ..., n_J)$ be the set of maturities of the bonds used in estimation and y_t be the $(J \times 1)$ vector of corresponding yields. For any full-rank matrix $W \in \mathbb{R}^{(N-m)\times J}$, Wy_t represents the associated $(N-m)$-dimensional set of portfolios of J $(\geq N)$ yields. Following Joslin, Singleton, and Zhu (2011), I let $\mathcal{P}_t$ denote the first $(N-m)$ principal components (PCs) of J yields with W being the weighting matrix of the PCs:

$$\mathcal{P}_t = Wy_t = A_W\left(\Theta^{\mathbb{Q}}\right) + B_W\left(\Theta^{\mathbb{Q}}\right)Z_t = A_W\left(\Theta^{\mathbb{Q}}\right) + B_{W,X}\left(\Theta^{\mathbb{Q}}\right)X_t + B_{W,V}\left(\Theta^{\mathbb{Q}}\right)V_t$$

Invoking Proposition 1, then, we can define the N observable pricing factors $\mathcal{Z}_t$ such that

$$\mathcal{Z}_t \equiv (\mathcal{P}_t', \mathcal{V}_t')' = \left((Wy_t)', \mathcal{V}_t'\right)' = U_0 + U_1(X_t', V_t')' \tag{10}$$

where

$$U_0 = \begin{bmatrix} A_W \\ \alpha \end{bmatrix}, \quad U_1 = \begin{bmatrix} B_{W,X} & B_{W,V} \\ 0_{m\times(N-m)} & \beta \end{bmatrix}$$

with α and β defined in Proposition 1. The dynamic of $\mathcal{Z}_t$ can be represented as a function of the observable factor $\mathcal{Z}_t$ after applying the invariant rotation of Dai and Singleton (2002) to the latent factor Z_t. Provided that the mapping between $\mathcal{Z}_t$ and Z_t is bijective (i.e. one-to-one mapping), the model with observable $\mathcal{Z}_t$ is observationally equivalent to the representation with the latent Z_t. The sufficient condition for the mapping to be bijective is a full rank matrix β. Once the Gaussian factor X_t is drift normalized, Joslin (2015) shows that the matrix $B_{W,X}$ should be full rank. Hence, the first $(N-m)$ columns of U_1 are linearly independent. With non-zero β, the last m columns of U_1 are not spanned by the first $(N-m)$ columns of U_1, which implies that a full rank matrix β guarantees U_1 to be not rank deficient.

The new representation of ADTSM with the observable factors $\mathcal{Z}_t$ in equation (10) follows the idea of Joslin, Singleton, and Zhu (2011), henceforth JSZ, and can be considered an extension of their work into general affine models. JSZ suggests a new representation of Gaussian ADTSM in which all the Gaussian pricing factors are observable as portfolios of yields, i.e. $\mathcal{P}_t$ in equation (10). Through their representation, the estimation of the Gaussian term structure model is extremely simplified, and becomes more reliable in terms of finding a global optimum in maximum likelihood estimation. In particular, simple ordinary least square estimation (OLS) can be utilized to estimate the $\mathbb{P}$ conditional mean parameters of the pricing factors which had been treated as one of the most challenging parts in estimating term structure models due to the high degree of persistence in yields.

In my representation, all pricing factors (including the variance risk), are observable. As a consequence, one can make use of generalized least square estimation (GLS) to pin down the drift of the pricing factor under $\mathbb{P}$. Since variance is directly observational up to its linear transformation via $\mathcal{V}_t$, it is also easy to estimate the parameters governing the time-series dynamics of $\mathcal{V}_t$. In addition, when volatility risk is identified from the cross-section of yields, one should solve a numerically unstable equation $AX = b$ where A is often nearly singular, with the possibility that the solution leads to negative values for volatility: see for example Piazzesi (2010) and Joslin (2014). Instead, the representation I posit here is unaffected by this issue. In sum, the representation helps us find the global optimum of maximum likelihood estimation by simplifying the two hardest parts of the $ADTSM$ with stochastic volatility estimation, namely, the identification of volatility, as well as the drift of the state under $\mathbb{P}$.

The parameterization scheme using portfolios of yields as pricing factors for $A_m(N)$ model is also explored in Joslin and Le (2014), where the variance factor in $A_m(N)$ is approximated by portfolios of yields. The model I posit here utilize portfolios of options rather than portfolios of yields, and the variance factor is known before the model estimation while their variance factors can only be identified after the model estimation. In addition, the approach here is robust even in the presence of unspanned stochastic volatility factors as in Section 4.2, while their approach only works for spanned stochastic volatility. Furthermore, as discussed extensively in their paper, extracting the variance factor from yields only (without options) results in undesirable properties in the factor dynamics under $\mathbb{P}$ - this issue is discussed further in Section 7.2.

6 Unspanned Macro Risk and the Likelihood Function

More recently, a large literature has been studying so-called hidden factors or unspanned macro factors, see e.g., Duffee (2011b), Chernov and Mueller (2012) and Joslin, Priebsch, and Singleton (2014), henceforth JPS. A factor is described as hidden if it plays an important role in determining investors' expectations for future yields, yet is not priced in the fixed income market. Hence, the hidden factor cannot be recovered from the cross-section of any fixed income assets. This section explains how to take into account the hidden factor inside the model described in the previous section.

Since all priced factors are observable due to the representation in the previous section, the same argument as in JPS can be applied in order to add hidden factors in the framework. Once both hidden and non-hidden factors are projected onto the space of fixed income asset returns as in JPS, we get the following factor dynamics under the physical measure $\mathbb{P}$ and the risk-neutral measure $\mathbb{Q}$. First, the factors are composed of (i) the priced risks in the fixed income market $\mathcal{Z}_t = (\mathcal{P}_t', \mathcal{V}_t)'$ and (ii) a non-priced (hidden) factor M_t. In discrete time setting, the dynamics of the non-variance factors, $(\mathcal{P}_t', M_t)$, can be represented as

$$\begin{bmatrix} \mathcal{P}_{t+1} \\ M_{t+1} \end{bmatrix} = \begin{bmatrix} K_{0\mathcal{P}}^{\mathbb{P}} \\ K_{0M}^{\mathbb{P}} \end{bmatrix} + \begin{bmatrix} K_{\mathcal{P}\mathcal{P}}^{\mathbb{P}} & K_{\mathcal{P}M}^{\mathbb{P}} \\ K_{M\mathcal{P}}^{\mathbb{P}} & K_{MM}^{\mathbb{P}} \end{bmatrix} \begin{bmatrix} \mathcal{P}_t \\ M_t \end{bmatrix} + \begin{bmatrix} K_{\mathcal{P}\mathcal{V}}^{\mathbb{P}} \\ K_{M\mathcal{V}}^{\mathbb{P}} \end{bmatrix} \mathcal{V}_t + \begin{bmatrix} \Sigma_{\mathcal{P}\mathcal{V}} \\ \Sigma_{M\mathcal{V}} \end{bmatrix} \epsilon_{\mathcal{V},t+1}^{\mathbb{P}} + \begin{bmatrix} \epsilon_{\mathcal{P},t+1} \\ \epsilon_{M,t+1} \end{bmatrix} \tag{11}$$

with

$$\begin{aligned} \left(\epsilon_{\mathcal{P},t+1}', \epsilon_{M,t+1}'\right)' &\sim N\left(0, \Sigma_t\right) \\ \Sigma_t &= \Sigma_0 + \Sigma_1\left(\mathcal{V}_t - \alpha\right) \\ \epsilon_{\mathcal{V},t+1}^{\mathbb{P}} &= \mathcal{V}_{t+1} - E_t\left(\mathcal{V}_{t+1}\right) \end{aligned} \tag{12}$$

The variance factor $\mathcal{V}_{t+1}$ follows a compound autoregressive gamma process

$$\mathcal{V}_{t+1}|\mathcal{V}_t \sim CAR\left(\rho^{\mathbb{P}}, c^{\mathbb{P}}, \nu^{\mathbb{P}}, \alpha\right)$$

where $c^{\mathbb{P}}$ is a scale parameter, $\nu^{\mathbb{P}}$ is a shape parameter, and $\rho^{\mathbb{P}}$ determines the autocorrelation of $\mathcal{V}_t$. The lower bound of $\mathcal{V}_t$ is set α, contrary to the standard lower bound of zero for a variance process. Indeed, the lower bound of the latent variance factor V_t should be set to zero for econometric identification. Then, the linear relationship between the observable $\mathcal{V}_t$ and the latent V_t, $\mathcal{V}_t = \alpha + \beta V_t$ with $\alpha\left(\Theta^{\mathbb{Q}}\right)$ defined in Proposition 1, implies that $\mathcal{V}_t$ should be greater than α. Furthermore, α should be positive since both $\alpha\left(\Theta^{\mathbb{Q}}\right)$ and $\beta\left(\Theta^{\mathbb{Q}}\right)$ capture the convexity components of yields - see equation (A-8) and Appendix D.2 for details. The non-zero lower bound of $\mathcal{V}_t$ also leads Σ_t in equation (12) to be $\Sigma_0 + \Sigma_1\left(\mathcal{V}_t - \alpha\right)$ rather than $\Sigma_0 + \Sigma_1 \mathcal{V}_t$. For a detailed explanation of compound autoregressive processes, see Gourieroux and Jasiak (2006), Le, Singleton, and Dai (2010) and Creal and Wu (2015).

Under the pricing measure $\mathbb{Q}$, the dynamics of $\mathcal{Z}_t$ are assumed to be

$$\begin{aligned} \mathcal{P}_{t+1} &= K^{\mathbb{Q}}_{0\mathcal{P}} + K^{\mathbb{Q}}_{\mathcal{P}\mathcal{P}}\mathcal{P}_t + K^{\mathbb{Q}}_{\mathcal{P}\mathcal{V}}\mathcal{V}_t + \Sigma_{\mathcal{P}\mathcal{V}}\epsilon^{\mathbb{Q}}_{\mathcal{V},t+1} + \epsilon_{\mathcal{P},t+1} \\ \mathcal{V}_{t+1}|\mathcal{V}_t &\sim CAR\left(\rho^{\mathbb{Q}}, c^{\mathbb{Q}}, \nu^{\mathbb{Q}}, \alpha\right) \end{aligned} \tag{13}$$

Hence, the specification of the price of risks follows that in Cheridito, Filipović, and Kimmel (2007), and yields can be represented as a linear function of $(\mathcal{P}_t', \mathcal{V}_t)'$ where yields' loadings on the pricing factors are determined by $\Theta^{\mathbb{Q}}$ (see Appendix D.1).

Furthermore, in order to maintain (i) the diffusion invariance property of the variance process $\mathcal{V}_t$ and (ii) non-exploding market price of risk in the continuous time limit (see Appendix B.4 in Joslin and Le (2014) for explanations), I impose the following two restrictions on parameters for $\mathcal{V}_t$:

$$c^{\mathbb{P}} = c^{\mathbb{Q}}, \quad \nu^{\mathbb{P}} = \nu^{\mathbb{Q}}$$

For the fitting of the cross-section, I assume that higher-order PCs, denoted by $\mathcal{P}_{e,t}$, are observed with i.i.d. uncorrelated Gaussian measurement errors with a common variance:

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

In sum, the likelihood function of the observed data, $\mathcal{L}$, is

$$\mathcal{L} = \sum_t f\left(\mathcal{P}_{t+1}, M_{t+1}|\mathcal{V}_{t+1}, \mathcal{I}_t\right) + f\left(\mathcal{V}_{t+1}|\mathcal{V}_t, \mathcal{I}_t\right) + f\left(\mathcal{P}_{e,t+1}|\mathcal{P}_{t+1}, \mathcal{V}_{t+1}\right)$$

where f denotes the log conditional density. The first two terms capture the density of the time-series dynamics, and the last term is the density of the cross-sectional fit on which the unspanned macro factors M_t have no impact. Particularly, the $\mathbb{P}$-feedback matrix of $\mathcal{P}_t$ and M_t can be concentrated out by running GLS of the following system:

$$\begin{bmatrix} \mathcal{P}_{t+1} - \Sigma_{\mathcal{P}\mathcal{V}}\epsilon^{\mathbb{P}}_{\mathcal{V},t+1} \\ M_{t+1} - \Sigma_{M\mathcal{V}}\epsilon^{\mathbb{P}}_{\mathcal{V},t+1} \end{bmatrix} = \begin{bmatrix} K^{\mathbb{P}}_{0\mathcal{P}} \\ K^{\mathbb{P}}_{0M} \end{bmatrix} + \begin{bmatrix} K^{\mathbb{P}}_{\mathcal{P}\mathcal{P}} & K^{\mathbb{P}}_{\mathcal{P}M} \\ K^{\mathbb{P}}_{M\mathcal{P}} & K^{\mathbb{P}}_{MM} \end{bmatrix} \begin{bmatrix} \mathcal{P}_t \\ M_t \end{bmatrix} + \begin{bmatrix} K^{\mathbb{P}}_{\mathcal{P}\mathcal{V}} \\ K^{\mathbb{P}}_{M\mathcal{V}} \end{bmatrix} \mathcal{V}_t + \begin{bmatrix} \epsilon_{\mathcal{P},t+1} \\ \epsilon_{M,t+1} \end{bmatrix}$$

The observable variance $\mathcal{V}_t$ can be either spanned by yields or unspanned (i.e. of the unspanned stochastic volatility type as in Section 4.2). However, this does not affect the estimation procedure, since the volatility factor is identified not via yields but via options, even before the estimation procedure. In contrast, without the observable volatility factor, one should choose a specific set of restrictions on the $\mathbb{Q}$ parameters (among many possible set of restrictions), in order to estimate a model with unspanned stochastic volatilities. Otherwise, the identification of the volatility factor is infeasible, because it has no effect on the price of bonds.

In the term structure literature, both the unspanned stochastic volatility and hidden factors have been considered important components driving the time variation in risk premia, although their mechanisms are totally different. The effect of hidden factors on changes in risk premia exactly cancels out its effect on expectations of future short rate while USV implies a cancelation of the convexity bias. The USV factor can be identified from interest rate derivatives while hidden factors cannot be identified from any financial instrument in the market. To the best of my knowledge, my model is the first one capable of accommodating both types of unspanned risks: the unspanned stochastic volatility factors as well as the hidden factors.

7 Model Comparison

7.1 Model Specifications and Data

The discussion in Section 4 indicates that the variance risks' explanatory power for the cross-section of yield is fairly limited when it is compared to the explanatory power of the three term structure factors, level, slope and curvature. Under the representation in Section 5, this implies that at least three Gaussian factors are required to adequately explain the shape of the yield curve. Hence, I study a model with three yield factors and one stochastic volatility, which I denote by $A_1(4)$ as in Section 4.2. Its corresponding specification with two unspanned macro risks, denoted as $UMA_1^2(6)$, is also investigated; $UMA_m^R(N)$ stands for the family of ADTSMs in Section 6, with $(\mathcal{P}_t', M_t', \mathcal{V}_t')'$ of dimension N, M_t of dimension R, and $\mathcal{V}_t$ of dimension m. Following JPS, I use measures of economic activity and inflation as the two unspanned macro risks. In particular, the three-month moving average of the Chicago Fed's National Activity Index (CFNAI), henceforth denoted as *GRO*, is used as the measure of the *growth* in real economic activity as in JPS. However, I use year-over-year growth in Consumer Price Index excluding food prices and energy prices (henceforth,

CPI) for the measure of inflation, contrary to JPS in which the measure of inflation is the expected rate of inflation from Blue Chip Financial Forecasts (henceforth, INF).

Moreover, the same $UMA_1^2(6)$ specification but with a different measure of economic activity - the unemployment gap - is also studied. The unemployment gap (henceforth $UGAP$) is the difference between the actual unemployment rate and the estimate of the natural rate of unemployment from the Congressional Budget Office (CBO). Hence, it gauges the *level* of economic activity rather than the *growth* of activity. Bauer and Rudebusch (2016), henceforth BR, argue that level indicators of activity such like $UGAP$ are largely related to the movement of the yield curves (i.e. weakly unspanned by yields) because these variables are relevant for setting the short-term policy rates; the authors also point out that the empirical monetary policy rules literature has identified level rather than growth variables as those which are most important for determining monetary policy (e.g. Taylor (1993), Taylor (1999), Orphanides (2003), Bean (2005) and Rudebusch (2006) among others). On the other hand, measures of growth in economic activity such as GRO are largely uncorrelated with the level of activity; see for example $UGAP$ and GRO in Figure 3. Furthermore, BR show that growth variables accompany low R^2s when (i) they are projected onto term structure factors or (ii) fed fund rates are regressed on them. Hence, they are strongly unspanned by yields. The different spanning properties of $UGAP$ and GRO induce significantly different estimates of risk premia for the $UMA_0^2(5)$ models in BR. BR qualitatively assess the relevance of two different estimates of risk premia and claim that $UGAP$ is a better measure of economic activity. I access four unspanned models $UMA_0^2(5)$ and $UMA_1^2(6)$ and evaluate their relevance based on whether they can match the pattern for violation of the expectations hypothesis.

[Insert Figure 3 here.]

The yield data set is the same as that in Section 4, but I only use yields with maturities of six months, 1 through 3 years, 5, 7, 9 and 10 years. As a measure of observable volatility, I make use of the 30-year TIV.

7.2 The Campbell and Shiller Regression

The most well-known stylized fact in the fixed income market is the failure of the expectations hypothesis (see, for example, Fama and Bliss (1987) or Cambpell and Shiller (1991) among many others). As pointed out by Dai and Singleton (2002), this prominent pattern of return predictability can serve as a measure to access the goodness-of-fit of term structure models - one can investigate whether a model implied data-generating-process can re-produce the observed pattern in the data. In this section, I assess the ability of each model posited in Section 7.1 by comparing the extent to which each of them can match the important stylized fact of bond yields. The investigation reveals that the identification of a variance factor through the portfolio of options $\mathcal{V}_t$ strikingly enhances the models' ability to reproduce important patterns in the data. Furthermore, it is shown that the usage of different unspanned macro risks also determines the models' ability to generate the stylized fact. Hence, this property of models can serve as a useful guidance in selecting unspanned macro risks.

The expectation hypothesis implies that the changes in yields are solely attributed to the revision of future expected interest rates. As a result, high yield spreads should proceed increases in long rates, and changes in risk premia play no role in determining the shape of yield curves. One way of testing the expectations hypothesis is to regress realized changes in yields onto yield spreads

$$y_{n-h,t+h} - y_{n,t} = \phi_0^{(n)} + \phi_1^{(n)} \left(\frac{h}{n-h} \left(y_{n,t} - y_{h,t} \right) \right) \tag{14}$$

as in Cambpell and Shiller (1991). While $\phi_1^{(n)}$ should be one for $n > h$ under the null, the estimated $\phi_1^{(n)}$'s are typically negative and their magnitudes are increasing with n, i.e. for longer yields maturities. Dai and Singleton (2002) named this property of linear projections of yields as LPY, and suggested treating the projections as descriptive statistics that any empirically desirable term structure model should replicate. They find that the population coefficients $\phi_1^{(n)}$ implied by estimated Gaussian models closely match their data counterparts. In contrast, the affine models with stochastic volatilities are not capable of generating this pattern, and counter-factually imply that the expectations hypothesis nearly holds: $\phi_1^{(n)}$'s typically stay close to one across all maturities. In other words, the affine models with stochastic volatilities fail to match the key empirical relationship between expected returns and the slope of the yields curve. However, Almeida, Graveline, and Joslin (2011) document that the stochastic volatility models can be as good as Gaussian models in generating the LPY property, once options data are incorporated into the estimation procedure.

[Insert Table 4 here.]

Table 4 reports the LPY property of each model. The stochastic volatility models with or without unspanned macro risks outperform the corresponding Gaussian models - this is in line with the findings in Almeida, Graveline, and Joslin (2011). The result first implies that the observational variance risk is beneficial in explaining the time variation in risk premia. Second, it indicates that the bond VIX is a properly identified measure of volatility risk in the Treasury market. As pointed out in Joslin and Le (2014), $\phi_1^{(n)}$ in equation (14) is mainly determined by the physical feedback matrix of factors. The population value of $\phi_1^{(n)}$ is

$$\phi_1^{(n)} = \frac{(n-h)}{h} \frac{\left(B_{n-h}\left(K_{1\mathcal{Z}}^{\mathbb{P}}\right)^h - B_n\right)\Sigma\left(B_n - B_h\right)'}{\left(B_n - B_h\right)\Sigma\left(B_n - B_h\right)'} \tag{15}$$

where Σ denote the unconditional covariance matrix of the time-series innovations and B_n is the yield's loadings on the observational factors $\mathcal{Z} = (\mathcal{P}_t', M_t', \mathcal{V}_t')'$. The loadings are almost identical across all the models - the variance factor's marginal impact on the cross-section is minimal as shown in Section 4, and the yield's loadings on unspanned macro risks are zero by construction. Since the covariance matrix Σ is in both the numerator and denominator of equation (15), its impact cancels out. Hence, $K_{1\mathcal{Z}}^{\mathbb{P}}$ is the key that causes the variation in the LPY property of each model. In the case of Gaussian models, the physical feedback matrix of $(\mathcal{P}_t', M_t')'$ is estimated by OLS. Hence, the estimates should be biased if the conditional volatility of $\mathcal{P}_t$ is time varying (which is strongly evident in the data). In the presence of $\mathcal{V}_t$, the bias of OLS estimates can be corrected because the physical feedback matrix of $(\mathcal{P}_t', M_t')'$ is estimated by GLS. However, the correction works only if the the instrument of conditional volatility can truly resemble the data generating process. The outperformance provides evidence that the bond VIX is a well-identified measure of volatility risk in the Treasury market.

[Insert Figure 4 and Figure 5 here.]

Figure 4 plots the $\phi_1^{(n)}$s from $A_0\,(3)$ and $A_1\,(4)$ models where six-month changes in yields are the dependent variables in the Campbell-Shiller regression. The $\phi_1^{(n)}$s from corresponding unspanned models, with GRO and CPI as the macro risks, are also displayed. They are notably worse than the three factor Gaussian model, $A_0\,(3)$. Figure 5 plots the same but with $UGAP$ as a measure of economic activity. Contrary to Figure 4, the unspanned models' performances are much improved, and it becomes even better than $A_0\,(3)$ model once the unspanned model incorporates the variance risk. Hence, the unemployment gap, a policy factor, can be considered a more relevant measure of real economic activity than GRO for a macro-finance term structure modeling. It also indicates that the impact of unspanned risk might not be as prominent as asserted by JPS in which CFNAI is utilized as a measure of output growth.

8 Risk Premia Accounting

This section studies the risk premia implied by $UMA_1^4\,(6)$ with UGAP and CPI as unspanned macro risks[9]. The risk premia on the risk factor $\mathcal{P}_t$ are the difference between the conditional expectation of $\mathcal{P}_{t+1}$ from the physical dynamics of equation (11) and the risk-neutral dynamics of equation (13). They are determined by the full set of the state $\mathcal{Z}_t^* \equiv (\mathcal{P}_t', M_t', \mathcal{V}_t)'$ rather than solely by pricing factors $(\mathcal{P}_t', \mathcal{V}_t)'$:

$$E_t^{\mathbb{P}}\left(\mathcal{P}_{t+1}\right) - E_t^{\mathbb{Q}}\left(\mathcal{P}_{t+1}\right) = \left[K_{0\mathcal{P}}^{\mathbb{P}} - K_{0\mathcal{P}}^{\mathbb{Q}}\right] + \left[\begin{array}{ccc} \left(K_{1\mathcal{P}\mathcal{P}}^{\mathbb{P}} - K_{1\mathcal{P}\mathcal{P}}^{\mathbb{Q}}\right) & K_{1\mathcal{P}M}^{\mathbb{P}} & \left(K_{1\mathcal{P}\mathcal{V}}^{\mathbb{P}} - K_{1\mathcal{P}\mathcal{V}}^{\mathbb{Q}}\right) \end{array}\right] \mathcal{Z}_t^*$$

Furthermore, the risk premia on $\mathcal{P}_t$ are, to a first-order approximation, the scaled excess returns on the yield portfolios whose value change locally one-to-one with changes in $\mathcal{P}_t$. In other words, the first row of $E_t^{\mathbb{P}}\left(\mathcal{P}_{t+1}\right) - E_t^{\mathbb{Q}}\left(\mathcal{P}_{t+1}\right)$ is the scaled excess return on the factor mimicking portfolio of $PC1$ while its value is unresponsive to changes in $PC2$, $PC3$, and $\mathcal{V}$ (see Appendix F for a detailed explanation - it extends the similar analysis of JPS for Gaussian unspanned models into affine models with stochastic volatilities.)

[9]Since both UGAP and CPI are weakly unspanned, the model also can be treated as a shortcut of a spanned macro-finance term structure model as pointed out by Bauer and Rudebusch (2016).

Denoting these PC-mimicking portfolios as $xPC = \Lambda_0 + \Lambda_1 \mathcal{Z}_t^*$, I impose a set of zero restrictions on Λ_0 and Λ_1 due to the concerns about over-parameterization caused by the large number of parameters of the model (see, for example, Duffee (2010)). Furthermore, the constraint is economically interpretable since all the factors are tradable portfolios and macro variables. As pointed out by JPS (for their Gaussian models), no such model-free interpretation is feasible with a latent factor model. To figure out an adequate set of zero restrictions, I initially estimate the fully flexible version of a model in which no element of Λ_0 and Λ_1 is constrained to be zero. Then, the elements of the first estimates without statistical significance at 5% level are set as zero for the next step of estimation[10]. This procedure is repeated until I find that every non-constrained element of Λ_0 and Λ_1 is statistically significant. Table 5 displays the resulting estimates of Λ_0 and Λ_1. It indicates that exposure to both level and slope risks is priced, while exposure to curvature risk is not. Increase in the level of uncertainty, $\mathcal{V}_t$, induces higher expected return on the level mimicking portfolio even though $\mathcal{V}_t$'s impact on the cross-section of bonds is noticeably small as expected from the exercises in Section 4: see Appendix G for a detailed description of the risk-neutral parameters and cross-sectional fit of the model. Also, positive shocks on the measure of inflation (CPI) raise the risk premia on both level and slope risks. On the other hand, the level measure of economic activity, $UGAP$, does not contribute to the evolution of risk premia at all. Its unspanned component provides no relevant information for the time-variation in risk premia, because $UGAP$ is largely spanned by yield curve components as documented in Bauer and Rudebusch (2016).

[Insert Table 5 and Figure 6 here.]

Figure 6 plots the estimates of one-year simple expected excess returns from the model with constraints on Λ_0 and Λ_1 as in Table 5. Henceforth, this model is denoted as $\mathcal{M}_C$. For comparison purposes, I also estimate the preferred model in JPS (henceforth, $\mathcal{M}_J$). Based on information criteria, they conduct model selection searches over Λ_0 and Λ_1 of $UMA_0^2(5)$ with GRO and Blue Chip inflation forecasts as unspanned macro risks. One-year expected excess returns on 2-year and 10-year bonds from $\mathcal{M}_C$ and $\mathcal{M}_J$ are plotted

[10] Asymptotic standard errors are computed by numerical approximation to the Hessian and using the delta method.

in Figure 7. The term premia from $\mathcal{M}_C$ peak early in the recovery or near the end of the recession, and they are more volatile than the term premia from $\mathcal{M}_J$, especially for long-term bonds. For example, expected excess returns on a 10-year bond implied by $\mathcal{M}_J$, $E_t\left(RX^{(10yr)}_{t\to t+12}|\mathcal{M}_J\right)$, is much less time-varying than the expected excess return from $\mathcal{M}_C$, $E_t\left(RX^{(10yr)}_{t\to t+12}|\mathcal{M}_C\right)$. Table 6 reports R^2s from projecting one-year realized excess returns of n-year bonds onto their corresponding model implied expected excess returns from $\mathcal{M}_C$ and $\mathcal{M}_J$. Expected excess returns from $\mathcal{M}_C$ explain, across all maturities, about 30% of time variation in realized excess returns. On the other hand, $\mathcal{M}_J$ is particularly good at capturing excess returns on short-term bonds, but its explanatory power diminishes along long-term bonds.

[Insert Table 6 and Figure 7 here.]

The estimates of expected returns from June 2004 to June 2006 are of particular interest. During this period, the Federal Open Market Committee raised the policy rates 25 basis points for 17 consecutive meetings while the long-end of the yield curve remained relatively constant. The puzzling behaviour of long-rates has been labeled as a “conundrum” by the former Chairman Greenspan, and subsequent studies have attributed the phenomenon to declining risk premia. A comparison of the top and bottom panels in Figure 7 indicates that incorporating time-varying variance induces more a prominent reduction in risk premia during the conundrum period. Moreover, Figure 6 and Figure 7 show that *negative* one-year expected returns are associated with the conundrum period. The negative expected bond return, especially implied by $\mathcal{M}_C$, has an interesting implication for the design of structural asset pricing models. As shown in Martin (2016), any expected gross return R_T can be decomposed into

$$E^{\mathbb{P}}_t\left(R_T - R_{f,t}\right) = \frac{1}{R_{f,t}}\mathrm{var}^{\mathbb{Q}}_t\left(R_T\right) - \mathrm{cov}_t\left(\mathcal{M}_T R_T, R_T\right)$$

where $\mathcal{M}_T$ is the pricing kernel that prices time T payoffs from the perspective of time t. Hence, if R^{sp}_T is the return on the S&P 500 index and the second component of the decomposition, $\mathrm{cov}_t\left(\mathcal{M}_T R^{sp}_T, R^{sp}_T\right)$, is negative, then the risk-neutral variance of return, $\frac{1}{R_{f,t}}\mathrm{var}^{\mathbb{Q}}_t\left(R^{sp}_T\right)$, gives the lower bound on the equity premium. Martin (2016) also argues that $\text{cov}_t\left(\mathcal{M}_T R_T^{sp}, R_T^{sp}\right)$ of the equity index is negative in most of macro-finance asset pricing models and estimates of covariance $\text{cov}(\mathcal{M}_T R_T^{sp}, R_T^{sp})$ are negative across various sample periods. As a consequence, the measure of the risk-neutral variance constructed from S&P 500 index options can serve as the lower bound on the equity premium. On the other hand, the *negative* expected bond returns from $\mathcal{M}_C$ imply that $\text{cov}_t\left(\mathcal{M}_{C,T} R_T^{(n)}, R_T^{(n)}\right)$ for n-year bond return $R_T^{(n)}$ can take a positive value even after explicitly incorporating $\text{var}_t^{\mathbb{Q}}\left(R_T\right)$ into the stochastic discount factor. Hence, in a desirable equilibrium asset pricing model, $\text{cov}_t\left(\mathcal{M}_T R_T, R_T\right)$ for bonds should be able to switch its sign while $\text{cov}_t\left(\mathcal{M}_T R_T, R_T\right)$ for the equity market remains negative. Given a preference of representative agent, for example, this condition gives a clue for the factor structure of the state. Alternatively, it can be utilized to restrict the parameter space of an equilibrium model.

9 Conclusion

This paper studies the impact of variance risk in the Treasury market on both term premia and the shape of the yield curve. Variance risk in the Treasury market can be observed via a portfolio of options given the assumptions that (i) the state of the economy is determined by a state variable following an affine diffusion process under the risk-neutral measure and (ii) the drift of a pricing kernel is affine in the state variable. This unique approach for the identification of variance risk in interest rates enables me to treat the bond VIX^2 as a measure of fixed income variance risk.

Using the observable proxy of variance risk, labeled bond VIX^2, this paper first proposes a novel return-forecasting factor. The return-forecasting factor is motivated by leading consumption-based asset pricing models, the long-run risk, rare disaster, and affine$^{\mathbb{Q}}$ habit formation models. In these frameworks, the set of risk factors underlying variation in risk premia is the sole source of time-varying variances in bond yields and should be captured by the bond VIX^2s. Projection of realized excess bond returns onto the space of VIX^2s gives a single return-forecasting factor that describes time-variation in the expected bond returns. The return-forecasting factor predicts excess returns of relatively short-term bonds well, and complements the Cochrane-Piazzesi factor.

Second, the observable measure of variance risk can be utilized to analyze the relationship between variance risk and the shape of the yield curve in a simple and parsimonious way. Its marginal impact on the cross-section of bonds is limited once I control for standard term structure factors. Furthermore, the hypothesis of unspanned stochastic volatility can be directly assessed by testing the spanning conditions of affine models, and I find the hypothesis is rejected. In sum, though the knife-edge conditions for USV effects do not hold in the data, it is true that identifying variance risk from the yield curve movements is very hard, and the variance risk can be effectively considered as unspanned risk.

Third, I propose a new representation of affine dynamic term structure models with time-varying variance risks in yields. Due to the observable proxy of variance risk, affine bond pricing models can be represented by observable and tradable factors. This simplifies the estimation procedure significantly while the information in volatility-sensitive instruments is readily incorporated. The estimated risk premia show that it is important to take into account both types of unspanned risk: the unspanned stochastic volatility factor and the hidden non-volatility factor.

References

Adrian, T., and H. S. Shin (2010): "Liquidity and Leverage," *Journal of Financial Intermediation*, 19, 418–437.

Ahn, D.-H., R. F. Dittmar, and A. R. Gallant (2002): "Quadratic Term Structure Models: Theory and Evidence," *Review of Financial Studies*, 15(1), 243–288.

Almeida, C., J. Graveline, and S. Joslin (2011): "Do Interest Rate Options Contain Information About Excess Bond Returns?," *Journal of Econometrics*, 164, 35–44.

Amengual, D., and D. Xiu (2014): "Resolution of Policy Uncertainty and Sudden Declines in Volatility," Working paper, CEMFI.

Andersen, T. G., 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.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006): "The Cross-Section of Volatility and Expected Returns," *Journal of Finance*, 61, 259 – 299.

Bansal, R., and I. Shaliastovich (2013): "A Long-Run Risks Explanation of Predictability Puzzles in Bond and Currency Markets," *Review of Financial Studies*, 26, 1–33.

Bansal, R., and A. Yaron (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59, 1481–1509.

Bao, J., J. Pan, and J. Wang (2011): "The Illiquidity of Corporate Bonds," *Journal of Finance*, 66, 911–946.

Bauer, M. D., and G. D. Rudebusch (2016): "Resolving the Spanning Puzzle in Macro-Finance Term Structure Models," *forthcoming, Review of Finance*.

Bean, C. (2005): "Commentary: separating the business cycle from other economic fluctuations," in *Proceedings-Economic Policy Symposium-Jackson Hole*, no. Aug, pp. 181–186. Federal Reserve Bank of Kansas City.

Bekaert, G., M. Hoerova, and M. L. Duca (2013): "Risk, Uncertainty and Monetary Policy," *Journal of Monetary Economics*, 60, 771–788.

Bibkov, R., and M. Chernov (2009): "Unspanned Stochastic Volatility in Affine Models: Evidence from EurodollarFutures and Options," *Management Science*, 55, 1292–1305.

——— (2011): "Yield Curve and Volatility: Lessons from Eurodollar Futures and Options," *Journal of Financial Econometrics*, 9, 66–105.

Björk, T. (2009): *Arbitrage Theory in Continuous Time*. Oxford University Press, 3rd edn.

Bollerslev, T. G., G. Tauchen, and H. Zhou (2009): "Expected Stock Returns and Variance Risk Premia," *Review of Financial Studies*, 22, 4463–4492.

Breeden, D., and R. H. Litzenberger (1978): "Prices of State-Contingent Claims Implicit in Option Prices," *Journal of Business*, 51, 621–651.

Brunnermeier, M., S. Nagel, and L. H. Pedersen (2009): "Carry Trades and Currency Crashes," *NBER Macroeconomics Annual 2008*, 23, 313–347.

CAMBPELL, J. Y., AND R. J. SHILLER (1991): “Yield Spreads and Interest Rate Movements: A Bird’s Eye View,” *Review of Economic Studies*, 58, 495–514.

CAMPBELL, J. Y. (1993): “Intertemporal Asset Pricing without Consumption Data,” *The American Economic Review*, 83(3), pp. 487–512.

CAMPBELL, J. Y., AND J. H. COCHRANE (1999): “By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior,” *Journal of Political Economy*, 107, 205–251.

CARR, P., AND D. MADAN (1998): “Towards a Theory of Volatility Trading,” in *Volatility: New Estimation Techniques for Pricing Derivatives*, ed. by R. Jarrow, pp. 417–427. RISK Publications.

CARR, P., AND L. WU (2009): “Variance Risk Premiums,” *Review of Financial Studies*, 22(3), 1311–1341.

CHEN, R.-R., AND L. SCOTT (1993): “Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates,” *The Journal of Fixed Income*, 3, 14–31.

CHERIDITO, P., D. FILIPOVIĆ, AND R. L. KIMMEL (2007): “Market price of risk specifications for affine models: Theory and evidence,” *Journal of Financial Economics*, 83(1), 123 – 170.

CHERNOV, M., AND P. MUELLER (2012): “The Term Structure of Inflation Expectations,” *Journal of Financial Economics*, 106, 367–394.

CHOI, H., P. MUELLER, AND A. VEDOLIN (2016): “Bond Variance Risk Premiums,” *forthcoming, Review of Finance.*

COCHRANE, J. H., AND M. PIAZZESI (2005): “Bond Risk Premia,” *American Economic Review*, 95, 138–160.

——— (2008): “Decomposing the Yield Curve,” Working Paper, Chicago Booth.

COLLIN-DUFRESNE, P., AND R. S. GOLDSTEIN (2002): “Do Bonds Span the Fixed-Income Markets?,” *Journal of Finance*, 57, 1685—1730.

COLLIN-DUFRESNE, P., R. S. GOLDSTEIN, AND C. S. JONES (2009): “Can interest rate volatility be extracted from the cross section of bond yields?,” *Journal of Financial Economics*, 94, 47–66.

CREAL, D. D., AND J. C. WU (2015): “Estimation of Affine Term Structure Models with Spanned or Unspanned Stochastic Volatility,” *Journal of Econometrics*, 185(1), 60–81.

DAI, Q., AND K. SINGLETON (2002): “Expectation Puzzles, Time-Varying Risk Premia, and Affine Models of the Term Structure,” *Journal of Financial Economics*, 63, 415–441.

DAI, Q., AND K. J. SINGLETON (2000): “Specification Analysis of Affine Term Structure Models,” *Journal of Finance*, 55(5), 1943–1978.

DRECHSLER, I., AND A. YARON (2011): “What’s Vol Got to Do With It,” *Review of Financial Studies*, 24, 1–45.

DUARTE, J. (2004): “Evaluating an Alternative Risk Preference in Affine Term Structure Models,” *The Review of Financial Studies*, 17(2), pp. 379–404.

Duffee, G. R. (2010): "Sharpe Ratios in Term Structure Models," Working Paper, Johns Hopkins University.

——— (2011a): "Forecasting with the Term Structure: the Role of No-arbitrage Restrictions," Working Paper, Johns Hopkins University.

——— (2011b): "Information in (and not in) the Term Structure," *Review of Financial Studies*, 24(9), 2895–2934.

Duffie, D., and L. G. Epstein (1992): "Asset Pricing with Stochastic Differential Utility," *Review of Financial Studies*, 5(3), 411–436.

Duffie, D., D. Filipović, and W. Schachermayer (2003): "Affine Processes and Applications in Finance," *The Annals of Applied Probability*, 13(3), 984–1053.

Duffie, D., and R. Kan (1996): "A yield-factor model of interest rates," *Mathematical finance*, 6, 379–406.

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

Gourieroux, C., and J. Jasiak (2006): "Autoregressive gamma processes," *Journal of Forecasting*, 25(2), 129–152.

Gürkaynak, R., B. Sack, and J. Wright (2007): "The U.S. Treasury Yield Curve: 1961 to the Present," *Journal of Monetary Economics*, 54, 2291–2304.

Hamilton, J. D., and J. C. Wu (2012): "Identification and Estimation of Gaussian Affine Term Structure Models," *Journal of Econometrics*, 168(2), 315 – 331.

Jagannathan, R., A. Kaplin, and S. Sun (2003): "An evaluation of multi-factor CIR models using LIBOR, swap rates, and cap and swaption prices," *Journal of Econometrics*, 116, 113 – 146.

Joslin, S. (2014): "Pricing and Hedging Volatility Risk in Fixed Income Markets," Working Paper, USC.

——— (2015): "Can Unspanned Stochastic Volatility Models Explain the Cross Section of Bond Volatilities?," Working Paper, USC.

Joslin, S., and A. Le (2014): "Interest Rate Volatility and No-Arbitrage Affine Term Structure Models," Working Paper, USC.

Joslin, S., A. Le, and K. J. Singleton (2013): "Why Gaussian Macro-Finance Term Structure Models Are (Nearly) Unconstrained Factor-VARs," *Journal of Financial Economics*, 109, 604 – 622.

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

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

Kelly, B., L. Pástor, and P. Veronesi (2016): "The Price of Political Uncertainty: Theory and Evidence From the Option Market," *forthcoming, Journal of Finance*.

Kim, D. H. (2007): "Spanned stochastic volatility in bond markets: a reexamination of the relative pricing between bonds and bond options," Unpublished working paper 239. BIS, Basel, Switzerland.

Kleibergen, F., and R. Paap (2006): "Generalized reduced rank tests using the singular value decomposition," *Journal of Econometrics*, 133(1), 97 – 126.

Le, A., and K. J. Singleton (2013): "The Structure of Risks in Equilibrium Affine Models of Bond Yields," Working Paper, UNC.

Le, A., K. J. Singleton, and Q. Dai (2010): "Discrete-Time Affine Term Structure Models with Generalized Market Prices of Risk," *Review of Financial Studies*, 23(5), 2184–2227.

Ludvigson, S. C., and S. Ng (2009): "Macro Factors in Bond Risk Premia," *Review of Financial Studies*, 22, 5027–5067.

Martin, I. (2016): "What is the Expected Return on the Market?," *forthcoming, Quarterly Journal of Economics*.

Mele, A., and Y. Obayashi (2013): "Volatility Indexes and Contracts for Government Bonds and Time Deposits," Working Paper, Applied Academics LLC.

Nagel, S. (2012): "Evaporating liquidity," *Review of Financial Studies*, 25, 2005–2039.

Newey, W. K., and K. D. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

Orphanides, A. (2003): "Historical monetary policy analysis and the Taylor rule," *Journal of Monetary Economics*, 50(5), 983 – 1022.

Pearson, N. D., and T.-S. Sun (1994): "Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model," *Journal of Finance*, 49(4), 1279–1304.

Piazzesi, M. (2010): "Affine Term Structure Models," in *Handbook of Financial Econometrics: Tools and Techniques*, ed. by Y. Aït-Sahalia, and L. Hansen, vol. 1 of *Handbooks in Finance*, pp. 691 – 766. North-Holland, San Diego.

Rudebusch, G. D. (2006): "Monetary Policy Inertia: Fact or Fiction?," *International Journal of Central Banking*, 2(4), 88–135.

Taylor, J. B. (1993): "Discretion versus policy rules in practice," *Carnegie-Rochester Conference Series on Public Policy*, 39(1), 195–214.

——— (1999): "A Historical Analysis of Monetary Policy Rules," in *Monetary Policy Rules*, NBER Chapters, pp. 319–348. National Bureau of Economic Research, Inc.

Trolle, A. B., and E. S. Schwartz (2009): "A General Stochastic Volatility Model for the Pricing of Interest Rate Derivatives," *Review of Financial Studies*, 22, 2007–2057.

Tsai, J. (2016): "Rare Disasters and the Term Structure of Interest Rates," Working Paper, University of Oxford.

Wachter, J. (2006): "A Consumption-Based Model of the Term Structure of Interest Rates," *Journal of Financial Economics*, 79, 365–399.

Wachter, J. A. (2013): "Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?," *Journal of Finance*, 68(3), 987–1035.

Zhou, G., and Y. Zhu (2015): "Macroeconomic Volatilities and the Long-run Risks of Asset Prices," *Management Science*, 61, 413 – 430.

Table 1
Predictive Regressions

Panel A reports adjusted R^2 from regressing twelve-month excess returns rx(n) of bonds with n years to maturity on CP factor, CIV, and both. Panel B presents estimated coefficients from predictive regressions from the mean of excess returns in Panel A onto CP factor or (and) CIV. Standard errors are in parentheses and adjusted according to Newey and West (1987). Data is monthly and runs from October 1990 to December 2007.

Panel A: Predictive Regressions and Adjusted R^2s

Excess returns from Fama-Bliss data				Excess returns from GSW data			
	CP	CIV	$CIV\&CP$		CP	CIV	$CIV\&CP$
rx(2)	0.20	0.27	0.36	rx(2)	0.15	0.26	0.31
rx(3)	0.22	0.28	0.38	rx(4)	0.21	0.28	0.36
rx(4)	0.24	0.28	0.40	rx(6)	0.25	0.27	0.38
rx(5)	0.22	0.28	0.38	rx(8)	0.27	0.26	0.39
				rx(10)	0.28	0.23	0.38

Panel B: Predictive Regression Coefficients

	CIV	CP	adj R^2
mean(FB rx)	0.53		0.28
	(0.10)		
mean(FB rx)	0.35	0.43	0.39
	(0.10)	(0.12)	
mean(GSW rx)	0.52		0.27
	(0.11)		
mean(GSW rx)	0.38	0.39	0.39
	(0.11)	(0.12)	

Table 2
Marginal Impact of Variance Risks onto the Shape of the Yield Curve

This table reports the test statistics of the likelihood ratio test. The first column presents the right hand side variables in the restricted models: equation (7). The remaining columns present an additional variable in each version of the unrestricted model and its corresponding LR statistics. The last column shows the 5% critical value of the test statistics, which follows $\chi^2(11)$ distribution. Data is monthly and runs from October 1990 to December 2007.

Restricted Model	Additional Variable in Unrestricted model				
	VPC1	VPC2	PC3	PC4	C.V.(5%)
PC1, PC2	4.2	14.2	5882.9		19.7
PC1, PC2, PC3	45.0	18.6		3678.7	19.7

Table 3
Tests for the USV effect

This table reports the test statistics of Kleibergen and Paap (2006) for the rank of the matrix $\mathcal{B}_{\mathcal{Z}}$ in equation (6). The null of the test is equation (9). Data is monthly and runs from October 1990 to December 2007.

Specification	Stat.	d.f.	C.V.(5%)	p-val
$A_1\,(3)$	20.82	9	16.92	0.01
$A_1\,(4)$	20.72	8	15.51	0.01
$A_2\,(4)$	14.90	8	15.51	0.06
$A_2\,(5)$	14.89	7	14.07	0.04

Table 4
Campbell-Shiller Regressions

Panel A reports the coefficients $\phi_1^{(n)}$ from the Campbell-Shiller regression in equation (14) with three month changes in yields as regressands. GRO and CPI are used to estimate $UMA_m^R(N)$ models. Panel B reports the coefficients $\phi_1^{(n)}$ from the Campbell-Shiller regression with six month changes in yields as regressands. $UGAP$ and CPI are used to estimate $UMA_m^R(N)$ models. Data is monthly and runs from October 1990 to December 2007.

Panel A: Campbell-Shiller Regression with Three-month Changes in Yields

Specification	Macro	Maturities										MSE
		1 yr	2 yr	3 yr	4 yr	5 yr	6 yr	7 yr	8 yr	9 yr	10 yr	
Data		0.40	-0.25	-0.67	-0.90	-1.03	-1.10	-1.15	-1.18	-1.20	-1.23	0.00
$A_0(3)$		0.40	0.01	-0.23	-0.37	-0.46	-0.53	-0.59	-0.64	-0.69	-0.74	2.31
$A_1(4)$		0.26	-0.29	-0.64	-0.84	-0.96	-1.07	-1.16	-1.25	-1.33	-1.41	0.09
$UMA_0^2(5)$	GRO,CPI	0.96	0.64	0.38	0.21	0.09	0.00	-0.06	-0.12	-0.17	-0.21	10.28
$UMA_1^2(6)$	GRO,CPI	0.96	0.60	0.27	0.04	-0.13	-0.25	-0.35	-0.43	-0.50	-0.57	6.45
$UMA_0^2(5)$	UGAP,CPI	0.56	0.16	-0.09	-0.25	-0.35	-0.42	-0.48	-0.54	-0.59	-0.64	3.44
$UMA_1^2(6)$	UGAP,CPI	0.33	-0.11	-0.40	-0.56	-0.67	-0.75	-0.83	-0.89	-0.96	-1.02	0.74

Panel B: Campbell-Shiller Regression with Six-month Changes in Yields

Specification	Macro	Maturities										MSE
		1 yr	2 yr	3 yr	4 yr	5 yr	6 yr	7 yr	8 yr	9 yr	10 yr	
Data		0.34	-0.39	-0.78	-0.98	-1.10	-1.17	-1.23	-1.27	-1.31	-1.34	0.00
$A_0(3)$		0.30	-0.03	-0.21	-0.30	-0.36	-0.41	-0.46	-0.50	-0.54	-0.59	4.36
$A_1(4)$		0.13	-0.37	-0.64	-0.78	-0.87	-0.96	-1.03	-1.11	-1.19	-1.26	0.29
$UMA_0^2(5)$	GRO,CPI	1.11	0.84	0.62	0.47	0.36	0.27	0.21	0.15	0.10	0.06	18.41
$UMA_1^2(6)$	GRO,CPI	1.06	0.76	0.46	0.25	0.10	-0.01	-0.11	-0.18	-0.25	-0.31	12.29
$UMA_0^2(5)$	UGAP,CPI	0.48	0.13	-0.07	-0.18	-0.24	-0.28	-0.33	-0.36	-0.40	-0.43	6.25
$UMA_1^2(6)$	UGAP,CPI	0.23	-0.18	-0.40	-0.51	-0.58	-0.64	-0.69	-0.75	-0.80	-0.85	2.04

Table 5
Risk Premia Parameters

This table presents the maximum likelihood estimation of the parameters Λ_0 and Λ_1 governing expected excess returns on the PC-mimicking portfolios. Standard errors are given in parentheses.

$\mathcal{P}$	const	$PC1$	$PC2$	$PC3$	$UGAP$	CPI	TIV^2_{30yr}
$PC1$	0	-0.636	-1.305	0	0	1.647	1.509
		(0.213)	(0.326)			(0.650)	(0.683)
$PC2$	-0.018	0	0	0	0	0.699	0
	(0.007)					(0.265)	
$PC3$	0	0	0	0	0	0	0

Table 6
Predictive Regressions

This table reports adjusted R^2 from regressing twelve-month excess returns xr(n) of bonds with n years to maturity on corresponding model implied expected excess returns.

Model	Specification	Macro	R^2			
			1 year	5 year	7 year	10 year
$\mathcal{M}_C$	$UMA^2_1(6)$	UGAP,CPI	0.26	0.28	0.29	0.30
$\mathcal{M}_J$	$UMA^2_0(5)$	GRO,CPI	0.38	0.22	0.18	0.14

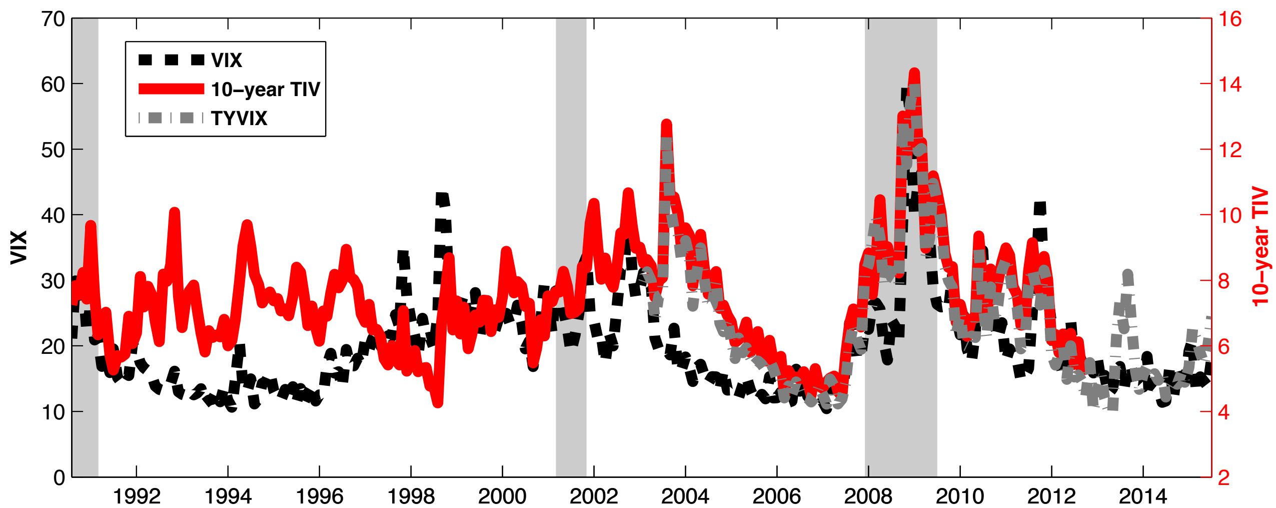

Figure 1. TIV,TYVIX and VIX

This figure plots the CBOE VIX, the CBOE TYVIX and the 10-year TIV from **?**. Volatilities are the square root from variances as constructed using option prices via equation (3). Numbers are annualized and expressed in percent. Gray bars indicate NBER recessions. The data is monthly and runs from July 1990 to June 2015; The 10-year TIV data ends in August 2012, and the TYVIX data starts from January 2003.

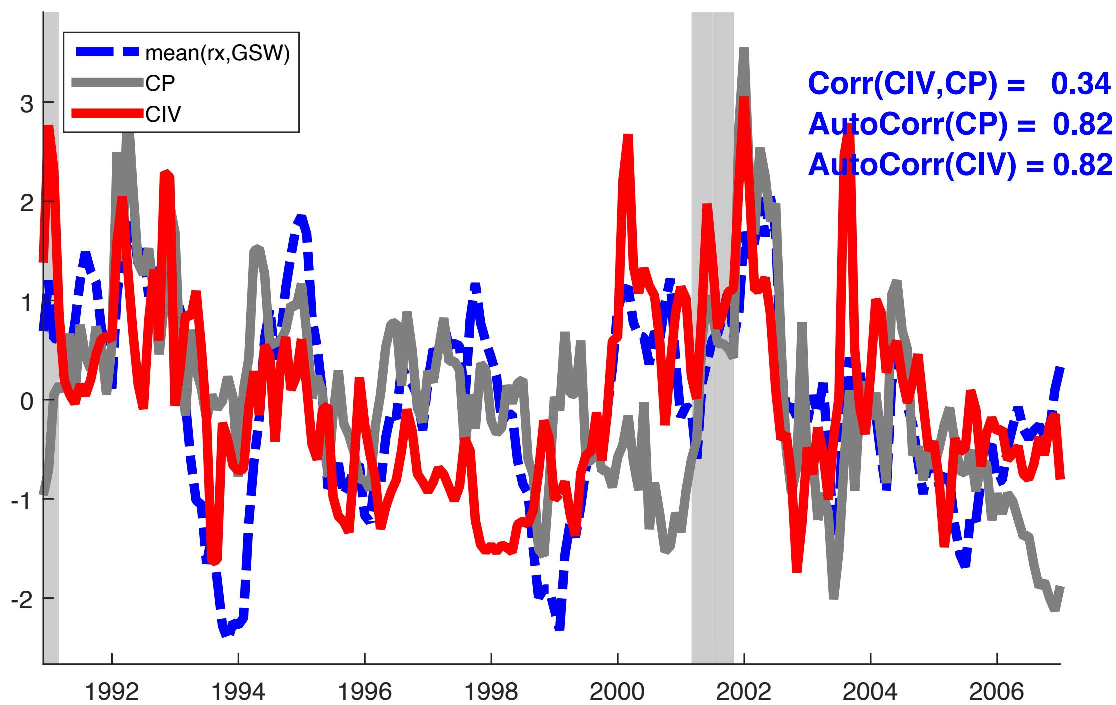

Figure 2. Excess Returns, CP and CIV

This figure plots the Cochrane-Piazzesi factor (CP), CIV, and the average of twelve-month excess returns on bonds with maturities of 1 through 10 years. Gray bars indicate NBER recessions, and blue bars represent financial crisis periods. The data is monthly and runs from October 1990 to December 2007.

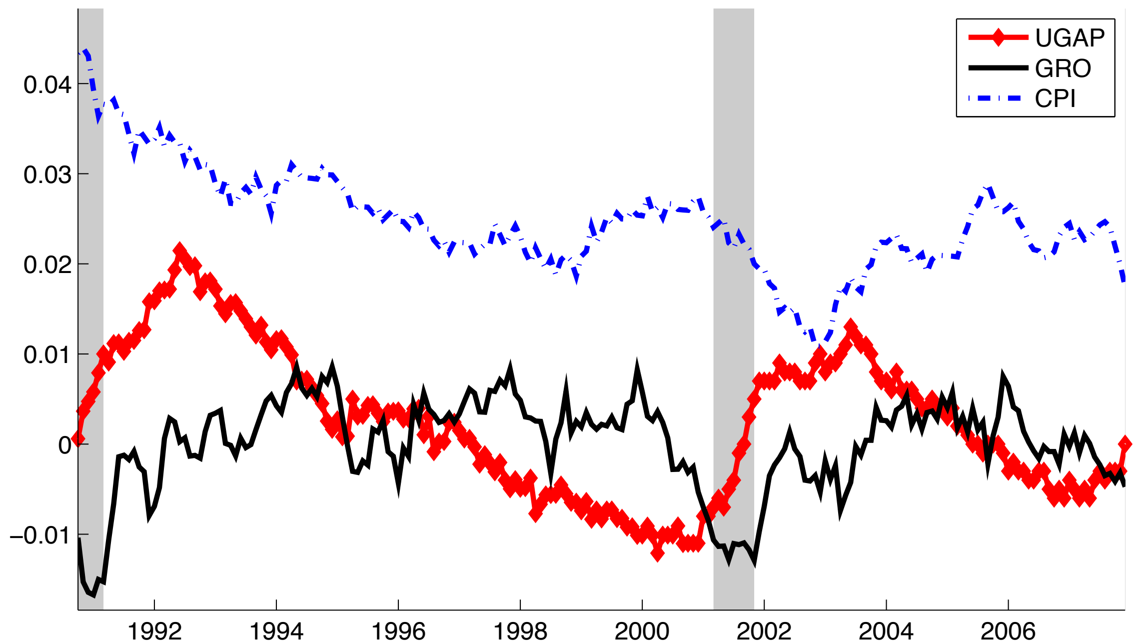

Figure 3. UGAP, GRO and CPI

This figure plots the unemployment gap ($UGAP$), the three-month moving average of the Chicago Fed's National Activity Index (GRO) and the year-over-year growth in Consumer Price Index, excluding food prices and energy prices (CPI). Gray bars indicate NBER recessions. The data is monthly and runs from October 1990 to December 2007.

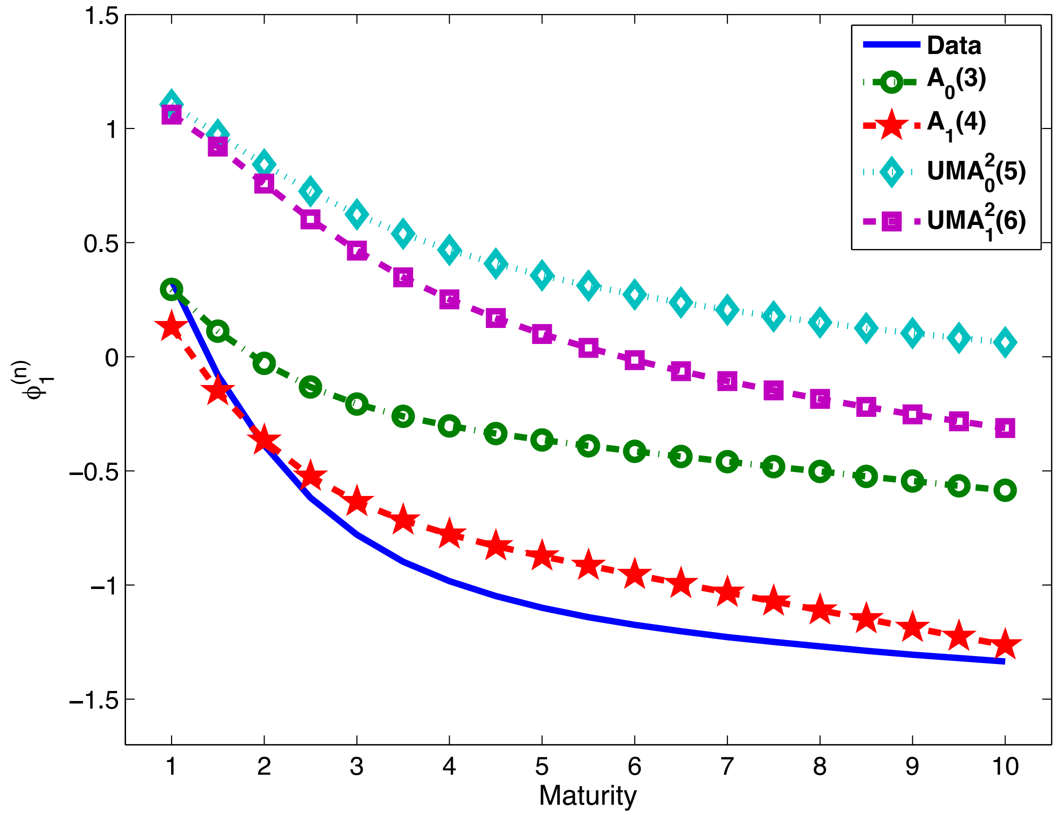

Figure 4. Campbell-Shiller Regression

This figure plots the coefficients $\phi_1^{(n)}$ from the Campbell-Shiller regression in equation (14) where the left-hand side variable is changes in yields over six months. The models $A_m(N)$ are models with $N - m$ Gaussian factors and m factors driving volatility. The models $UMA_m^R(N)$ are models with GRO and CPI as unspanned macro risks. Data is monthly and runs from October 1990 to December 2007.

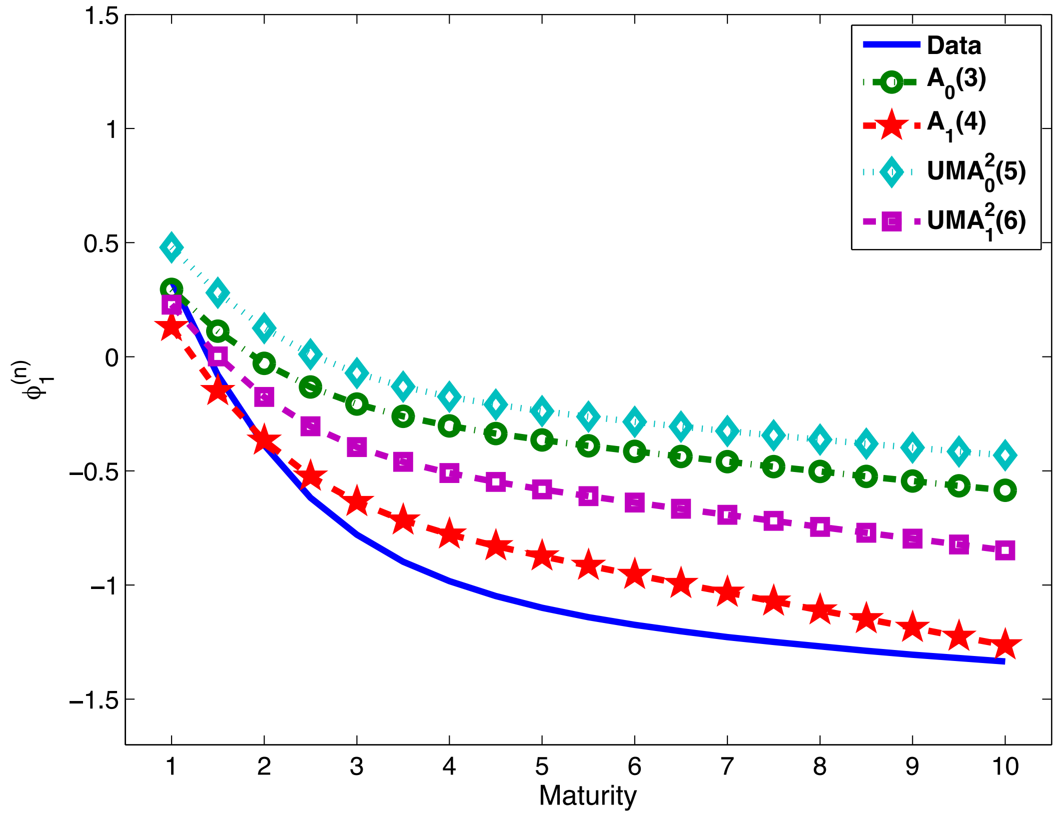

Figure 5. Campbell-Shiller Regression

This figure plots the coefficients $\phi_1^{(n)}$ from the Campbell-Shiller regression in equation (14) where the left-hand side variable is changes in yields over six months. The models $A_m(N)$ are models with $N - m$ Gaussian factors and m factors driving volatility. The models $UMA_m^R(N)$ are models with $UGAP$ and CPI as unspanned macro risks. Data is monthly and runs from October 1990 to December 2007.

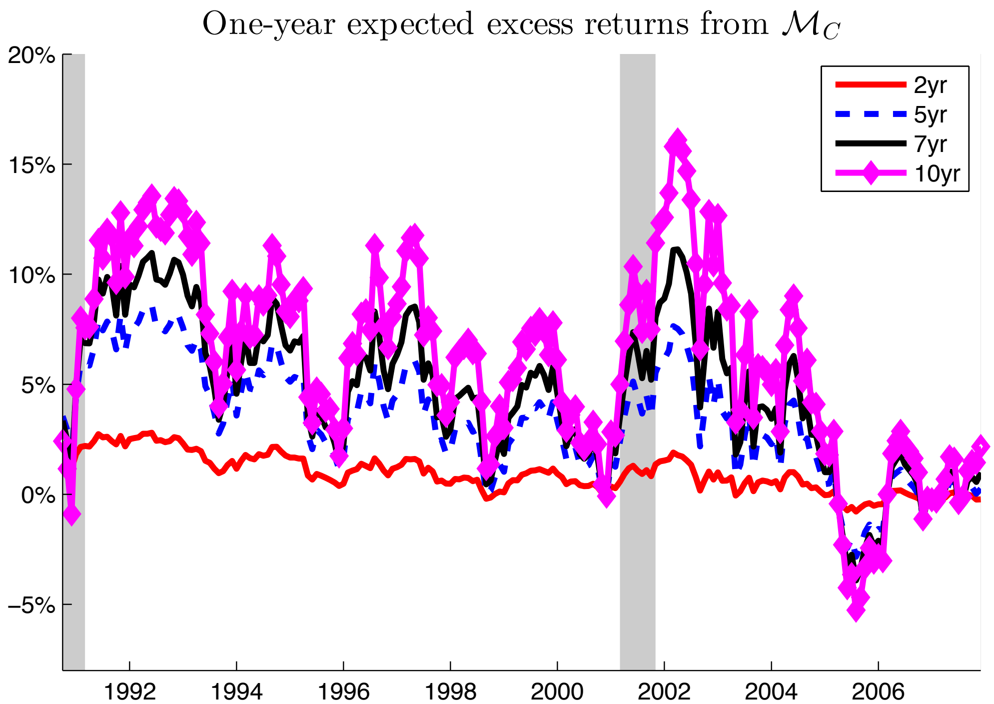

Figure 6. One-year Expected Excess Returns from $\mathcal{M}_C$

This figure plots the estimates of one-year simple expected excess returns from $\mathcal{M}_C$. The constraints on risk premia dynamics in Section 8 are imposed. The data is monthly and runs from October 1990 to December 2007.

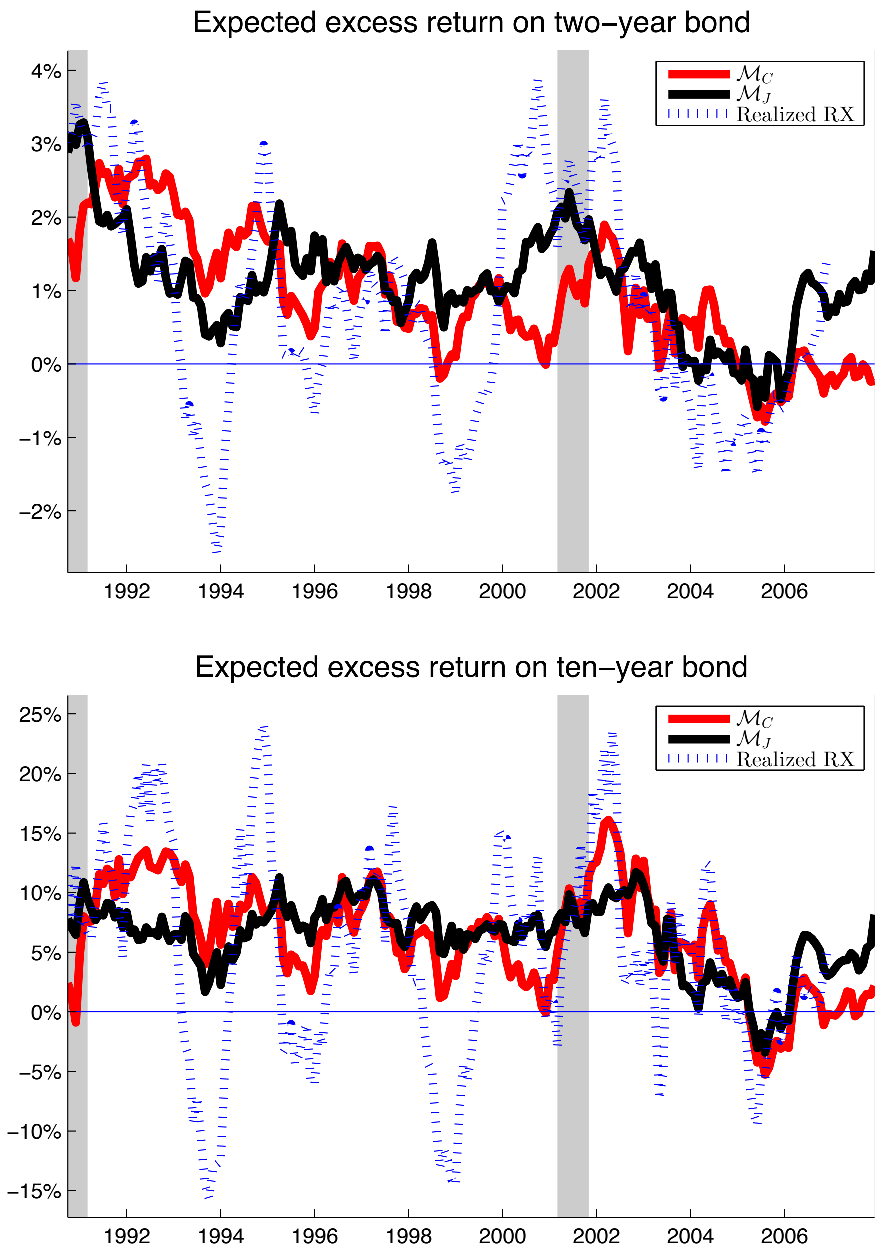

Figure 7. One-year Expected Excess Returns from $\mathcal{M}_C$ and $\mathcal{M}_J$

This figure plots the estimates of one-year simple expected excess returns from $\mathcal{M}_C$ and $\mathcal{M}_J$. The constraints on risk premia dynamics in Section 8 are imposed to estimate each specification. The data is monthly and runs from October 1990 to December 2007.

Appendix A Proof of Proposition 1

Lemma 1. *Suppose that $F_t(T, \mathbb{T})$ is a diffusion process*

$$\frac{dF_t(T, \mathbb{T})}{F_t(T, \mathbb{T})} = v_t' dB_t^{\mathbb{Q}_T}$$

Then,

$$E_t^{\mathbb{Q}_T}\left[\int_t^T v_s v_s' ds\right] = \frac{2}{P_{t,T}}\left[\int_0^{F_t(T,\mathbb{T})} \frac{Put_t(K,T,\mathbb{T})}{K^2} dK + \int_{F_t(T,\mathbb{T})}^{\infty} \frac{Call_t(K,T,\mathbb{T})}{K^2} dK\right]$$

Proof. See for example Choi, Mueller, and Vedolin (2016). □

Lemma 2. *Suppose that the numeraire of $\mathbb{Q}_T$ forward measure, $P_{t,T}$, has the following dynamics under $\mathbb{Q}$*

$$dP_{t,T} = P_{t,T}\left(r_t dt + \sigma_{t,T}' dB_t^{\mathbb{Q}}\right) \tag{A-1}$$

Then,

$$dB_t^{\mathbb{Q}_T} = -\sigma_{t,T} dt + dB_t^{\mathbb{Q}}$$

Proof. See for example Björk (2009). □

Proof of Proposition 1. Under the two assumptions that (i) the state follows affine diffusion and (ii) the short rate is an affine function of the state, Duffie and Kan (1996) show that $P_{t,T}$ is exponentially affine in the state.

$$P_{t,T} = \exp\left[\bar{A}\left(\Theta^{\mathbb{Q}}, t, T\right) + \bar{B}\left(\Theta^{\mathbb{Q}}, t, T\right)' Z_t\right] \tag{A-2}$$

Since no-arbitrage condition implies $F_t = P_{t,\mathbb{T}}/P_{t,T}$, the forward price is also exponentially affine in the state

$$F_t(T, \mathbb{T}) = P_{t,\mathbb{T}}/P_{t,T} = \exp\left[\psi_0 + \psi_1' Z_t\right]$$

where ψ_0 and ψ_1 are functions of $\left\{\Theta^{\mathbb{Q}}, T, \mathbb{T}\right\}$. Applying Ito's lemma to F_t under $\mathbb{Q}_T$, we have the following diffusion process

$$\frac{dF_t(T, \mathbb{T})}{F_t(T, \mathbb{T})} \equiv v_t' dB_t^{\mathbb{Q}_T} = \psi_1' \Sigma_{Z,t} dB_t^{\mathbb{Q}_T}$$

The expected quadratic variation of the forward up to time T is

$$E_t^{\mathbb{Q}_T}\left[\int_t^T v_s' v_s ds\right] = \int_t^T \psi_1' \Sigma_{Z0} \psi_1 ds + \sum_{i=1}^m \int_t^T \psi_1' \Sigma_{Zi} \psi_1 E_t^{\mathbb{Q}_T}\left[V_{is}\right] ds \tag{A-3}$$

The proof is completed by Lemma 1 once the right-hand side of equation (A-3) is an affine function of V_t only. Intuitively, the Gaussian factor X_t cannot affect $E_t^{\mathbb{Q}_T}[V_{is}]$, otherwise the expected value of V_s under $\mathbb{Q}_T$ could have a negative value. As a last step, apply Ito's lemma to equation (A-2). Then, the diffusion term of it, the $\sigma_{t,T}$ term in case of equation (A-1), is given by

$$\sigma_{t,T} = \Sigma_{Z,t} \bar{B}\left(\Theta^{\mathbb{Q}}, t, T\right)$$

As a consequence of Lemma 2, the dynamics of the state under $\mathbb{Q}_T$ can be written as

$$dZ_t = d\left(X_t', V_t'\right)' = \left[\left(\mu_{X,t}^{\mathbb{Q}\prime}, \mu_{V,t}^{\mathbb{Q}\prime}\right)' + \Sigma_{Z,t}\Sigma_{Z,t}'\bar{B}\left(\Theta^{\mathbb{Q}}, t, T\right)\right]dt + \Sigma_{Z,t}dB^{\mathbb{Q}_T}$$

Since $\mu_{V,t}^{\mathbb{Q}}$ is assumed to be affine in V_t only (i.e. as in Duffie, Filipović, and Schachermayer (2003), K_{1VX} in equation (1) is set to be zero for admissibility) and $\Sigma_{Z,t}\Sigma_{Z,t}'\bar{B}\left(\Theta^{\mathbb{Q}}, t, T\right)$ is a linear function of V_t solely, $\mu_{V,t}^{\mathbb{Q}_T}$ is also an affine function of V_t. □

Appendix B The bond VIX^{2}s in a long-run risk framework

I initially solve the model of Bansal and Shaliastovich (2013) in the continuous-time framework, and then demonstrate the linear mapping between the bond VIXs and the two macroeconomic uncertainties in the model. The dynamics of consumption C_t, inflation π_t ,and their long-run risks as well as quantity of risk are specified as

$$\begin{aligned}
\frac{dC_t}{C_t} &= (\mu_c + X_{ct})\,dt + \sigma_c dZ_{1t} \\
\frac{d\pi_t}{\pi_t} &= (\mu_\pi + X_{\pi t})\,dt + \sigma_\pi dZ_{2t} \\
dX_{ct} &= (-\rho_c X_{ct} - \rho_{c\pi} X_{\pi t})\,dt + \sqrt{V_{ct}}dW_{1t} \\
dX_{\pi t} &= -\rho_\pi X_{\pi t}dt + \sqrt{V_{\pi t}}dW_{2t} \\
V_{ct} &= \kappa_c\left(\bar{V}_c - V_{ct}\right)dt + w_c\sqrt{V_{ct}}dB_{1t} \\
V_{\pi t} &= \kappa_\pi\left(\bar{V}_\pi - V_{\pi t}\right)dt + w_\pi\sqrt{V_{\pi t}}dB_{2t}
\end{aligned}$$

where $Z_{1t}, Z_{2t}, W_{1t}, W_{2t}, B_{1t}$ and B_{2t} are independent Brownian motions. Following Duffie and Epstein (1992), the representative agent's objective is

$$J_t = \max_{\{C_s\}} E_t\left[\int_t^T f\left(C_s, J_s\right)ds\right]$$

where the normalized aggregator $f\left(C, J\right)$ is given by

$$f\left(C, J\right) = \beta\left(\frac{1-\gamma}{1-1/\psi}\right)J\left[\left(\frac{C}{\left(\left(1-\gamma\right)J\right)^{1/(1-\gamma)}}\right)^{1-1/\psi} - 1\right] \tag{A-4}$$

with β the rate of time preference, γ the relative risk version, and ψ the elasticity of intertemporal substitution. Conjecture J as

$$J\left(W_t, Z_t\right) = \exp\left(a_0 + a_1X_{ct} + a_2X_{\pi t} + a_3V_{ct} + a_4V_{\pi t}\right)\frac{W_t^{1-\gamma}}{1-\gamma}. \tag{A-5}$$

Since the envelop condition, $f_C = J_W$, can be written as

$$C = J_W^{-\psi}\left[\left(1-\gamma\right)J\right]^{\frac{1-\gamma\psi}{1-\gamma}}\beta^{\psi}, \tag{A-6}$$

substituting equation (A-5) into equation (A-6) enables us to express the log consumption-wealth ratio in terms of the state variables:

$$\log\left(\frac{C_t}{W_t}\right) = \psi \log \beta + \frac{1-\psi}{1-\gamma}\left(a_0 + a_1 X_{ct} + a_2 X_{\pi t} + a_3 V_{ct} + a_4 V_{\pi t}\right)$$

In addition, substituting equation (A-6) into equation (A-4) gives

$$f = \left(\frac{1-\gamma}{1-1/\psi}\right)\left(\frac{C_t}{W_t} - \beta\right) J$$

Applying log-linear approximation from Campbell (1993) to the consumption-wealth ratio

$$\frac{C_t}{W_t} \approx g_1 + g_1 \log g_1 + g_1 \log\left(\frac{C_t}{W_t}\right) \tag{A-7}$$

where g_1 is the long-term mean of the consumption-wealth ratio. Then

$$\begin{aligned} f &\approx \left(\frac{1-\gamma}{1-1/\psi}\right)\left[g_1 + g_1 \log g_1 + g_1 \log\left(\frac{C_t}{W_t}\right) - \beta\right] J \\ &= \left(\frac{1-\gamma}{1-1/\psi}\right)\left[\xi + \frac{g_1(1-\psi)}{1-\gamma}\left(a_0 + a_1 X_{ct} + a_2 X_{\pi t} + a_3 V_{ct} + a_4 V_{\pi t}\right)\right] J \end{aligned}$$

where $\xi = g_1 + g_1 \log g_1 + g_1 \psi \log \beta - \beta$. As shown in Duffie and Epstein (1992), the state price process is identified as $\zeta_t = \exp\left[\int_0^t f_J(C_s, J_s)\, ds\right] f_C(C_t, J_t)$ and the corresponding nominal pricing kernel is defined as $\tilde{\zeta}_t = \frac{\zeta_t}{\pi_t}$. Applying Ito's lemma for the nominal pricing kernel results in

$$\begin{aligned} \frac{d\tilde{\zeta}_t}{\tilde{\zeta}_t} &= -\left(r_0 + r_1 X_{ct} + r_2 X_{\pi t} + r_3 V_{ct} + r_4 V_{\pi t}\right) dt \\ &\quad -\lambda_{1t} dZ_{1t} - \lambda_{2t} dZ_{2t} - \lambda_{3t} dW_{1t} - \lambda_{4t} dW_{2t} - \lambda_{5t} dB_{1t} - \lambda_{6t} dB_{2t} \end{aligned}$$

where

$$\begin{aligned} \lambda_{1t} &= \gamma \sigma_c, & \lambda_{2t} &= \sigma_\pi \\ \lambda_{3t} &= -\frac{1-\gamma\psi}{1-\gamma} a_1 \sqrt{V_{ct}}, & \lambda_{4t} &= -\frac{1-\gamma\psi}{1-\gamma} a_2 \sqrt{V_{\pi t}} \\ \lambda_{5t} &= -\frac{1-\gamma\psi}{1-\gamma} a_3 w_c \sqrt{V_{ct}}, & \lambda_{6t} &= -\frac{1-\gamma\psi}{1-\gamma} a_4 w_\pi \sqrt{V_{\pi t}} \end{aligned}$$

and

$$
\begin{aligned}
r_0 &= -\xi_1 - \frac{1-\psi}{1-\gamma}\left(a_1\mu_\pi + a_3\kappa_c\bar{V}_c + a_4\kappa_\pi\bar{V}_\pi\right) + \gamma\mu_c - \frac{1}{2}\gamma\left(\gamma-1\right)\sigma_c^2 + \mu_\pi - \sigma_\pi^2 \\
r_1 &= \gamma + a_1\left(\rho_c + g_1\right)\frac{1-\gamma\psi}{1-\gamma} \\
r_2 &= 1 + \left(a_1\rho_{c\pi} + a_2\left(\rho_\pi + g_1\right)\right)\frac{1-\gamma\psi}{1-\gamma} \\
r_3 &= a_3\left(\kappa_c + g_1\right)\frac{1-\gamma\psi}{1-\gamma} - \frac{1}{2}\left(a_1^2 + a_3^2 w_c^2\right)\left(\frac{1-\gamma\psi}{1-\gamma}\right)^2 \\
r_4 &= a_4\left(\kappa_\pi + g_1\right)\frac{1-\gamma\psi}{1-\gamma} - \frac{1}{2}\left(a_2^2 + a_4^2 w_\pi^2\right)\left(\frac{1-\gamma\psi}{1-\gamma}\right)^2
\end{aligned}
$$

Because of the Girsanov theorem, the price of risk dynamics implies that the state variable also follows affine diffusion under $\mathbb{Q}$, and the short rate is affine in the state as can be seen.

Appendix C The bond VIX^{2}s and Gaussian quadratic term structure models

In Gaussian quadratic term structure models, the state variable X_t is assumed to follow the Ornstein-Uhlenbeck process under the risk-neutral measure $\mathbb{Q}$:

$$
dX_t = \left[K_{0X}^{\mathbb{Q}} + K_{1X}^{\mathbb{Q}} X_t\right] dt + \sqrt{\Omega} dB_t^{\mathbb{Q}}
$$

where $\sqrt{\Omega}$ represents the Cholesky decomposition of a positive definite matrix Ω. The short rate is a quadratic function of the Gaussian affine-diffusion state

$$
r_t = \Psi_0 + \Psi_1 X_t + X_t' \Psi_2 X_t
$$

Then, bond prices are represented as

$$
P_{t,T} = \exp\left[\tilde{A}\left(\Theta^{\mathbb{Q}}, t, T\right) + X_t'\tilde{B}\left(\Theta^{\mathbb{Q}}, t, T\right) + X_t'\tilde{C}\left(\Theta^{\mathbb{Q}}, t, T\right) X_t\right]
$$

with a symmetric matrix $\tilde{C}$ (see for example Ahn, Dittmar, and Gallant (2002)). Applying Ito's lemma, the dynamics of the forward are

$$
\frac{dF_t\left(T, \mathbb{T}\right)}{F_t\left(T, \mathbb{T}\right)} \equiv v_t' dB_t^{\mathbb{Q}_T} = \left[\xi_0 + 2\xi_1 X_t\right]' \sqrt{\Omega} dB_t^{\mathbb{Q}_T}
$$

where

$$
\begin{aligned}
\xi_0 &= \tilde{B}\left(\Theta^{\mathbb{Q}}, t, \mathbb{T}\right) - \tilde{B}\left(\Theta^{\mathbb{Q}}, t, T\right) \\
\xi_1 &= \tilde{C}\left(\Theta^{\mathbb{Q}}, t, \mathbb{T}\right) - \tilde{C}\left(\Theta^{\mathbb{Q}}, t, T\right)
\end{aligned}
$$

Then, the expected quadratic variation can be written as

$$
E_t^{\mathbb{Q}_T}\left[\int_t^T v_s' v_s ds\right] = E_t^{\mathbb{Q}_T}\left[\int_t^T \left(\xi_0 + 2\xi_1 X_s\right)' \Omega \left(\xi_0 + 2\xi_1 X_s\right) ds\right]
$$

Suppose that the i-th element of X_t has no impact on the volatilities in yields. This implies that i-th row and column of $\tilde{C}\left(\Theta^{\mathbb{Q}}, t, T\right)$ is zero for all $T > 0$. Since the instantaneous volatility of $P_{t,T}$ is

$$\sigma_{t,T} = \sqrt{\Omega}' \left[\tilde{B}\left(\Theta^{\mathbb{Q}}, t, T\right) + 2\tilde{C}\left(\Theta^{\mathbb{Q}}, t, T\right) X_t\right],$$

the dynamics of the state under $\mathbb{Q}_T$ is

$$dX_t = \left[K_{0X}^{\mathbb{Q}_T} + K_{1X}^{\mathbb{Q}_T} X_t\right] dt + \sqrt{\Omega} dB_t^{\mathbb{Q}_T}$$

where

$$\begin{aligned} K_{0X}^{\mathbb{Q}_T} &= K_{0X}^{\mathbb{Q}} + \Omega\tilde{B}\left(\Theta^{\mathbb{Q}}, t, T\right) \\ K_{1X}^{\mathbb{Q}_T} &= K_{1X}^{\mathbb{Q}} + 2\Omega\tilde{C}\left(\Theta^{\mathbb{Q}}, t, T\right) \end{aligned}$$

Appendix D Discrete-time term structure model with stochastic volatility

Appendix D.1 Zero-coupon bonds' loading on pricing factors

Denote the price of zero-coupon bond with maturity of n as $P_t^{(n)}$. Then, we can show that $\ln P_t^{(n)} = -\bar{A}_n - \bar{B}_{V,n} V_t - \bar{B}_{X,n} X_t$ with loadings given by

$$\begin{aligned} \bar{A}_n &= \delta_0 + \bar{A}_{n-1} + K_{0X}^{\mathbb{Q}\prime} \bar{B}_{X,n} + c^{\mathbb{Q}} \nu^{\mathbb{Q}\prime} \bar{B}_{V,n-1} - \frac{1}{2}\alpha_{n-1} \\ \bar{B}_{V,n} &= \delta_V + K_{1XV}^{\mathbb{Q}\prime} \bar{B}_{X,n-1} + \rho^{\mathbb{Q}\prime} \bar{B}_{V,n-1} - \frac{1}{2}\beta_{n-1} \\ \bar{B}_{X,n} &= \delta_X + K_{1X}^{\mathbb{Q}\prime} \bar{B}_{X,n-1} \end{aligned} \tag{A-8}$$

where

$$\alpha_n = \bar{B}_{X,n}' \Sigma_{0X} \Sigma_{0X}' \bar{B}_{X,n} - 2 v_V^{\mathbb{Q}\prime} \left[\log \mathcal{J}_n - c_V^{\mathbb{Q}\prime} \bar{B}_{XV,n}\right] \tag{A-9}$$

$$\beta_n = -2\mathcal{G}_n + \left(\bar{B}_{X,n}' \Sigma_{1X} \Sigma_{1X}' \bar{B}_{X,n}, \cdots, \bar{B}_{X,n}' \Sigma_{mX} \Sigma_{mX}' \bar{B}_{X,n}\right)' \tag{A-10}$$

and

$$\begin{aligned} \mathcal{G}_{n-1} &= \rho^{\mathbb{Q}\prime} c^{\mathbb{Q}-1\prime} \left(\left[\operatorname{diag}\left(\mathcal{J}_{n-1}\right)\right]^{-1} - I_m\right) c^{\mathbb{Q}\prime} \bar{B}_{XV,n-1} \\ \mathcal{J}_{n-1} &= \iota_m + c^{\mathbb{Q}\prime} \bar{B}_{XV,n-1} \\ \bar{B}_{XV,n-1} &= \Sigma_{XV}' \bar{B}_{X,n-1} + \bar{B}_{V,n-1} \end{aligned}$$

The initial condition of the difference equation is $\bar{A}_0 = \bar{B}_{V,0} = \bar{B}_{X,n} \equiv 0$.

In addition, denoting the yield on a zero-coupon bond of maturity n as $y_{n,t}$, then, we have

$$y_{n,t} = A_n + B_n Z_t$$

where $A_n = \frac{1}{n}\bar{A}_n$, $B_{V,n} = \frac{1}{n}\bar{B}_{V,n}$ and $B_{V,n} = \frac{1}{n}\bar{B}_{X,n}$. Stacked yields, y_t, can be represented as

$$y_t = A + B_V V_t + B_X X_t$$

where A, B_V and B_X are corresponding stacked A_n, $B_{V,n}$ and $B_{X,n}$. Furthermore it can be written as a function of the observable factor $\mathcal{Z}_t$ as in Section 5. With $W \in \mathbb{R}^{(N-m)\times J}$ being the weighting matrix to construct $\mathcal{P}_t$,

$$y_t = A + B_V V_t + B_X X_t = \mathcal{A} + \mathcal{B}_{\mathcal{P}}\mathcal{P}_t + \mathcal{B}_{\mathcal{V}}\mathcal{V}_t$$

where

$$\begin{aligned} \mathcal{B}_{\mathcal{P}} &= B_X (W B_X)^{-1} \\ \mathcal{B}_{\mathcal{V}} &= (I - \mathcal{B}_{\mathcal{P}} W) B_V \beta^{-1} \\ \mathcal{A} &= (I - \mathcal{B}_{\mathcal{P}} W) A - \mathcal{B}_{\mathcal{V}} \alpha \end{aligned} \tag{A-11}$$

Appendix D.2 $\mathcal{V}_t$'s loading on the latent variance factor

As shown in Choi, Mueller, and Vedolin (2016),

$$\begin{aligned} \mathcal{V}_t &\equiv \frac{2}{P_{t,T}}\left[\int_0^{F_t(T,\mathbb{T})} \frac{\text{Put}_t(K,T,\mathbb{T})}{K^2}dK + \int_{F_t(T,\mathbb{T})}^{\infty} \frac{\text{Call}_t(K,T,\mathbb{T})}{K^2}dK\right] \\ &= 2\left[\ln E_t^{\mathbb{Q}_T}(F_T(T,\mathbb{T})) - E_t^{\mathbb{Q}_T}(\ln F_T(T,\mathbb{T}))\right] \end{aligned}$$

In the case of one-month TIVs or TYVIX, T is equal to $t+1$. Hence, the calculation for α and β should be taken under $\mathbb{Q}_{t+1}$ forward measure (i.e. the risk neutral measure). I use the notation $\mathbb{Q}$ instead of $\mathbb{Q}_{t+1}$ according to the convention in the literature, and denote the time to maturity of the underlying bond on the expiration date of the forward as $n = \mathbb{T} - (t+1)$ for notational simplicity. Then,

$$\begin{aligned} \ln E_t^{\mathbb{Q}}[F_{t+1}(t+1,\mathbb{T})] &= \ln F_t(t+1,\mathbb{T}) = \ln P_t^{(n+1)} - \ln P_t^{(1)} = -(A_{n+1} - A_1) - \left(B'_{n+1} - B'_1\right) Z_t \\ E_t^{\mathbb{Q}}[\ln F_{t+1}(t+1,\mathbb{T})] &= E_t^{\mathbb{Q}}\left(\ln P_{t+1}^{(n)}\right) = -A_n - B'_n E_t^{\mathbb{Q}}(Z_{t+1}) \\ &= -A_n - B'_{X,n}\left(K_{0X}^{\mathbb{Q}} + K_{1X}^{\mathbb{Q}} X_t + K_{1XV}^{\mathbb{Q}} V_t\right) - B'_{V,n}\left(c^{\mathbb{Q}}\nu^{\mathbb{Q}} + \rho^{\mathbb{Q}} V_t\right) \end{aligned}$$

Using the difference equations in (A-8), we have

$$\begin{aligned} \mathcal{V}_t^{(n)} &= 2\ln E_t^{\mathbb{Q}}[F_{t+1}(t+1,\mathbb{T})] - 2E_t^{\mathbb{Q}}[\ln F_{t+1}(t+1,\mathbb{T})] \\ &= \alpha_n + \beta'_n V_t \end{aligned}$$

where α_n and β_n are given in equations (A-9) and (A-10).

Appendix E Test for the USV effect

Denote R as the number of priced factors. For $J \geq R$, a $(J \times 1)$ vector of yields can be written as

$$y_t = A + B_X X_t + B_V V_t = \mathcal{A} + \mathcal{B}_{\mathcal{P}}\mathcal{P}_t + \mathcal{B}_{\mathcal{V}}\mathcal{V}_t$$

where $\mathcal{P}_t\,(\equiv W y_t)$ is the first $(R-m)$ principal components of y_t, $\mathcal{V}_t$ is the observable measure of variance risk, and

$$\mathcal{B}_{\mathcal{V}} = \left(I - B_X (W B_X)^{-1} W\right) B_V \beta^{-1}$$

as in equation (A-11). Suppose that, without loss of generality, the first volatility factor V_{1t} is an unspanned volatility. Then, there exists Φ such that $B_{V_1} = B_X \Phi$ (see, for example, Lemma 2 in Joslin (2015)), which implies

$$\begin{aligned}\mathcal{B}_{\mathcal{V}} &= \left(I - B_X \left(W B_X\right)^{-1} W\right) \left[\begin{array}{cc} B_X \Phi, & B_{V_{2:m}} \end{array}\right] \beta^{-1} \\ &= \left[\begin{array}{cc} 0, & \left(I - B_X \left(W B_X\right)^{-1} W\right) B_{V_{2:m}} \end{array}\right] \beta^{-1}\end{aligned}$$

Hence, $\mathcal{B}_{\mathcal{Z}} \equiv \left[\begin{array}{cc} \mathcal{B}_{\mathcal{P}}, & \mathcal{B}_{\mathcal{V}} \end{array}\right]$ cannot be a full rank matrix in the presence of USV.

Appendix F Returns on generalized mimicking portfolios

As in Joslin, Priebsch, and Singleton (2014), we have

$$\mathbb{E}_t^{\mathbb{P}} \left[P_{t+1}^{(n-1)} / P_t^{(n)}\right] = \exp\left[k_t^{\mathbb{P}} \left(\mathcal{Z}_{t+1}; -B_{n-1}\right) - k_t^{\mathbb{Q}} \left(\mathcal{Z}_{t+1}; -B_{n-1}\right) + r_t\right]$$

where $k_t \left(\mathcal{Z}_{tt+1}; u\right)$ denotes the conditional cumulant generating function of $\mathcal{Z}_{t+1}$ at time t, and B_n is the corresponding factor loading. In addition, the Laplace transform of the mixture of Gaussian and multivariate non-central gamma distribution is given by

$$\begin{aligned}\mathbb{E}\left[\exp\left(u' \mathcal{Z}_{t+1}\right)\right] &= \exp\left(u'_{\mathcal{P}} \left(K_{0\mathcal{P}} + K_{1\mathcal{P}} \mathcal{P}_t + K_{1\mathcal{P}\mathcal{V}} \mathcal{V}_t - \Sigma_{\mathcal{P}\mathcal{V}} \left[\mu + c\nu + \rho\left(\mathcal{V}_t - \mu\right)\right]\right) + \frac{1}{2} u'_{\mathcal{P}} \Sigma_{\mathcal{P},t} \Sigma'_{\mathcal{P},t} u_{\mathcal{P}}\right) \\ &\quad \times \exp\left(u'_{\mathcal{P}\mathcal{V}} \mu - \sum_{i=1}^{m} v_i \log\left(1 - e'_i c' u_{\mathcal{P}\mathcal{V}}\right) + \sum_{i=1}^{m} \frac{e'_i c' u_{\mathcal{P}\mathcal{V}}}{1 - e'_i c' u_{\mathcal{P}\mathcal{V}}} e'_i c^{-1} \rho\left(\mathcal{V}_t - \mu\right)\right)\end{aligned}$$

where $u_{\mathcal{P}\mathcal{V}} = \Sigma'_{\mathcal{P}\mathcal{V}} u_{\mathcal{P}} + u_{\mathcal{V}}$ and e_i is a zero vector except its i-th element is 1. μ denotes the lower bound of $\mathcal{V}_t$, and is equal to α_n in Appendix D.2 when the lower bound on latent V_t is set to be zero. Under the normalization scheme of $c^{\mathbb{P}} = c^{\mathbb{Q}}$ and $\Sigma_{XV}^{\mathbb{P}} = \Sigma_{XV}^{\mathbb{Q}}$, we have

$$\begin{aligned}& k_t^{\mathbb{P}} \left(\mathcal{Z}_{t+1}; -B_{n-1}\right) - k_t^{\mathbb{Q}} \left(\mathcal{Z}_{t+1}; -B_{n-1}\right) \\ = \;& B'_{\mathcal{P},n-1} \left[\left(K_{0\mathcal{P}}^{\mathbb{Q}} - K_{0\mathcal{P}}^{\mathbb{P}}\right) + \left(K_{1\mathcal{P}}^{\mathbb{Q}} - K_{1\mathcal{P}}^{\mathbb{P}}\right) \mathcal{P}_t + \left(K_{1\mathcal{P}\mathcal{V}}^{\mathbb{Q}} - K_{1\mathcal{P}\mathcal{V}}^{\mathbb{P}}\right) \mathcal{V}_t\right] \\ & - B'_{\mathcal{P},n-1} \Sigma_{\mathcal{P}\mathcal{V}} \left[c\left(\nu^{\mathbb{Q}} - \nu^{\mathbb{P}}\right) + \left(\rho^{\mathbb{Q}} - \rho^{\mathbb{P}}\right)\left(\mathcal{V}_t - \mu\right)\right] \\ & + \sum_{i=1}^{m} \left(v_i^{\mathbb{Q}} - v_i^{\mathbb{P}}\right) \log\left(1 + \mathcal{A}_i\right) + \sum_{i=1}^{m} \frac{\mathcal{A}_i}{1 + \mathcal{A}_i} e'_i c^{-1} \left(\rho^{\mathbb{Q}} - \rho^{\mathbb{P}}\right)\left(\mathcal{V}_t - \mu\right)\end{aligned}$$

where the constant $\mathcal{A}_i$ is given by

$$\mathcal{A}_i = e'_i c' \left(\Sigma'_{\mathcal{P}\mathcal{V}} B_{\mathcal{P},n-1} + B_{\mathcal{V},n-1}\right)$$

Since $\frac{1}{1+x} \approx x$ and $\log(1+x) \approx x$ for small x, the above can be approximated by

$$\begin{aligned}
&B'_{\mathcal{P},n-1}\left[\left(K^{\mathbb{Q}}_{0\mathcal{P}} - K^{\mathbb{P}}_{0\mathcal{P}}\right) + \left(K^{\mathbb{Q}}_{1\mathcal{P}} - K^{\mathbb{P}}_{1\mathcal{P}}\right)\mathcal{P}_t\right] + B'_{\mathcal{P},n-1}\left(K^{\mathbb{Q}}_{1\mathcal{P}\mathcal{V}} - K^{\mathbb{P}}_{1\mathcal{P}\mathcal{V}}\right)\mathcal{V}_t \\
&+ \left[B'_{\mathcal{P},n-1}\Sigma_{\mathcal{P}\mathcal{V}} - \sum_{i=1}^{m}\mathcal{A}_i e'_i c^{-1}\right]\left(\rho^{\mathbb{P}} - \rho^{\mathbb{Q}}\right)(\mathcal{V}_t - \mu) \\
&+ B'_{\mathcal{P},n-1}\Sigma_{\mathcal{P}\mathcal{V}} c\left(\nu^{\mathbb{P}} - \nu^{\mathbb{Q}}\right) - \sum_{i=1}^{m}\mathcal{A}_i\left(v^{\mathbb{P}}_i - v^{\mathbb{Q}}_i\right)
\end{aligned}$$

When $m = 1$ and $\nu^{\mathbb{Q}}_i = \nu^{\mathbb{P}}_i$, this can be further simplified as

$$B'_{\mathcal{P},n-1}\left[\left(K^{\mathbb{Q}}_{0\mathcal{P}} - K^{\mathbb{P}}_{0\mathcal{P}}\right) + \left(K^{\mathbb{Q}}_{1\mathcal{P}} - K^{\mathbb{P}}_{1\mathcal{P}}\right)\mathcal{P}_t + \left(K^{\mathbb{Q}}_{1\mathcal{P}\mathcal{V}} - K^{\mathbb{P}}_{1\mathcal{P}\mathcal{V}}\right)\mathcal{V}_t\right] + B_{\mathcal{V},n-1}\left(\rho^{\mathbb{Q}} - \rho^{\mathbb{P}}\right)(\mathcal{V}_t - \mu)$$

Now consider a linear combination of yields $y^a_t = \sum_{i=1}^{N} a_i y^{n_i}_t$ where n_i denotes the maturity of the i-th yield. To construct a mimicking portfolio of it, we need to find $\{w_i\}_{i=1}^{N}$ such that

$$\frac{dP^w_t}{dy^a_t} = \sum_{i=1}^{N}\frac{dP^w_t}{dy^{n_i}_t}\frac{dy^{n_i}_t}{dy^a_t} = -\sum_{i=1}^{N} w_i n_i P^{n_i}_t \frac{1}{a_i} = 1$$

which will hold for weights

$$w_i = -\frac{a_i}{N n_i P^{n_i}_t}$$

Consider the one-period excess return on portfolio P^w_t:

$$\frac{\sum_i w_i\left(P^{n_i-1}_{t+1} - e^{r_t}P^{n_i}_t\right)}{|\sum_i w_i P^{n_i}_t|} = \frac{-\sum_i a_i/n_i\left(P^{n_i-1}_{t+1}/P^{n_i}_t - e^{r_t}\right)}{|\sum_i w_i P^{n_i}_t|}$$

Using $e^x \approx 1 + x$, we have

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}_t\left[P^{n_i-1}_{t+1}/P^{n_i}_t\right] &= \exp\left[k^{\mathbb{P}}_t(\mathcal{Z}_{t+1}; -B_{n-1}) - k^{\mathbb{Q}}_t(\mathcal{Z}_{t+1}; -B_{n-1}) + r_t\right] \\
&\approx 1 + k^{\mathbb{P}}_t(\mathcal{Z}_{t+1}; -B_{n-1}) - k^{\mathbb{Q}}_t(\mathcal{Z}_{t+1}; -B_{n-1}) + r_t
\end{aligned}$$

which implies that

$$\frac{-\sum_i a_i/n_i E^{\mathbb{P}}_t\left[P^{n_i-1}_{t+1}/P^{n_i}_t - e^{r_t}\right]}{|\sum_i w_i P^{n_i}_t|} = \frac{-\sum_i a_i/n_i\left[k^{\mathbb{P}}_t(\mathcal{Z}_{t+1}; -B_{n-1}) - k^{\mathbb{Q}}_t(\mathcal{Z}_{t+1}; -B_{n-1})\right]}{|\sum_i w_i P^{n_i}_t|}$$

Hence, the expected excess return on portfolio P^w_t, to a first-order approximation, is given by

$$\frac{\sum_i a_i/n_i B_{n_i-1}\left[\Lambda_0 + \Lambda_1\mathcal{Z}_t\right]}{|\sum_i a_i/n_i|} \tag{A-12}$$

where

$$\Lambda_0 = \begin{bmatrix} K_{0\mathcal{P}}^{\mathbb{P}} - K_{0\mathcal{P}}^{\mathbb{Q}} \\ -\left(\rho^{\mathbb{P}} - \rho^{\mathbb{Q}}\right)\mu \end{bmatrix}$$

$$\Lambda_1 = \begin{bmatrix} \left(K_{1\mathcal{PP}}^{\mathbb{P}} - K_{1\mathcal{PP}}^{\mathbb{Q}}\right) & K_{1\mathcal{PM}}^{\mathbb{P}} & \left(K_{1\mathcal{PV}}^{\mathbb{P}} - K_{1\mathcal{PV}}^{\mathbb{Q}}\right) \\ 0 & 0 & \rho^{\mathbb{P}} - \rho^{\mathbb{Q}} \end{bmatrix}$$

and

$$B_{n_i-1} = \left(B'_{\mathcal{P},n-1}, B'_{\mathcal{V},n-1}\right)'$$

Since the first $(N-m)$ elements of $\mathcal{Z}_t$ correspond to the first $(N-m)$ principal components of yields

$$PCj_t = \sum_{i=1}^{N} l_i^j y_t^{n_i} = \sum_{i=1}^{N} l_i^j \left(A_{n_i}/n_i + B_{n_i}/n_i \mathcal{Z}_t\right)$$

it follows that $\sum_{i=1}^{N} l_i^j B_{n_i}/n_i$ is the selection vector for the j-th element (e.g. $(1,0,0)$ for $j=1$) for $j \leq N-m$. Replacing a_i of equation (A-12) with l_i^j and approximating B_{n_i-1} with B_{n_i}, we have

$$\frac{\sum_i l_i^j B_{n_i-1}/n_i \left[\Lambda_0 + \Lambda_1 \mathcal{Z}_t\right]}{\left|\sum_i l_i^j/n_i\right|} = \frac{\overbrace{\sum_i l_i^j B_{n_i}/n_i}^{\text{selction vector}} \left[\Lambda_0 + \Lambda_1 \mathcal{Z}_t\right]}{\left|\sum_i l_i^j/n_i\right|}$$

which implies that $xPCj$ is given by the j-th row of $\Lambda_0 + \Lambda_1 \mathcal{Z}_t$ scaled by $\left|\sum_i l_i^j/n_i\right|$ for $j \leq N-m$.

Appendix G Estimates of $\mathcal{M}_C$

Under the risk-neutral measure $\mathbb{Q}$, the latent state variable is drift-normalized as in Joslin (2015) or Creal and Wu (2015), for econometric identification. Then, the short rate equation for $\mathcal{M}_C$ is[11]

$$r_t = r_\infty^{\mathbb{Q}} + \iota' X_t + \delta_V V_t \tag{A-13}$$

where ι denotes a vector of ones. δ_V can take ± 1 or 0, and each possible value induces different local maxima: see for example Creal and Wu (2015). The conditional mean of the normalized state $Z_t \equiv (X_t', V_t)'$ is

$$\begin{bmatrix} E_t^{\mathbb{Q}}(X_{t+1}) \\ E_t^{\mathbb{Q}}(V_{t+1}) \end{bmatrix} = \begin{bmatrix} 0 \\ K_{0V}^{\mathbb{Q}} \end{bmatrix} + \begin{bmatrix} diag\left(\lambda^{\mathbb{Q}}\right) & 0 \\ 0 & \rho^{\mathbb{Q}} \end{bmatrix} \begin{bmatrix} X_t \\ V_t \end{bmatrix} \tag{A-14}$$

where $K_{0V}^{\mathbb{Q}}$ is the product of the scale and shape parameters of V_t.

Table A-1 presents the estimates of the $\mathbb{Q}$ parameters for $\mathcal{M}_C$ and $\mathcal{M}_J$. The persistency of Gaussian factors, measured by $\lambda^{\mathbb{Q}}$, is similar across the two models. For $\mathcal{M}_J$, δ_V in equation (A-13) should be zero and V_t does not affect the conditional variance of X_t. Then, $r_\infty^{\mathbb{Q}}$ can be interpreted as the long-run $\mathbb{Q}$ mean of the short rate, since the long-run mean of X_t is set to zero under $\mathbb{Q}$

[11] For ease of explanation, $K_{1X}^{\mathbb{Q}}$ is assumed to have real distinct eigenvalues - this is overidentifying. For details, see Joslin, Singleton, and Zhu (2011).

as in equation (A-14). For $\mathcal{M}_C$, the long-run $\mathbb{Q}$ mean of the short rate is $r_\infty^{\mathbb{Q}} + \delta_V K_{0V}^{\mathbb{Q}} / \left(1 - \rho^{\mathbb{Q}}\right)$ rather than $r_\infty^{\mathbb{Q}}$, which induces the difference between the $r_\infty^{\mathbb{Q}}$s of $\mathcal{M}_C$ and $\mathcal{M}_J$. In addition, the likelihood of $\mathcal{M}_C$ is maximized with $\delta_V = 1$ among the three possible values of δ_V. This is in line with $r_\infty^{\mathbb{Q}}$ of $\mathcal{M}_C$ being slightly less than $r_\infty^{\mathbb{Q}}$ of $\mathcal{M}_J$ for the two models to have similar levels in the long-run $\mathbb{Q}$ mean of the short rate.

For each maturity n, Table A-2 reports the fraction $\mathrm{var}(B_n \iota_i \iota_i' Z_t) / \mathrm{var}(B_n Z_t)$ where ι_i denotes a vector of zero with i-th element being one. The table thus represents the relative contribution of each latent factor toward the yields curve movement. Note that the exercises in Section 4 analyze the marginal impact of $\mathcal{V}_t$ after controlling the yield curve factors. Because the yield curve factors themselves, $\mathcal{P}_t$, are linear functions of latent factors X_t and V_t, the sole impact of $\mathcal{V}_t$ on the shape of the yield cannot be assessed in this setting. Within a fully-fledged ADTSM, the impact of V_t on the cross-section of yields can be completely isolated from the impact of latent Gaussian factor X_t. The table performs this exercise, and shows that the shape of the yields curve is largely unexplained by V_t or $\mathcal{V}_t$. Note that the interpretations of $\mathcal{V}_t$ and V_t are freely interchangeable in the context, since one is an invariant transformation of the other: one can freely scale up or down V_t, then yield loadings on V_t are adjusted accordingly so that its impact on the yield curve still remains the same.

Table A-1
Persistence Parameters

This table reports the estimates of persistence parameters for the each model $\mathcal{M}_C$ and $\mathcal{M}_J$. Standard errors are given in parentheses.

Model	$r_\infty^{\mathbb{Q}}$	$\lambda_1^{\mathbb{Q}}$	$\lambda_2^{\mathbb{Q}}$	$\lambda_3^{\mathbb{Q}}$	$\rho^{\mathbb{Q}}$
$\mathcal{M}_C$	0.098	0.996	0.963	0.903	0.949
	(0.012)	(0.000)	(0.003)	(0.010)	(0.036)
$\mathcal{M}_J$	0.108	0.996	0.963	0.906	
	(0.003)	(0.000)	(0.002)	(0.008)	

Table A-2
Yield Curve Decomposition

This table reports the fraction $\mathrm{var}(B_n \iota_i \iota_i' Z_t) / \mathrm{var}(B_n Z_t)$ where ι_i denotes a vector of zero with i-th element being one.

$\iota_i' Z_t$	n						
	6 mon	1 yr	2 yr	3 yr	5 yr	7 yr	10 yr
X_1 ($\equiv \iota_1' Z_t$)	1.20	1.22	1.28	1.37	1.47	1.49	1.45
X_2 ($\equiv \iota_2' Z_t$)	3.60	3.02	2.24	1.73	1.07	0.69	0.40
X_3 ($\equiv \iota_3' Z_t$)	0.64	0.39	0.18	0.10	0.05	0.03	0.01
V ($\equiv \iota_4' Z_t$)	0.03	0.03	0.02	0.01	0.01	0.00	0.00