Modeling Volatility in Dynamic Term Structure Models

Hitesh Doshi
University of Houston

Kris Jacobs
University of Houston

Rui Liu
Duquesne University

June 26, 2020

Abstract

We propose a class of no-arbitrage term structure models in which the volatility factors follow GARCH processes. The models' tractability is similar to that of canonical affine term structure models, but they capture the conditional variances of yields much more accurately. We estimate a model with one volatility factor using 1971-2019 Treasury yield data. Relative to standard affine term structure models with stochastic volatility, the model improves the fit of yield volatility substantially, especially for long-maturity yields. This improvement does not come at the expense of a deterioration in yield fit. We conclude that the ability of no-arbitrage term structure models to extract and model conditional volatility critically depends on the specification of the volatility factors. Modeling volatility as a function of (lagged) squared innovations to factors improves on models where volatility is a linear function of the factors.

JEL Classification: G12, C58, E43

Keywords: term structure; stochastic volatility; GARCH.

1 Introduction

There is a wealth of evidence in the literature indicating that volatility implied by standard no-arbitrage term structure models corresponds poorly to measures of realized volatility or other model-free estimates. For state-of-the-art affine term structure models (ATSMs) with stochastic volatility, simultaneously matching the properties of the conditional means and variances of yields is indeed the key empirical challenge. These models contain an inherent tension between fitting yields and fitting yield volatility, partly because the mean and variance of yields are driven by the levels of the same state variables.[1]

This paper proposes a class of no-arbitrage term structure models in which the volatility factors follow GARCH processes. We provide analytical solutions for bond prices for a large class of models. The models' tractability is thus similar to that of canonical affine volatility models, but they capture the time variation in the conditional variances of yields much more accurately. The model is motivated by the well-established literature on ARCH and GARCH models (Engle, 1982; Bollerslev, 1986). There is ample evidence that ARCH/GARCH models provide an adequate characterization of interest rate volatility (see Koedijk, Nissen, Schotman, and Wolff 1997; Brenner, Harjes, and Kroner, 1996; Christiansen, 2005). We incorporate the superior fit of GARCH models into state-of-the-art term structure models by relating the conditional volatility to the lagged squared *residuals* of the factors driving yields. From a modeling perspective, this is a critical difference with standard ATSMs with stochastic volatility, which model volatility as a linear combination of the *level* of the yield factors.

While this may appear to be a trivial distinction, we show that it greatly matters for the models' ability to capture the stylized facts of the time series and cross-section of conditional volatility. In the empirical analysis, we follow the existing literature and focus on a version of the model with three latent factors, but for simplicity and parsimony we only endow the volatility of

[1] See Dai and Singleton (2000, 2002), Duffee (2002), Collin-Dufresne, Goldstein, and Jones (2009), and Duarte (2004) for evidence on the volatility fit of these models and the trade-off between fitting yields and yield volatility. Joslin and Le (2013) discuss in depth why there is a trade-off between fitting yields and yield volatilities.

the residuals of the first factor with a GARCH dynamic. We estimate the model using monthly Treasury yields from November 1971 to October 2019. In our main empirical results, we use an EGARCH(1, 1) model as a measure of model-free volatility, but we show that the results are very similar when using realized volatility instead.[2] We find that the model-implied conditional volatility performs well in capturing model-free volatility, especially at longer maturities (4-5 and 10 years). The proposed model captures the high volatility periods of the early 1980s well at longer maturities, and fits the low volatility periods well for all bond maturities. The unconditional correlation between model-implied and model-free volatility is 90% on average across maturities. We also compute the root mean squared errors (RMSEs) and the mean absolute percentage errors (MAPEs) based on the model-implied and model-free volatilities. The RMSEs on average across all maturities is about 14 basis points. For yield volatilities at 3-5 and 10 years, the RMSEs are below 10 basis points. In summary, the proposed model performs well in fitting the conditional second moment of yields, even though it contains only one volatility factor.

We compare the implications of the proposed model with those from standard affine stochastic volatility models. We estimate a three-factor affine model with one factor driving the volatility process, the essentially affine $A_1(3)$ model (see Dai and Singleton, 2000; Duffee, 2002). To facilitate the comparison with our proposed model, we also consider a restricted version of the canonical $A_1(3)$ model, in which the dynamics of the yield curve factors are exactly the same as those in our proposed model. The critical difference with the new model is that in both the canonical and restricted $A_1(3)$ models, the volatility process depends on the level of the factors driving the yield curve. Consistent with the literature, we find that the two benchmark models do not perform as well in capturing the time-variation in the conditional yield volatilities in our sample. On average across all maturities, the unconditional correlation with the EGARCH

[2] GARCH and EGARCH estimates are widely used to measure the "true" conditional volatility in the literature on term structure volatilities (Bikbov and Chernov, 2010; Collin-Dufresne, Goldstein, and Jones, 2009; Dai and Singleton, 2003).

estimates is 67% for the canonical $A_1(3)$ model and 66% for the restricted $A_1(3)$ model.[3] The proposed no-arbitrage model with GARCH volatility therefore substantially outperforms the benchmark models.

The benchmark models cannot capture the high volatility during the monetary experiment in the early 1980s. The estimated short-term volatilities are similar before and after the monetary experiment for both benchmark models. The canonical and restricted $A_1(3)$ models also overestimate yield volatility during the low-volatility period between the mid-1980s to 2000, and also exhibit very limited time-variation across the maturity spectrum at the start of the sample. The proposed no-arbitrage model with GARCH volatility outperforms the two benchmark models by 33% on average across maturities in terms of yield volatility RMSE, and by approximately 40% on average across maturities in terms of yield volatility MAPE. The improvement over the canonical $A_1(3)$ model is more significant at longer maturities. For example, for the 5-year yield, the improvement in RMSEs is about 55%, and the improvement in MAPEs is about 60%. These findings suggest that we can substantially improve the ability of ATSMs to fit yield volatility by endowing the volatility factor with a GARCH process. The state-of-the art stochastic volatility models in the term structure literature are not able to accurately model the dynamics of volatility, because by design a linear combination of the yield factors is not able to capture the second moment of yields. We emphasize that it is this aspect of the model structure that drives the differences in performance, and that the differences are not driven by the discrete- versus continuous-time specification of the models. Note also that the GARCH variance dynamic is a function of lagged innovations, while stochastic volatility contains an independent innovation. By design this model feature favors the stochastic volatility models. Therefore, the improved performance of the GARCH models is due to their specification of the variance as a function of

[3] The performance of ATSMs in matching yields volatilities is somewhat model- and sample-dependent. We find a positive correlation as in Jacobs and Karoui (2009), because we also use a relative long sample of Treasury yields and include the high inflation period. In Andersen and Benzoni (2010) and Collin-Dufresne, Goldstein, and Jones (2009), the performance of these models is worse.

(lagged) squared *innovations* rather than as a function of the *levels* of the yield factors.

To demonstrate that the proposed model's improved ability in fitting conditional yield volatility does not come at the expense of a poor fit for the level of yields, we compare the in-sample yield fit of the proposed model and the benchmark models. The yield fit of the proposed model is very similar to that of the two benchmark stochastic volatility models. Moreover, we find that the proposed model captures the empirical patterns in bond risk premia, as characterized by the regression coefficients in the Campbell and Shiller (1991) regressions, very well. We confirm the finding of Dai and Singleton (2000) that benchmark no-arbitrage stochastic volatility models are not able to capture these deviations from the expectations hypothesis in the data. These findings indicate that the proposed model provides an adequate fit to the conditional means as well as the variances of yields.

We contribute to several strands of literature. As mentioned above, an extensive literature has questioned the ability of ATSMs with stochastic volatility to model conditional volatility.[4] We show that the performance of term structure models in fitting conditional volatility can be improved greatly with a different specification of the volatility factor. This improvement does not come at the expense of the fit of the yield level. Our contribution is closely related to previous attempts to build ARCH/GARCH volatility into no-arbitrage term structure models, in particular Heston and Nandi (2003).[5] In our empirical analysis, we estimate a more general model using a long sample of Treasury yields data, while Heston and Nandi (2003) calibrate their model using zero-coupon bond prices for a two-week sample. The limited empirical exercise does not allow them analyze the modeling of yield volatility or the model's potential to resolve the tension between modeling yield levels and volatilities. We also explicitly compare the performance of the proposed model to that of state-of-the-art stochastic volatility models.

[4]See for example Collin-Dufresne and Goldstein (2002), and Andersen and Benzoni (2010). See also Heidari and Wu (2003), Fan, Gupta, and Ritchken (2003), Jagannathan, Kaplin, and Sun (2003), and Li and Zhao (2006). See Bikbov and Chernov (2010), and Tang and Xia (2007) for studies using different fixed-income data.

[5]In other related work, the volatility factors in Longstaff and Schwartz (1992) have been interpreted as following a GARCH process. Haubrich, Pennacchi, and Ritchken (2012) introduce GARCH volatility into a macro-finance term structure model.

Our work also complements a rich literature that provides alternative solutions to model interest rate volatility in a no-arbitrage framework. For example, Collin-Dufresne and Goldstein (2002) and Collin-Dufresne, Goldstein, and Jones (2009) propose an unspanned stochastic volatility model by imposing a set of parametric restrictions to break the spanning of conditional volatilities by yields.[6] However, Joslin (2018) shows that the unspanned stochastic volatility restrictions unduly constrain other aspects of model dynamics, and that these restrictions are rejected by the data. Ghysels, Le, Park, and Zhu (2014) impose a component GARCH volatility structure on the no-arbitrage term structure model. Their approach results in unspanned volatility under the physical measure, while our approach is much simpler and falls under the class of spanned volatility models. Finally, the class of models we propose is also related to an existing literature that goes beyond the affine paradigm. Examples include affine-quadratic models (see Ahn, Dittmar, and Gallant, 2002; Ahn, Dittmar, Gao, and Gallant, 2003; Leippold and Wu, 2002), regime-switching models (see for example Dai, Singleton, and Yang, 2007; Bansal and Zhou, 2002; Bansal, Tauchen, and Zhou, 2004; Ang and Bekaert, 2002), and other nonlinear models (see for example Ahn and Gao, 1999; Feldhutter, Heyerdahl-Larsen, and Illeditsch, 2018). We complement these studies by offering a parsimonious yet flexible model for capturing interest rate volatility.

The paper proceeds as follows. Section 2 presents the specification of term structure models with GARCH volatility and the benchmark ATSMs with stochastic volatility. Section 3 provides the data and discusses the estimation method. Section 4 discusses the parameter estimates, the models' performance in fitting the conditional volatility of yields, the model-implied term structure of unconditional yield volatility, and the trade-off between fitting the conditional means and variances of yields. Section 5 presents robustness results, and Section 6 concludes.

[6] Creal and Wu (2015) provide new estimation procedures for ATSMs with spanned or unspanned stochastic volatility. They find that models with spanned volatility fit the cross section of the yield curve better, while those with unspanned volatility fit volatility better.

2 Models

In this section, we first discuss the structure of the proposed ATSMs with GARCH volatility. Subsequently, we briefly discuss the models we use as empirical benchmarks throughout the paper. we use the canonical affine stochastic volatility models as specified in Dai and Singleton (2000), in which the conditional covariance of the state variable is an affine function of the state variables. This class of models is motivated by a rich body of literature showing that the volatility of yield curve is, at least partially related to the shape of the yield cure. For example, the volatility of interest rate is usually high when interest rates are high and when the yield curve exhibits higher curvature (see Cox, Ingersoll, and Ross, 1985; Litterman, Scheinkman, and Weiss, 1991; Longstaff and Schwartz, 1992). We also consider a restricted version of the canonical affine stochastic volatility model as an additional benchmark. In this model, the volatility process is an affine function of the state variable, and the dynamics of the state variables are the same as in the model we propose. The only difference between this benchmark and the newly proposed model is the specification of the volatility process.

2.1 ATSMs with GARCH Volatility

We propose a discrete-time term structure model with analytical solutions for bond prices, in which the volatility factors follow a GARCH process. The GARCH literature is formulated in discrete time, which facilitates model implementation. This specification of the volatility process is motivated by the large literature on ARCH and GARCH modeling. There is considerable evidence that ARCH/GARCH processes provide a good description of interest rate volatility (Koedijk, Nissen, Schotman, and Wolff, 1997; Brenner, Harjes, and Kroner, 1996; Christiansen, 2005).

Our approach retains the tractability of the affine models while inheriting the ability of the GARCH models to accurately capture the time variation of yield volatility. The existing literature has concluded that at least three factors are needed to explain term structure dynamics (see for example Litterman and Scheinkman, 1991; Knez, Litterman, and Scheinkman, 1994). Accordingly, we use ATSMs with three latent state variables, with the following dynamics under the physical measure P and the risk-neutral measure Q:

$$X_{t+1} = K_0^P + K_1^P X_t + \sqrt{\Sigma_{t+1}}\epsilon_{t+1}, \tag{2.1}$$

$$X_{t+1} = K_0^Q + K_1^Q X_t + \sqrt{\Sigma_{t+1}}\epsilon_{t+1}, \tag{2.2}$$

$$r_t = \rho_0 + \rho_1 X_t, \tag{2.3}$$

where X_{t+1}, K_0^P and ϵ_{t+1} are 3×1 vectors, and K_1^P is a 3×3 diagonal matrix. r_t denotes the short rate. ρ_0 is a scalar, and ρ_1 is a 1×3 vector. ϵ_{t+1} is assumed to be distributed $N(0,\, I)$. The conditional covariance matrix Σ_{t+1} is a 3×3 diagonal matrix with the ith diagonal element $\sigma_{i,t+1}^2$ governed by a GARCH$(1,1)$ dynamic:

$$\sigma_{i,t+1}^2 = \omega_i + \beta_i \sigma_{i,t}^2 + \alpha_i \epsilon_{i,t}^2, \tag{2.4}$$

where $\epsilon_{i,t}$ is the ith element of vector ϵ_t. ω_i, β_i and α_i are scalars. To ensure that $\sigma_{i,t+1}^2$ is positive, we restrict ω_i, β_i and α_i to be positive numbers. $\sigma_{i,t+1}^2$ is known as of time t, given the history of $X_{i,t}$ and initial variance $\sigma_{i,0}$ as follows

$$\sigma_{i,t+1}^2 = \omega_i + \beta_i \sigma_{i,t}^2 + \alpha_i \frac{(X_{i,t} - K_{0(i)}^P - K_{1(i,i)}^P X_{i,t-1})^2}{\sigma_{i,t}^2}, \tag{2.5}$$

where $X_{i,t}$ is the ith state variable, $K_{0(i)}^P$ is the ith element of K_0^P, and $K_{1(i,i)}^P$ is the ith diagonal element of K_1^P. If $\beta_i = \alpha_i = 0$, the volatility process for the ith state variable is constant over time. The time-varying volatility can be easily filtered from the observations of the state variables.

To link the physical and risk-neutral measures, we specify the pricing kernel to take the form

$$m_{t+1} = \exp(-r_t - \frac{1}{2}\lambda_t^{'}\lambda_t - \lambda_t^{'}\epsilon_{t+1}), \tag{2.6}$$

where λ_t is a 3×1 vector. We use an essentially affine specification for the price of risk (Duffee, 2002; Dai and Singleton, 2002; Cheridito, Filipović, and Kimmel, 2007), which gives:

$$\lambda_t = \left(\sqrt{\Sigma_t}\right)^{-1} (\lambda_0 + \lambda_1 X_t), \tag{2.7}$$

where λ_0 is a 3×1 vector, and λ_1 is a 3×3 diagonal matrix. The P- and Q-parameters in equations (2.1) and (2.2) are therefore related as follows:

$$\begin{aligned} K_0^Q &= K_0^P - \lambda_0, \\ K_1^Q &= K_1^P - \lambda_1. \end{aligned} \tag{2.8}$$

The model-implied price of a zero coupon bond $\widehat{P}_t^n$ with maturity n is given by

$$\widehat{P}_t^n = \exp\left(A_n(\Theta^Q) + B_n^{'}(\Theta^Q)X_t + \sum_i C_{i,n}(\Theta^Q)\sigma_{i,t+1}^2\right), \tag{2.9}$$

where $A_n(\Theta^Q)$, $B_n(\Theta^Q)$ and $C_{i,n}(\Theta^Q)$ are functions of the parameters $\Theta^Q = \{K_0^Q, K_1^Q, \rho_0, \rho_1, \omega_i, \beta_i, \alpha_i\}$ under the Q-dynamics, satisfying the following recursive relations

$$A_n = -\rho_0 + A_{n-1} + B_{n-1}^{'}K_0^Q + \sum_i \left(C_{i,n}\omega_i - \frac{1}{2}\log(1 - 2\alpha_i C_{i,n-1})\right), \tag{2.10}$$

$$B_n = -\rho_1^{'} + B_{n-1}^{'}K_1^Q, \tag{2.11}$$

$$C_{i,n} = \frac{B_{i,n-1}^2}{2(1 - 2\alpha_i C_{i,n-1})} + \beta_i C_{i,n-1}, \tag{2.12}$$

where A_n and $C_{i,n}$ are scalars, and B_n is a 3×1 vector with $B_{i,n}$ is the ith element. The initial conditions are $A_1 = -\rho_0$, $B_1 = -\rho_1^{'}$ and $C_{i,1} = 0$. The derivation of the recursive relations is provided in Appendix A. For the model with a single time-varying volatility factor, $\beta_i = \alpha_i = 0$ for $i = 2$ and 3. Therefore $C_{i,n} = \frac{1}{2}B_{i,n-1}^2$ for $i = 2$ and 3. For the model with two time-varying volatility factors, $\beta_i = \alpha_i = 0$ and $C_{i,n} = \frac{1}{2}B_{i,n-1}^2$ for $i = 3$.

The model-implied continuously compounded n-maturity yield $\widehat{y}_t^n$ is given by

$$\begin{aligned} \widehat{y}_t^n &= \overline{A}_n + \overline{B}_n^{'} X_t + \sum_i \overline{C}_{i,n} \sigma_{i,t+1}^2 \\ &= \overline{A}_n + \overline{B}_n^{'} X_t + \sum_i \overline{C}_{i,n} \left(\omega_i + \beta_i \sigma_{i,t}^2 + \alpha_i \frac{(X_{i,t} - K_{0(i)}^P - K_{1(i,i)}^P X_{i,t-1})^2}{\sigma_{i,t}^2} \right), \end{aligned} \tag{2.13}$$

where $\overline{A}_n = -\frac{A_n}{n}$, $\overline{B}_n^{'} = -\frac{B_n^{'}}{n}$, and $\overline{C}_{i,n} = -\frac{C_{i,n}}{n}$.

We extract the conditional volatilities of yields using the filtered time-series of X_t and the estimated model parameters. The model-implied conditional variance of the n-maturity yield is given by

$$\widehat{var}_t(y_{t+1}^n) = \overline{B}_n^{'} var_t(X_{t+1}) \overline{B}_n + \sigma_e^2, \tag{2.14}$$

where σ_e^2 is the variance of the pricing errors. Appendix B provides the computation of the conditional variance based on the Kalman filter algorithm.

2.2 Canonical ATSMs with Stochastic Volatility

The main benchmark model considered in the paper is the widely used canonical affine stochastic volatility model. Using the classification of Dai and Singleton (2000), we denote the class of affine stochastic volatility models as $A_j(3)$, with $j = 1, 2$ or 3 factors driving the conditional variance of the three state variables. The instantaneous spot interest rate r_t is given by

$$r_t = \rho_0 + \rho_1 X_t, \tag{2.15}$$

where ρ_0 is a scalar, and ρ_1 is a 1×3 vector. Most of this literature uses continuous-time specifications, and we follow this approach.[7] The state variables X_t follows an affine diffusion under the risk-neutral measure Q

$$dX_t = K_{1\Delta}^Q (K_{0\Delta}^Q - X_t)dt + \sqrt{\Sigma_t} dW_t^Q, \tag{2.16}$$

where $K_{0\Delta}^Q$ is a 3×1 vector, $K_{1\Delta}^Q$ is a 3×3 matrix, W_t^Q is a 3×1 vector of independent standard Brownian motions under the risk-neutral measure Q, and Σ_t is the conditional covariance matrix of X_t, and Σ_t is a 3×3 diagonal matrix with the ith diagonal element given by

$$\sigma_{i,t}^2 = a_i + b_i^{'} X_t, \tag{2.17}$$

where a_i is a scalar, and b_i is a 3×1 vector. We define $a = [a_1, a_2, a_3]$, which is a 3×1 vector, and $b = [b_1, b_2, b_3]$, which is a 3×3 matrix. In the $A_1(3)$ model, b_i is a vector of zeros for $i = 2$ and $i = 3$, and in the $A_2(3)$ model, b_i is a vector of zeros for $i = 3$. In the canonical $A_j(3)$ model, all three state variables have a time-varying conditional variance, which is an affine function of $j = 1, 2$ or 3 state variables.

The model-implied price of n-maturity zero coupon bond $\widehat{P}_t^n$ is given by (see Duffie and Kan, 1996)

$$\widehat{P}_t^n = \exp\left(A_n(\Theta^Q) + B_n^{'}(\Theta^Q) X_t\right), \tag{2.18}$$

where $A_n(\Theta^Q)$ and $B_n(\Theta^Q)$ are functions of the parameters under the Q-dynamics, $\Theta^Q = \{K_{0\Delta}^Q, K_{1\Delta}^Q, \rho_0, \rho_1, a, b\}$, through a set of Ricatti ordinary differential equations. The model-implied continuously compounded n-maturity yield $\widehat{y}_t^n$ is given by

$$\widehat{y}_t^n = \overline{A}_n + \overline{B}_n^{'} X_t, \tag{2.19}$$

[7] See Le, Singleton, and Dai (2010) for related discrete-time models.

where $\overline{A}_n = -\frac{A_n}{n}$, and $\overline{B}_n^{'} = -\frac{B_n'}{n}$.

The pricing kernel π_t is given by

$$\frac{d\pi_t}{\pi_t} = -r_t dt - \Lambda_t^{'} dW_t^P, \tag{2.20}$$

where W_t^P is a 3×1 vector of independent standard Brownian motions under the physical measure P, and Λ_t, a 3×1 vector, denotes the market price of risk. We adopt the essentially affine specification for the price of risk as in Duffee (2002) and Dai and Singleton (2002).[8]

$$\Lambda_t = \sqrt{\Sigma_t}\lambda_0 + \sqrt{\Sigma_t^-}\lambda_1 X_t, \tag{2.21}$$

where λ_0 is a 3×1 vector and λ_1 is a 3×3 matrix. In $A_j(3)$ models, the diagonal matrix Σ_t^- has zeros in its first j entries and $(a_i + b_i^{'} X_t)^{-1}$ for $i = j + 1$,...,3.

The dynamics of the state variables under the physical measure P can be written in terms of Λ_t and equation (2.16)

$$dX_t = K_{1\Delta}^Q (K_{0\Delta}^Q - X_t)dt + \sqrt{\Sigma_t}\Lambda_t dt + \sqrt{\Sigma_t} dW_t^P. \tag{2.22}$$

The physical dynamic in the essentially affine model is then given by

$$dX_t = K_{1\Delta}^P (K_{0\Delta}^P - X_t)dt + \sqrt{\Sigma_t} dW_t^P, \tag{2.23}$$

where

$$K_{1\Delta}^P = K_{1\Delta}^Q - \begin{pmatrix} \lambda_{01} b_1^{'} \\ \lambda_{02} b_2^{'} \\ \lambda_{03} b_3^{'} \end{pmatrix} - I^- \lambda_1,$$

[8] Jacobs and Karoui (2009) show that the specification of the price of risk has a minimal impact on modeling conditional volatility.

$$K^P_{1\Delta} K^P_{0\Delta} = K^Q_{1\Delta} K^Q_{0\Delta} + \begin{pmatrix} a_1 \lambda_{01} \\ a_2 \lambda_{02} \\ a_3 \lambda_{03} \end{pmatrix}.$$

We denote element i of λ_0 by λ_{0i}. We define I^- as a 3×3 diagonal matrix. The ith diagonal element $I^-_i = 1$ if the ith diagonal element of Σ^-_t is nonzero. $I^-_i = 0$ if the ith diagonal element of Σ^-_t is zero.

We follow the Dai and Singleton identification scheme to ensure the $\sigma^2_{i,t}$ are strictly positive for all i.[9] The $A_j(3)$ models are different from our proposed model in the parameterization for both the state variables and the volatility process. In the proposed GARCH model, the feedback matrix K^P_1 is a diagonal matrix. In the canonical $A_j(3)$ models, following the admissibility constraints of Dai and Singleton (2000),[10]

$$K^P_{1\Delta} = \begin{bmatrix} K^P_{1\Delta j \times j} & 0_{j \times (3-j)} \\ K^P_{1\Delta (3-j) \times j} & K^P_{1\Delta (3-j) \times (3-j)} \end{bmatrix}, \tag{2.24}$$

where $K^P_{1\Delta j \times j}$ is a $j \times j$ matrix, $K^P_{1\Delta (3-j) \times j}$ is a $(3-j) \times j$ matrix, and $K^P_{1\Delta (3-j) \times (3-j)}$ is a $(3-j) \times (3-j)$ matrix. For the volatility process in equation (2.17),

$$a = \begin{bmatrix} 0_{j \times 1} \\ 1_{(3-j) \times 1} \end{bmatrix}, \tag{2.25}$$

$$b = \begin{bmatrix} I_{j \times j} & b_{j \times (3-j)} \\ 0_{(3-j) \times j} & 0_{(3-j) \times (3-j)} \end{bmatrix}. \tag{2.26}$$

We provide more details on the estimation of these stochastic volatility models in Appendix C.

[9] The identification constraints can be applied to either the P- or Q-parameters, see Dai and Singleton (2000), and Singleton (2006).

[10] We refer to Dai and Singleton (2000) equations 15-19 for details on the admissibility restrictions. Joslin and Le (2013) shows that for no-arbitrage affine term structure models, these admissibility constraints give rise to a tension in simultaneous fitting of the physical and risk-neutral yields.

To facilitate the comparison with our proposed model, we also consider a restricted version of the canonical affine stochastic volatility model as a benchmark model. In the restricted model, we constrain $K^P_{1\Delta}$ and $K^Q_{1\Delta}$ to be diagonal matrices.[11] This model is nested within the canonical specification. The dynamics of the state variables under the restricted model are similar to those in our proposed model. The only difference is the specification of the volatility process. In the restricted $A_j(3)$ model, the conditional variance of the state variables is a linear combination of the levels of $j = 1, 2$ or 3 state variables. The difference with our model is therefore a rather subtle and technical one, and exclusively due to the volatility dynamic.

3 Data and Estimation Method

3.1 Data

We use monthly data on continuously compounded zero-coupon bond yields with maturities of three and six months, and one, two, three, four, five, and ten years. The three- and six-months yields are obtained from the Federal Reserve Economic Data. The one to five, and ten year yields are from the Gürkaynak, Sack, and Wright (2007, GSW) dataset. The sample period is from November 1971 to October 2019.[12]

Table 1 reports the summary statistics of the yields (Panel A) and the EGARCH$(1, 1)$ estimated volatilities of changes in yields (Panel B). The EGARCH$(1, 1)$ is estimated assuming that the conditional mean of changes in monthly yields is generated by an AR(1) process. On average, the yield curve is upward sloping, and the volatility of yields is relatively lower for three- and six-months maturities. The yields for all maturities are highly persistent, and exhibit mild excess kurtosis and positive skewness. The volatilities for all maturities are also very persistent,

[11] To get diagonal $K^Q_{1\Delta}$, restrictions are imposed on the specification of the market price of risk.

[12] The GSW dataset is obtained from the Federal Reserve, available at `http://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html`. We use November 1971 as the start date because it is the earliest date with uninterrupted availability of 10-year yield data.

with slightly higher autocorrelation for short-term yields than for long-term yields. Volatilities exhibit excess kurtosis and positive skewness for all maturities.

3.2 Estimation Method

The term structure model with GARCH volatility can be expressed using a state-space representation. The observed yield curve $y_t = \widehat{y}_t + e_t$ is the measurement equation, where $\widehat{y}_t$ is the model-implied yield as specified in equation (2.13), and e_t is a vector of measurement errors that is assumed to be $i.i.d.$ normal. We assume that the errors of each maturity yield have equal variance σ_e^2 to ensure that all maturities are weighted similarly in the likelihood. The state equation is given by equation (2.1). We apply the Kalman filter to the state-space representation of the model and estimate the parameters $\Theta = \{K_0^P, K_1^P, K_0^Q, K_1^Q, \rho_0, \rho_1, \omega_i, \beta_i, \alpha_i\}$ and filter the state variables X_t using the quasi-maximum likelihood (QML) method. The log likelihood of the tth observation is

$$\begin{aligned} \log f_t(\Theta) \;=\; & -\frac{N}{2}\log(2\pi) - \log(|\det(J)|) - \frac{N}{2}\log(\sigma_e^2) - \frac{1}{2}\frac{\|e_t\|^2}{\sigma_e^2} - \frac{1}{2}\log(\det(\Sigma_t)) \\ & -\frac{1}{2}(X_t - K_0^P - K_1^P X_{t-1})'\Sigma_t(X_t - K_0^P - K_1^P X_{t-1}). \end{aligned} \tag{3.1}$$

N denotes the number of available yields in the term structure. In our sample, $N = 8$. The Jacobian term is given by

$$J = \begin{pmatrix} I_3 & 0_{3\times 8} \\ \overline{B} & \Sigma_e \end{pmatrix}, \tag{3.2}$$

where Σ_e is a 8×8 diagonal matrix with σ_e^2 as the diagonal term. Recall that $\overline{B}_n(\Theta^Q) = -\frac{B_n(\Theta^Q)}{n}$. $\overline{B}_n$ is a 3×1 vector, and $\overline{B}$ is a $3\times N$ matrix. $\|e_t\|$ denotes the Euclidean norm of the vector of measurement errors. Appendix B provides more detailed information on the Kalman filter algorithm.[13] The estimation of the benchmark $A_1(3)$ model is discussed in Appendix C.

4 Empirical Results

To retain a low-dimensional structure for the affine term structure model, we focus on a model with one time-varying volatility factor in our empirical analysis. That is, we set $i = 1$ in equation (2.4) in the proposed term structure model with GARCH volatility. Only the first state variable has a time-varying variance. We refer to this model as the no-arbitrage GARCH model. Our main benchmark model is therefore the canonical $A_1(3)$ model with one factor driving the conditional variances of the state variables. Dai and Singleton (2000) conclude that this model offers the best characterization of unconditional yield volatilities and a sufficiently flexible correlation structure among three-factor models with stochastic volatility. Note that this model has a time-varying variance for all three state variables. However, the dynamic of the variance is driven by only one state variable. We also consider a restricted version of the canonical $A_1(3)$ model, where the feedback matrix of the state variables is diagonal, as in the no-arbitrage GARCH model.

In this section, we first present the parameter estimates of the models. Then we focus on comparing the models' ability to predict the conditional volatility of the yield curve. Subsequently, we examine the unconditional volatility implied by these models. We also document the tension between matching yields and yield volatilities in these models, and we discuss their implications for the expectations hypothesis.

4.1 Parameter Estimates

Table 2 presents parameter estimates for the no-arbitrage GARCH model (Panel A), the canonical $A_1(3)$ model (Panel B), and the restricted $A_1(3)$ model (Panel C). The no-arbitrage GARCH model has 22 parameters, including the variance σ_e^2 of the measurement errors. The canonical

[13] See Duffee and Stanton (2012) and Christoffersen, Dorion, Jacobs and Karoui (2014) for estimation using the Kalman filter.

$A_1(3)$ model has 24 parameters, and the restricted $A_1(3)$ model has 14 parameters. In the no-arbitrage GARCH model and the restricted $A_1(3)$ model, the conditional mean under the P-measure is more restricted than in the canonical $A_1(3)$ model. For all models, the estimates satisfy the admissibility conditions under the physical measure P. The state variables in the canonical and restricted $A_1(3)$ models follow a first order VAR process when sampled monthly. To facilitate the comparison with the estimates from the discrete-time no-arbitrage GARCH model, we report the estimated parameters of the discretized VAR process for both benchmark models.

The time series properties of the state variables critically depend on the speed of mean reversion in the feedback matrix K_1^P. For the no-arbitrage GARCH model, the first state variable, which has time-varying volatility, is highly persistent. The second and third state variables are less persistent than the first. The third state variable reverts more quickly under the Q-measure, indicating that the bond loadings on this factor decay rapidly as maturity increases.

Some of the implications of the canonical and restricted $A_1(3)$ models are similar to the no-arbitrage GARCH model. Once again the first state variable is the most persistent and plays the role of level factor, whereas the third variable is strongly mean reverting. However, in contrast to the no-arbitrage GARCH model, the third factor is more mean-reverting under the P-measure for the canonical and restricted $A_1(3)$ models. Also, the effect of the second factor on the short rate is negative in the no-arbitrage GARCH model but positive in the $A_1(3)$ models.

For the volatility process, the no-arbitrage GARCH model requires a persistent volatility dynamic for the level factor, which drives the volatility dynamics of the term structure of yields. The estimated autocorrelation coefficient is $\beta = 0.9064$. In the canonical and restricted $A_1(3)$ models on the other hand, the level factor mainly affects the volatility of the slope of the yields. Figure 1 plots the time series of the conditional variance for the level factor $\sigma_{1,t}^2$ implied by the different models. For comparison, we also plot the EGARCH$(1,1)$ estimated variance of the 3-month yield. The EGARCH model is estimated assuming that the conditional mean of changes in 3-month yield follows an AR(1) process. The variance factor from the no-arbitrage GARCH model has the highest unconditional correlation with the EGARCH estimates (84%). The correlations for the two benchmark models are approximately 54%.[14]

4.2 Conditional Yield Volatility

In this section, we examine the properties of the model-implied conditional yield volatilities. We present the model-implied one-month conditional volatilities together with the EGARCH(1, 1) volatilities in Figure 2. Appendices B and C discuss the derivation of the implied one-month conditional volatilities from the no-arbitrage GARCH model and the $A_1(3)$ model respectively. The EGARCH model is estimated assuming that the conditional mean of changes in monthly yields follows an AR(1) process. The estimated conditional volatilities from the canonical and restricted $A_1(3)$ models are very similar and much less variable than the EGARCH volatilities. At the very short end (3-month and 6-month) and very long end (10-year) of the yield curve, the estimated conditional volatilities from the restricted $A_1(3)$ model exhibit excess movement in the last decade of the sample. The estimated volatility at the short end is similar before and after the monetary experiment in early 1980s for both benchmark models. In addition, both benchmark models overestimate yield volatility when volatility is low, from the mid-1980s to 2000. Both benchmark models also do not exhibit sufficient yield volatility at the beginning of the sample.

The estimated conditional volatilities from the no-arbitrage GARCH model are more variable than the estimates from the two benchmark models. The estimates from the no-arbitrage GARCH model comove closely with the EGARCH volatilities at longer maturities (4-5 and 10 years), but do not perform as well for short maturities. For longer maturities, the no-arbitrage GARCH model fits the high volatility periods of the early 1980s well. The canonical and re-

[14] In the robustness section, we use realized volatility as an alternative measure of the true variance. Andersen and Benzoni (2010) and Christensen, Lopez, and Rudebusch (2014) use realized volatility as a benchmark to compare the volatility fit of ATSMs. Cieslak and Povala (2016) use realized covariance to extract stochastic volatility from a term structure model with multivariate volatility components.

stricted $A_1(3)$ models, on the other hand, cannot capture the high volatility periods for any maturity. The no-arbitrage GARCH model also does a better job in fitting the periods of low volatility in our sample (from the mid-1980s to 2000) than the two benchmark models for all maturities. Moreover, the no-arbitrage GARCH model is able to capture the time variation in yield volatility at the beginning of our sample for all maturities. In summary, the no-arbitrage GARCH model appears to capture the time-variation in the second moment of yields quite well, especially at longer maturities.

To further assess the quality of the estimated conditional volatilities, Table 3 reports the unconditional correlation between model-implied and EGARCH$(1, 1)$ volatilities. A first observation is that the correlation between model-implied and EGARCH volatilities is positive at all maturities for all models. Also, the correlations are slightly lower at the very short- and long-end of the yield curve for all models. The estimated conditional volatilities from the no-arbitrage GARCH model have the highest correlation with the EGARCH estimates at all maturities. For example, the unconditional correlation is as high as 95% for the 3-year yield volatility, while it is about 71% for the two benchmark models. The two benchmark models have similar correlations with the EGARCH model at all maturities. On average across all maturities, the unconditional correlation for the no-arbitrage GARCH model is 90%, for the canonical $A_1(3)$ model it is 67%, and for the restricted $A_1(3)$ model it is 66%. As discussed in Jacobs and Karoui (2009), the performance of the stochastic volatility model in fitting yield volatility is sensitive to the model and sample under consideration. We find a positive correlation as in Jacobs and Karoui (2009), because we also use a relative long sample of Treasury yields and include the high inflation period. Andersen and Benzoni (2010), and Collin-Dufresne, Goldstein, and Jones (2009) do not find a significant positive correlation.

To provide additional insight into the models' ability to fit volatility, we also examine the root mean squared errors (RMSEs) and the mean absolute percentage errors (MAPEs) between model-implied and EGARCH volatilities. Panel A of Table 4 reports the RMSEs in basis points, and Panel B of Table 4 reports the MAPEs in percentages. The no-arbitrage GARCH model outperforms both benchmark models in fitting the volatility across all maturities for both performance measures. The model does a particularly good job at the intermediate and long end of the yield curve, although it only has the single volatility factor. For example, for 3-5 and 10 years, the RMSEs are below 10 basis points.

The RMSE improvement of the no-arbitrage GARCH model over the two benchmark models is about 33% on average across maturities. The MAPEs improvement of the no-arbitrage GARCH model over the canonical $A_1(3)$ model is about 39% on average across maturities. The improvement over the restricted $A_1(3)$ model on average across maturities is about 37%. Moreover, the improvement over the two $A_1(3)$ models for both fit measures is more significant at longer maturities. For example, for 5-year yield, the improvement in RMSEs is about 55%, and the improvement in MAPEs is about 60%. In summary, the no-arbitrage GARCH model performs better in capturing the time variability of conditional volatility despite using a single volatility factor. This finding suggests that the improvements mainly result from the GARCH specification of the volatility process.

4.3 The Term Structure of Unconditional Volatility

We investigate model-implied unconditional volatility in our sample. We present the term structure of unconditional yield volatility implied by the EGARCH$(1, 1)$ model, the no-arbitrage GARCH model, and the canonical and restricted $A_1(3)$ models. The unconditional model volatilities are computed as the averages of the conditional volatility path generated by each model. Figure 3 shows that the no-arbitrage GARCH implied term structure matches the EGARCH$(1, 1)$ implied term structure much better than the benchmark models. The term structures implied by the two benchmark models are very similar. The volatility curve from these models monotonically decreases as a function of maturity. In contrast, the no-arbitrage GARCH model exhibits a non-monotonic pattern as a function of the term structure, consistent with the pattern observed in the data. In summary, the no-arbitrage GARCH model does an excellent job fitting the term structure of unconditional yield volatility.

4.4 The Trade-off Between Fitting Yield Levels and Volatilities

The tension between matching the first and second moments of Treasury yields under the ATSMs has been documented in many studies (Dai and Singleton, 2000, 2002; Duffee, 2002; Duarte, 2004; Joslin and Le, 2013). In particular, Dai and Singleton (2002) note "a tension in matching simultaneously the historical properties of the conditional means and variances of yields". Joslin and Le (2013) study the exact mechanism that underlies this tension, and argue that imposing a spanning condition could prevent a no-arbitrage model from fully capturing the predictability patterns of bond yields in the data.

The no-arbitrage GARCH model belongs to the class of spanned affine models. Since the GARCH volatility is one of the factors that determine bond yields, it is spanned by the yields. The estimation of this model should therefore be subject to the same tension. We now show that the improved volatility fit is not obtained at the expense of poor yield fit, by examining the tension between the fit of the first and second moments in two ways. First, we compare the models' ability to fit yields. Second, we examine the models' ability to capture patterns in the data from the perspective of the expectations hypothesis.

4.4.1 Yield Fit

Panel A of Table 5 reports the RMSEs of yields in basis points for the no-arbitrage GARCH model and the two benchmark models. On average across all maturities, the fits of yields are similar for the three models. The in-sample RMSE of yields for the no-arbitrage GARCH model with a single volatility factor is about 19 basis points on average across different maturities. This finding suggests that the improvement of the no-arbitrage GARCH model in fitting conditional volatility, as shown in Section 4.2, does not come at the cost of fitting the conditional mean of yields.

For comparison, we also present the fit of yields for Gaussian models with constant variance-covariance matrix. We consider the canonical representation of Joslin, Singleton, and Zhu (2011, henceforth referred to as JSZ), which allows for stable and tractable estimation of the $A_0(3)$ three-factor Gaussian model.[15] Panel B of Table 5 shows the RMSEs for the maximum flexible specification of the JSZ model and also for three restricted models. In the first restricted model, we use a diagonal variance-covariance matrix with constant variance. In the second restricted model, we set the $(1,2)$ and $(1,3)$ entries of the feedback matrix (K_1) to zero and we also restrict the variance-covariance matrix to be diagonal and the variance to be constant. This restricted version is comparable to the canonical $A_1(3)$ model. In the third restricted JSZ specification, we restrict the feedback matrix (K_1) and the constant variance-covariance matrix to be diagonal. This restricted form is more comparable to the no-arbitrage GARCH model. Recall that under the newly proposed model, the feedback matrix is a diagonal matrix.

The maximum flexible $A_0(3)$ model provides the best in-sample fit of yields. This is not surprising, since the model has the richest specification for the conditional mean of the state variables. All three restricted forms have marginally higher average RMSEs than the maximum flexible model.

Overall the Gaussian models outperform the models with time-varying volatility in Panel A for the purpose of fitting yields. However, the Gaussian models are unable to capture the time variation of yield volatility. These findings show that a tension remains between matching yields and yield volatilities in the proposed model. In comparison to the Gaussian models, all models in Panel A sacrifice the fitting of the cross-section of yield levels to acquire flexibility in fitting conditional variances. This is inevitable, since in these non-Gaussian models, the state variables driving both yields and yield volatilities are the same and are spanned by the cross section of yields. In the canonical $A_1(3)$ model, all three state variables are non-Gaussian. The

[15] We refer to JSZ (2011) for implementation details.

non-Gaussian state variables must be positive and enter the conditional variance. As discussed in Joslin and Le (2013), this admissibility constraint creates an asymmetry between fitting the yields and yield volatilities. In the no-arbitrage GARCH model, we have a different specification for the factor with time-varying volatility. The variance of this factor follows a GARCH specification, which results in a very simple admissibility constraint. This simpler constraint does not seem to result in a deterioration of the fit for the conditional mean of yields compared to the benchmark models.

4.4.2 The Expectations Hypothesis

In-sample fit is an important aspect of models' ability to capture the level of yields. We now examine the ability of the models to capture the time series properties of the yields data by focusing on the predictability patterns observed in the data. Campbell and Shiller (1991) show that under the expectations hypothesis, a regression coefficient of $\varphi_n = 1$ obtains in the following regression

$$y_{t+1}^{n-1} - y_t^n = \psi_n + \varphi_n \left(\frac{y_t^n - y_t^1}{n-1} \right) + e_{t+1}^n, \tag{4.1}$$

where y_t^n is the n-month yield at time t. However, Dai and Singleton (2002) show that actual estimates are negative and more so for longer maturities. They then show that the Gaussian model is consistent with this downward sloping pattern in the data. However, no-arbitrage stochastic volatility models are not able to match this pattern.

We conduct this regression analysis using our sample and the yields implied by the different models we investigate. Figure 4 presents the results. We find deviations from the expectations hypothesis, consistent with the existing literature. The estimated coefficients are all negative, and more negative for shorter maturities. Consistent with Dai and Singleton (2002), we find that the Gaussian model is able to capture this pattern, but the canonical and restricted $A_1(3)$ models are not. In contrast to the canonical stochastic volatility models, the proposed no-arbitrage GARCH model can match the empirical patterns of bond risk premia as characterized by the regression coefficients. These results confirm that the improvements provided by the no-arbitrage GARCH for the purpose of fitting yield volatilities do not come at the expense of matching the time series properties in the yields data. By specifying variance as a GARCH process within a no-arbitrage model, we are able to rationalize the deviations from the expectations hypothesis observed in the data.

5 Robustness

In this section, we first investigate the robustness of our findings when estimating the model exclusively based on yield volatility. Subsequently we use realized volatility as an alternative to the EGARCH model to evaluate models' performance in matching conditional volatilities of yields. Finally, we discuss the performance of no-arbitrage GARCH models with multiple volatility factors.

5.1 Estimation Based on Yield Volatility

We provide additional evidence on the performance of the no-arbitrage GARCH model by estimating with an objective function that is exclusively based on yield volatilities. We apply the Kalman filter to the state-space representation of the model. The parameters are estimated and the state variables are filtered by minimizing the sum squared errors S

$$S = \sum_{t=1}^{T} \sum_{n=1}^{N} \left(var_t(y_{t+1}^n) - \widehat{var}_t(y_{t+1}^n) \right)^2 , \tag{5.1}$$

where $var_t(y_{t+1}^n)$ represents the EGARCH$(1,1)$ estimated variance of changes in n-maturity yield. The cross sectional fit of the yield volatilities is the only objective in this estimation exercise, which therefore removes the tension inherent in the models between estimating yield levels and volatilities.

Figure 5 plots the implied one-month conditional volatilities together with the EGARCH$(1,1)$

volatilities. The no-arbitrage GARCH model provides the best fit. The estimated conditional volatilities from the proposed model match the EGARCH$(1, 1)$ estimates very well. The two benchmark models cannot capture the high volatility periods well for all maturities. At the long end of the yield curve, the canonical and restricted $A_1(3)$ models imply excess volatility for the last decade of the sample. These findings are consistent with those from the likelihood-based estimation discussed in Section 4.2. The two benchmark models are unable to capture the volatility well even when the objective of the estimation is exclusively to match volatility.

We also present the unconditional correlations and the measures of volatility fit for these models in Table 6. Panel A shows that the implied volatilities from the no-arbitrage GARCH model have the highest unconditional correlations with the EGARCH estimates for all maturities. On average across maturities, the correlation is approximately 86% for the no-arbitrage GARCH model. Compared to the results in Table 3, which are based on maximum likelihood estimation, the correlations for the no-arbitrage GARCH model are slightly higher at the short end, but marginally lower at the intermediate and long end of the yield curve. For the two benchmark models, the correlations are smaller than those in Table 3 for all maturities.

The no-arbitrage GARCH model also provides the best fit in terms of the RMSEs and MAPEs of yield volatilities. The volatility-based objective function leads to substantial improvements for the no-arbitrage GARCH model in terms of MAPEs. Figure 2 shows that volatility from the likelihood-based estimation exceeds the EGARCH estimates most of the time in our sample, especially for short maturities. When we use volatility information in the estimation, this bias is avoided. This leads to a significant improvement in MAPEs. The improvement of the no-arbitrage GARCH model over the canonical $A_1(3)$ model on average across maturities is about 41% for RMSEs and 50% for MAPEs. These improvements exceed those from the maximum likelihood estimation.

In summary, when estimating the models using an objective function based on yield volatilities, the no-arbitrage GARCH model continues to outperform the two benchmark models. This finding is consistent with our findings in Section 4.2. The no-arbitrage GARCH is better suited to model yield volatility regardless of the information used in the estimation.

5.2 Matching Realized Volatility

In this section, we use realized volatility instead of EGARCH volatility as the measure of model-free yield volatility. We do not have high-frequency data available to construct realized volatility for an extended sample period. We therefore follow the technique pioneered by Schwert (1989) in the equity return literature and construct measures of monthly realized volatility using within-month squared changes in yields. Assuming that M observations are available within a month, the estimate of the monthly variance for the n-maturity yield is computed as

$$var(y_{t+1}^n) = \sum_{m=1}^{M} \left(\Delta y_{t+m/M}^n\right)^2 . \tag{5.2}$$

It has been shown that an ARMA$(1,1)$ provides a good fit to the logarithm of the realized variance:

$$\log(var(y_{t+1}^n)) = \gamma \log(var(y_t^n)) + \delta\varepsilon_t + \varepsilon_{t+1}, \tag{5.3}$$

where ε_{t+1} is assumed to be distributed $N(0,\, \sigma_{t,\varepsilon}^2)$. We refer this model as the realized variance model. The model-implied one-month conditional variance of the n-maturity yield is then given by

$$\widehat{var}_t(y_{t+1}^n) = (var(y_t^n))^\gamma \exp\left(\delta\varepsilon_t + \frac{\sigma_{t,\varepsilon}^2}{2}\right). \tag{5.4}$$

Figure 6 presents the one-month conditional volatilities implied by the realized variance model and the three models with time-varying volatility.[16] The estimates from the no-arbitrage GARCH model comove closely with those from the realized variance model. Both benchmark models overestimate yield volatilities for most of the sample between 1980 to 2000. These findings are

[16] We plot the results for one- to five- and ten-year maturities, because three- and six-month daily yields are not available for our sample.

consistent with those in Figure 2 in which EGARCH(1, 1) is used as a measure of model-free volatility. This is not surprising because the two measures are highly correlated. Panel A of Table 7 shows that the unconditional correlation between the two measures is on 87% on average across maturities. The two measures major differ somewhat during the high volatility periods and for one- and two-year maturity yields: the EGARCH estimates exceed the estimates from the realized variance model.

Panel A of Table 7 also reports the unconditional correlations between the estimate of realized variance and the variance estimates implied by the models. Correlations are again positive at all maturities for all models,but the he no-arbitrage GARCH model has a much higher unconditional correlation than the canonical and restricted $A_1(3)$ models for all maturities. Panels B and C of Table 7 present the RMSEs and the MAPEs of model-implied volatilities when realized variance is used as a model-free variance measure. Consistent with the findings in Table 4, the no-arbitrage GARCH model outperforms both the canonical and restricted $A_1(3)$ models in fitting conditional volatility for all maturities and for both performance measures.

Overall, we conclude that our results in Section 4.2 are robust to the choice of measure of model-free yield volatility. The no-arbitrage GARCH model performs well for the purpose of modeling conditional volatility.

5.3 Models with Multiple Volatility Factors

In this section, we investigate the performance of no-arbitrage GARCH models with multiple volatility factors. Figure 7 plots the model-implied one-month conditional volatilities together with the EGARCH(1, 1) volatilities. The performance of no-arbitrage GARCH models with two and three volatility factors is similar to that of the model with a single volatility factor. The implied conditional volatilities of yields are closely related to the EGARCH estimates, especially at longer maturities. This finding suggests that the first volatility factor plays a dominant role in fitting conditional yield volatility. Table 8 presents the parameter estimates for the no-arbitrage

GARCH models with two and three volatility factors as well as the log-likelihood. The estimates of the feedback matrix for the state variables are similar to those in Table 2 for the model with one volatility factor. The likelihood ratio tests indicate that there is no strong evidence that the models with two and three volatility factors fit the data better than the model with one volatility factor.

In the model with three volatility factors, the second and third volatility factors mean revert more quickly than the first factor. The estimated mean reversion parameters for the three volatility factors are $\beta_1 = 0.9021$, $\beta_2 = 0.7314$, and $\beta_3 = 0.8431$ respectively. Figure 8 plots the time series of the three variance factors $\sigma^2_{i,t}$ as well as the EGARCH variance for the 3-month yield. The first variance factor has the highest unconditional correlation with the EGARCH estimates (83.41%). For the second and third variance factors, the correlations are 64.01% and 61.40% respectively. However, note that the second variance factor is scaled by $1e^7$ and the third variance factor is scaled by $1e^4$ in Figure 8. These variance factors exhibit substantial time variation, but their magnitudes are small.

6 Conclusion

In the term structure literature, state-of-the-art models face difficulties in simultaneously fitting the time variation in yield levels and volatilities. We propose a parsimonious yet flexible class of models with closed-form solutions that outperform benchmark models in this dimension. Our results suggest that the performance of ATSMs in matching yield volatility critically depends on the specification of the volatility dynamics. In standard ATSMs with stochastic volatility, the volatility dynamic is a linear combination of the *levels* of the yield curve factors. We instead propose a no-arbitrage term structure model where the volatility factor is written as a function of the (lagged) squared *innovations* to the yields.

The model combines the tractability of ATSMs with improved modeling of yield volatility. We estimate the model using monthly yield data from 1971 to 2019, and find that the model-implied conditional volatility performs well, especially at longer maturities. The correlation between model-implied and model-free yield volatility is between 85% and 95%. The proposed model significantly outperforms benchmark stochastic volatility models for the purpose of fitting yield volatility. The model also provides a good fit to the conditional mean of yields, suggesting that the improved volatility fit is not obtained at the expense of yield fit. These findings are robust to various variations in the empirical setup.

It is worth emphasizing that our approach may not the only one that provides improved modeling of conditional volatility. Indeed it may be possible to construct better models, and we plan to address this in future work. Our objective in this paper is merely to show that it is possible to write down parsimonious term structure models that allow for improved modeling of yield volatility as compared to state-of-the art models in this literature, without sacrificing the model's ability to fit the level of yields.

Appendix A. Bond Valuation in ATSMs with GARCH Volatility

To derive the recursions in equations (2.10), (2.11) and (2.12), we first note that the price of a one-period bond, $n = 1$, is as follows

$$\begin{aligned} P(t, t+1) &= E_t^Q \left[\exp(-r_t) \right] \\ &= \exp(-\rho_0 - \rho_1 X_t). \end{aligned} \tag{A.1}$$

Suppose that the price of a n-period bond is given by $P_t^n = \exp\left(A_n + B_n' X_t + \sum_i C_{i,n} \sigma_{i,t+1}^2 \right)$. Matching coefficients gives $A_1 = -\rho_0$, $B_1 = -\rho_1'$ and $C_{i,1} = 0$. In order to solve for A_n, B_n and

$C_{i,n}$ we derive the bond price under the risk neutral probability measure

$$
\begin{aligned}
P_t^n &= E_t^Q[\exp(-r_t)P_{t+1}^{n-1}] \qquad (A.2)\\
&= E_t^Q\left[\exp\left(-\rho_0-\rho_1 X_t\right)\exp\left(A_{n-1}+B_{n-1}^{'}X_{t+1}+\sum_i C_{i,n-1}\sigma_{i,t+2}^2\right)\right]\\
&= \exp\left(-\rho_0-\rho_1 X_t+A_{n-1}+B_{n-1}^{'}(K_0^Q+K_1^Q X_t)+\sum_i C_{i,n-1}\omega_i+\sum_i C_{i,n-1}\beta_i\sigma_{i,t+1}^2\right)\\
&\quad\times E_t^Q\left[\exp\left(\sum_i\left(B_{i,n-1}\sigma_{i,t+1}\epsilon_{i,t+1}+C_{i,n-1}\alpha_i\epsilon_{i,t+1}^2\right)\right)\right].
\end{aligned}
$$

Completing the square in the portion to which the expectation applies, using the fact for a standard normal z, $E\left(a(z+b)^2\right)=\exp\left(-\frac{1}{2}\log(1-2a)+\frac{ab^2}{1-2a}\right)$ and matching coefficients results in the recursive relations in equations (2.10), (2.11) and (2.12).

Appendix B. Conditional Volatility in ATSMs with GARCH Volatility

This appendix summarizes the derivation of the conditional variance for the affine model in a GARCH framework. The contemporaneous forecast of the state vector and its corresponding covariance matrix are denoted by $X_{t|t}$ and $P_{t|t}$. The Kalman filter algorithm works as follows at any time t:

1. Given $X_{t|t}$ and $P_{t|t}$, compute the one-period ahead forecast of the state vector and its corresponding covariance matrix[17]

$$X_{t+1|t}=K_0^P+K_1^P X_t, \qquad (B.1)$$

[17] The unconditional two first moments are used in the first step of recursion.

and

$$P_{t+1|t} = K_1^{P'} P_{t|t} K_1^P + \Sigma_{t+1|t+1}. \tag{B.2}$$

The volatility factor can be computed based on equation (2.5)

$$\sigma_{i,t+2|t}^2 = \omega_i + \beta_i \sigma_{i,t+1|t}^2 + \alpha_i \frac{(X_{i,t+1|t} - K_{0(i)}^P - K_{1(i,i)}^P X_{i,t})^2}{\sigma_{i,t+1|t}^2}, \tag{B.3}$$

where

$$\sigma_{i,t+1|t}^2 = \omega_i + \beta_i \sigma_{i,t|t}^2 + \alpha_i \frac{(X_{i,t} - K_{0(i)}^P - K_{1(i,i)}^P X_{i,t-1})^2}{\sigma_{i,t|t}^2}.$$

$\sigma_{i,t|t}^2$ is the ith diagonal element of the 3×3 matrix $\Sigma_{t|t}$.

2. Compute the one-period ahead forecast of y_{t+1} and its corresponding covariance matrix

$$\begin{aligned} y_{t+1|t} &= \overline{A} + \overline{B}' X_{t+1|t} + \sum_i \overline{C}_i \sigma_{i,t+2|t}^2 \\ &= \overline{A} + \overline{B}' X_{t+1|t} + \sum_i \overline{C}_i \left(\omega_i + \beta_i \sigma_{i,t+1|t}^2 + \alpha_i\right), \end{aligned} \tag{B.4}$$

where $y_{t+1|t}$ is a $N \times 1$ vector, $\overline{A}$ and $\overline{C}_i$ are $N \times 1$ vectors, and $\overline{B}$ is a $3 \times N$ matrix, N denotes the number of available yields in the term structure.

$$V_{t+1|t} = \overline{B}' P_{t+1|t} \overline{B} + R, \tag{B.5}$$

where R is a $N \times N$ diagonal matrix. The diagonal terms are the same, as denoted by σ_e^2. We assume that the pricing errors on each maturity have equal variance σ_e^2.

3. Compute the forecast error of y_{t+1}, $e_{t+1|t} = y_{t+1} - y_{t+1|t}$.

4. Update the contemporaneous forecast of the state vector and its corresponding covariance

matrix

$$X_{t+1|t+1} = X_{t+1|t} + P_{t+1|t}\overline{B}V_{t+1|t}^{-1}e_{t+1|t}, \tag{B.6}$$

$$P_{t+1|t+1} = P_{t+1|t} - P_{t+1|t}\overline{B}V_{t+1|t}^{-1}\overline{B}^{'}P_{t+1|t}. \tag{B.7}$$

and compute the smoothed volatility factor

$$\sigma_{i,t+2|t+1}^2 = \omega_i + \beta_i\sigma_{i,t+1}^2 + \alpha_i\frac{(X_{i,t+1|t+1} - K_{0(i)}^P - K_{1(i,i)}^P X_{i,t})^2}{\sigma_{i,t+1}^2}. \tag{B.8}$$

5. Return to the first step.

The log quasi-likelihood of observation $t+1$ is then

$$\log f_t(\Theta) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log(\det(V_{t+1|t})) - \frac{1}{2}e_{t+1|t}^{'}V_{t+1|t}e_{t+1|t}. \tag{B.9}$$

The model-implied conditional variance of yields is computed using the filtered state vector $X_{t|t}$

$$\widehat{var}_t(y_{t+1}) = diag(V_{t+1|t}). \tag{B.10}$$

In the empirical investigation, we use the conditional variance of yield differences, which is also equal to equation (B.10).

Appendix C. Estimation of Canonical ATSMs with Stochastic Volatility

This appendix summarizes the estimation method used for canonical affine stochastic volatility models. A three-factor latent model can be expressed as a state-space representation. The observed yield curve $y_t = \widehat{y}_t + e_t$ is the measurement equation, where $\widehat{y}_t$ is the model-implied yield as specified in equation (2.19), and e_t is a vector of measurement errors that is assumed to be *i.i.d.* normal. We assume that the errors on each maturity have equal variance σ_e^2 so that the likelihood tries equally hard to match the yield curve. The state equation (2.23) can be discretized as $X_{t+1} = K_0^P + K_1^P X_t + \epsilon_{t+1}^P$, where $\epsilon_{t+1|t}^P$ is assumed to be distributed $N(0, \Sigma_t)$. We estimate the P- and Q-parameters simultaneously by applying the Kalman filter to the state-space representation. We use the quasi-maximum likelihood (QML) method as implemented by Jacobs and Karoui (2009).

The contemporaneous forecast of the state vector and its corresponding covariance matrix are denoted by $X_{t|t}$ and $P_{t|t}$. The Kalman filter algorithm works as follows at any time t:

1. Given $X_{t|t}$ and $P_{t|t}$, compute the one-period ahead forecast of the state vector and its corresponding covariance matrix[18]

$$X_{t+1|t} = K_0^P + K_1^P X_t, \tag{C.1}$$

and

$$P_{t+1|t} = K_1^{P'} P_{t|t} K_1^P + \Sigma_{t|t}. \tag{C.2}$$

The volatility factor can be computed based on equation (2.17)

$$\sigma_{i,t|t}^2 = a_i + b_i' X_t, \tag{C.3}$$

where $\sigma_{i,t|t}^2$ is the ith diagonal element of the 3×3 matrix $\Sigma_{t|t}$.

2. Compute the one-period ahead forecast of y_{t+1} and its corresponding covariance matrix

$$y_{t+1|t} = \overline{A} + \overline{B}' X_{t+1|t}, \tag{C.4}$$

where $y_{t+1|t}$ is a $N \times 1$ vector, $\overline{A}$ is a $N \times 1$ vector, and $\overline{B}$ is a $3 \times N$ matrix, N denotes

[18] The unconditional two first moments are used in the first step of recursion.

the number of available yields in the term structure.

$$V_{t+1|t} = \overline{B}^{'} P_{t+1|t} \overline{B} + R, \tag{C.5}$$

where R is a $N \times N$ diagonal matrix. The diagonal terms are the same, as denoted by σ_e^2. We assume that the pricing errors on each maturity have equal variance σ_e^2.

3. Compute the forecast error of $y_{t+1,}$ $e_{t+1|t} = y_{t+1} - y_{t+1|t}$.

4. Update the contemporaneous forecast of the state vector and its corresponding covariance matrix

$$X_{t+1|t+1} = X_{t+1|t} + P_{t+1|t} \overline{B} V_{t+1|t}^{-1} e_{t+1|t}, \tag{C.6}$$

and

$$P_{t+1|t+1} = P_{t+1|t} - P_{t+1|t} \overline{B} V_{t+1|t}^{-1} \overline{B}^{'} P_{t+1|t}. \tag{C.7}$$

5. Return to the first step.

The log quasi-likelihood of observation $t+1$ is then

$$\log f_t(\Theta) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log(\det(V_{t+1|t})) - \frac{1}{2} e_{t+1|t}^{'} V_{t+1|t} e_{t+1|t}. \tag{C.8}$$

The model-implied conditional variance of yields is computed using the filtered state vector $X_{t|t}$

$$\widehat{var}_t(y_{t+1}) = diag(V_{t+1|t}). \tag{C.9}$$

In the empirical investigation, we use the conditional variance of yield differences, which is also equal to equation (C.9).

References

[1] Ahn, D. H., Dittmar, R. F., and Gallant, A. R. (2002). Quadratic term structure models: Theory and evidence. The Review of financial studies, 15(1), 243-288.

[2] Ahn, D. H., Dittmar, R. F., Gallant, A. R., and Gao, B. (2003). Purebred or hybrid?: Reproducing the volatility in term structure dynamics. Journal of Econometrics, 116(1-2), 147-180.

[3] Ahn, D. H., and Gao, B. (1999). A parametric nonlinear model of term structure dynamics. The Review of Financial Studies, 12(4), 721-762.

[4] Anderson, T., and Benzoni, L. (2010). Do bonds span volatility risk in the US treasury market? A specification test for affine monetary policy regimes and the term structure of interest rates. Journal of Finance, 65, 603-653.

[5] Ang, A., and Bekaert, G. (2002). Regime switches in interest rates. Journal of Business & Economic Statistics, 20(2), 163-182.

[6] Bansal, R., Tauchen, G., & Zhou, H. (2004). Regime-Shifts in Term Structure, Expectations Hypothesis Puzzle, and the Real Business Cycle. Journal of Business and Economic Statistics, 22(4), 396-409.

[7] Bansal, R., and Zhou, H. (2002). Term structure of interest rates with regime shifts. The Journal of Finance, 57(5), 1997-2043.

[8] Bikbov, R., and Chernov, M. (2010). Yield curve and volatility: Lessons from eurodollar futures and options. Journal of Financial Econometrics, nbq019.

[9] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of econometrics, 31(3), 307-327.

[10] Brenner, R. J., Harjes, R. H., and Kroner, K. F. (1996). Another look at models of the short-term interest rate. Journal of Financial and Quantitative Analysis, 31(01), 85-107.

[11] Campbell, J. Y., and Shiller, R. J. (1991). Yield spreads and interest rate movements: A bird's eye view. The Review of Economic Studies, 58(3), 495-514.

[12] Cheridito, P., Filipović, D., and Kimmel, R. L. (2007). Market price of risk specifications for affine models: theory and evidence. Journal of Financial Economics, 83(1), 123-170.

[13] Christensen, J. H., Lopez, J. A., and Rudebusch, G. D. (2014). Can spanned term structure factors drive stochastic yield volatility? Federal Reserve Bank of San Francisco.

[14] Christiansen, C. (2005). Multivariate term structure models with level and heteroskedasticity effects. Journal of Banking and Finance, 29(5), 1037-1057.

[15] Christoffersen, P., Dorion, C., Jacobs, K., and Karoui, L. (2014). Nonlinear Kalman filtering in affine term structure models. Management Science, 60(9), 2248-2268.

[16] Cieslak, A., and Povala, P. (2016). Information in the term structure of yield curve volatility. The Journal of Finance, 71(3), 1393-1436.

[17] Collin-Dufresne, P., and Goldstein, R. S. (2002). Do bonds span the fixed income markets? Theory and evidence for unspanned stochastic volatility. The Journal of Finance, 57(4), 1685-1730.

[18] Collin-Dufresne, P., Goldstein, R. S., and Jones, C. S. (2009). Can interest rate volatility be extracted from the cross section of bond yields? Journal of Financial Economics, 94(1), 47-66.

[19] Cox, J. C., Ingersoll Jr, J. E. and Ross, S. A. 1985. A Theory of the Term Structure of Interest Rates. Econometrica: Journal of the Econometric Society, 385-407.

[20] Creal, D. D., and Wu, J. C. (2015). Estimation of affine term structure models with spanned or unspanned stochastic volatility. Journal of Econometrics, 185(1), 60-81.

[21] Dai, Q., and Singleton, K. J. (2000). Specification analysis of affine term structure models. The Journal of Finance, 55(5), 1943-1978.

[22] Dai, Q., and Singleton, K. J. (2002). Expectation puzzles, time-varying risk premia, and affine models of the term structure. Journal of Financial Economics, 63(3), 415-441.

[23] Dai, Q., and Singleton, K. J.(2003). Term structure dynamics in theory and reality. Review of Financial Studies, 16(3), 631-678.

[24] Dai, Q., Singleton, K. J., and Yang, W. (2007). Regime shifts in a dynamic term structure model of US treasury bond yields. The Review of Financial Studies, 20(5), 1669-1706.

[25] Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. The Journal of Finance, 57(1), 405-443.

[26] Duffee, G. R., and Stanton, R. H. (2012). Estimation of dynamic term structure models. The Quarterly Journal of Finance, 2(2).

[27] Duffie, D. and Kan, R. 1996. A Yield Factor Model of Interest Rates. Mathematical finance, 6, 379-406.

[28] Duarte, J. (2004). Evaluating an alternative risk preference in affine term structure models. Review of Financial Studies, 17(2), 379-404.

[29] Engle, R. (1982). ARCH with estimates of variance of United Kingdom inflation. Econometrica, 50(4), 987-1007.

[30] Fan, R., Gupta, A., and Ritchken, P. (2003). Hedging in the possible presence of unspanned stochastic volatility: Evidence from swaption markets. The Journal of Finance, 58(5), 2219-2248.

[31] Feldhütter, P., Heyerdahl-Larsen, C., and Illeditsch, P. (2018). Risk premia and volatilities in a nonlinear term structure model. Review of Finance, 22(1), 337-380.

[32] Fleming, M. J., Mizrach, B., and Nguyen, G. (2018). The microstructure of a US Treasury ECN: The BrokerTec platform. Journal of Financial Markets, 40, 2-22.

[33] Ghysels, E., Le, A., Park, S., and Zhu, H. (2014). Risk and return trade-off in the U.S. treasury market. Working Paper, University of North Carolina.

[34] Gürkaynak, R. S., Sack, B., and Wright, J. H. (2007). The US Treasury yield curve: 1961 to the present. Journal of monetary Economics, 54(8), 2291-2304.

[35] Haubrich, J., Pennacchi, G., and Ritchken, P. (2012). Inflation expectations, real rates, and risk premia: Evidence from inflation swaps. The Review of Financial Studies, 25(5), 1588-1629.

[36] Heidari, M., and Wu, L. (2003). Are interest rate derivatives spanned by the term structure of interest rates?. The Journal of Fixed Income, 13(1), 75-86.

[37] Heston, S. L., and Nandi, S. (2003). A two-factor term structure model under garch volatility. The Journal of Fixed Income, 13(1), 87-95.

[38] Jacobs, K., and Karoui, L. (2009). Conditional volatility in affine term-structure models: Evidence from Treasury and swap markets. Journal of Financial Economics, 91(3), 288-318.

[39] Jagannathan, R., Kaplin, A., and Sun, S. (2003). An evaluation of multi-factor CIR models using LIBOR, swap rates, and cap and swaption prices. Journal of Econometrics, 116(1-2), 113-146.

[40] Joslin, S. (2018). Can unspanned stochastic volatility models explain the cross section of bond volatilities?. Management Science, 64(4), 1707-1726.

[41] Joslin, S., and Le, A. (2013). Interest Rate Volatility and No-Arbitrage Affine Term Structure Models. Working Paper, University of Southern California.

[42] Joslin, S., Singleton, K. J., and Zhu, H. (2011). A new perspective on Gaussian dynamic term structure models. The Review of Financial Studies, 24(3), 926-970.

[43] Knez, P. J., Litterman, R., and Scheinkman, J. (1994). Explorations into factors explaining money market returns. The Journal of Finance, 49(5), 1861-1882.

[44] Koedijk, K. G., Nissen, F. G., Schotman, P. C., and Wolff, C. C. (1997). The dynamics of short-term interest rate volatility reconsidered. European Finance Review, 1(1), 105-130.

[45] Le, A., Singleton, K. J., & Dai, Q. (2010). Discrete-Time Affine$\mathbb{Q}$ Term Structure Models with Generalized Market Prices of Risk. The Review of Financial Studies, 23(5), 2184-2227.

[46] Leippold, M., and Wu, L. (2002). Asset pricing under the quadratic class. Journal of Financial and Quantitative Analysis, 37(2), 271-295.

[47] Li, H., and Zhao, F. (2006). Unspanned stochastic volatility: Evidence from hedging interest rate derivatives. The Journal of Finance, 61(1), 341-378.

[48] Litterman, R. B., and Scheinkman, J. (1991). Common factors affecting bond returns. The Journal of Fixed Income, 1(1), 54-61.

[49] Longstaff, F. A., and Schwartz, E. S. (1992). Interest rate volatility and the term structure: A two-factor general equilibrium model. The Journal of Finance, 47(4), 1259-1282.

[50] Mizrach, B., and Neely, C. J. (2006). The transition to electronic trading in the secondary treasury market (No. 2006-03). Federal Reserve Bank of St. Louis Review.

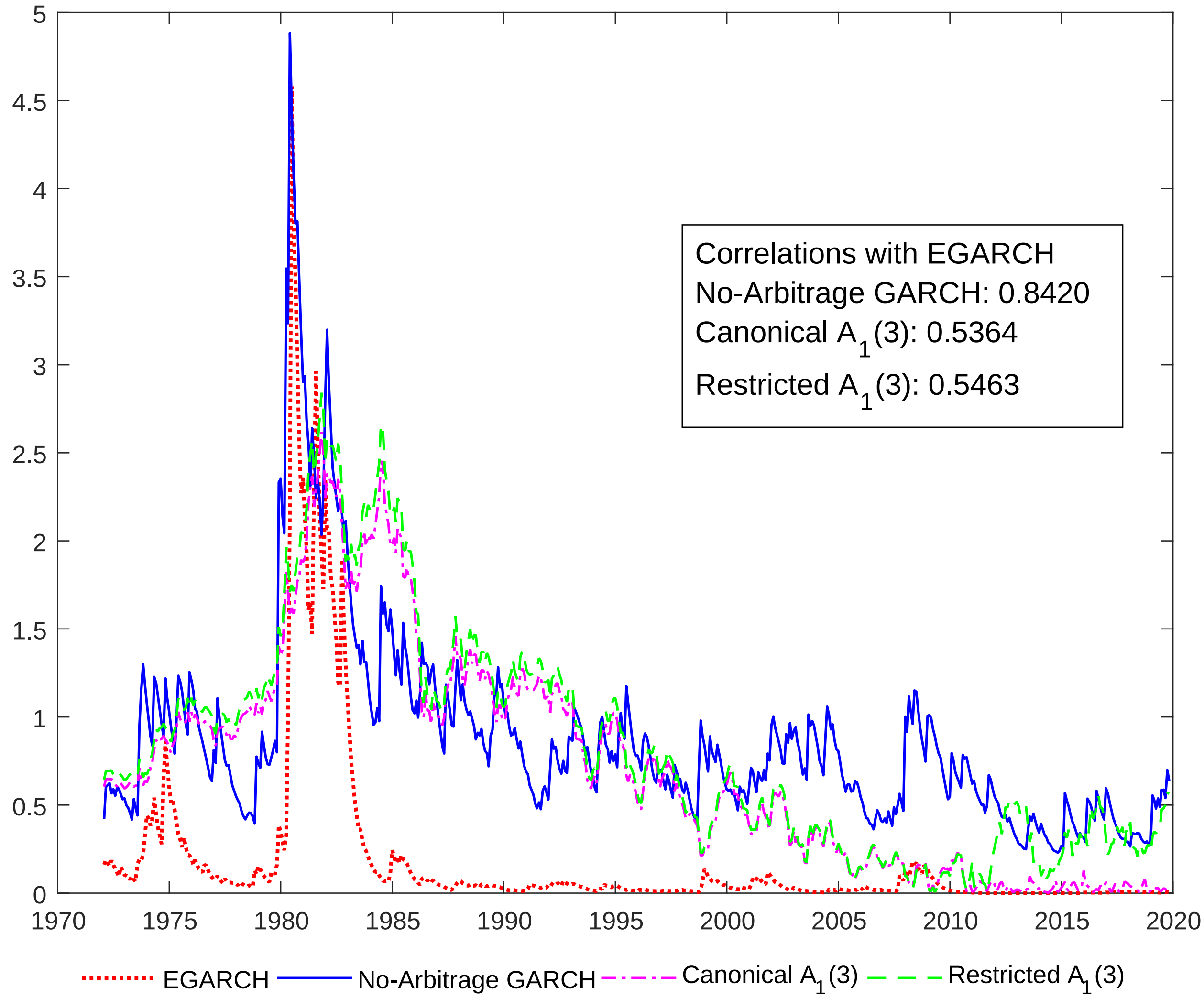
[51] Phoa, W. (1997). Can you derive market volatility forecasts from the observed yield curve convexity bias?. The Journal of Fixed Income, 7(1), 43-54.

[52] Schwert, G. W. (1989). Why does stock market volatility change over time? The journal of finance, 44(5), 1115-1153.

[53] Singleton, Kenneth J. (2006). Empirical dynamic asset pricing. Princeton University Press.

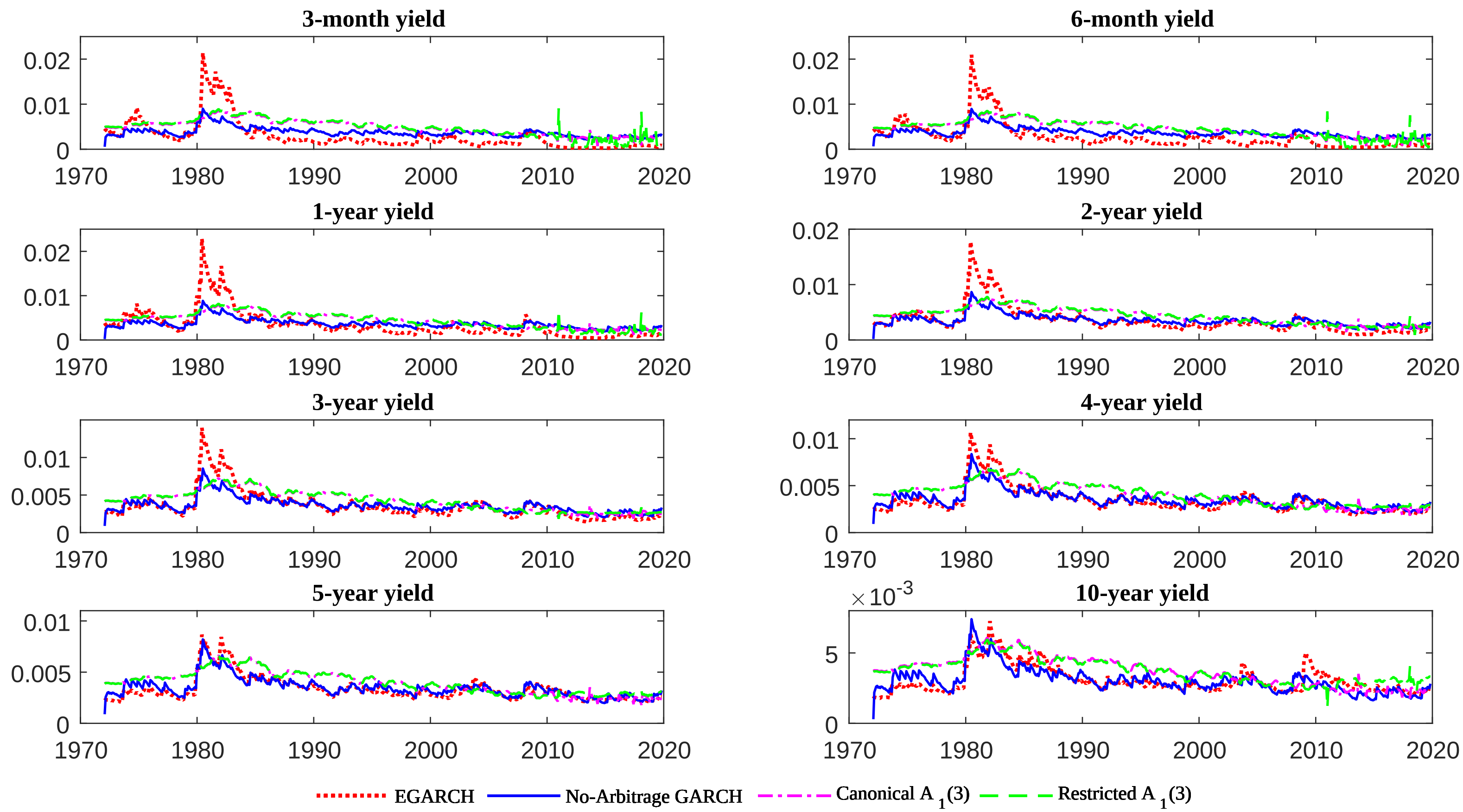
[54] Tang, H., and Xia, Y. (2007). An international examination of affine term structure models and the expectations hypothesis. Journal of Financial and Quantitative Analysis, 42(1), 41-80.

Figure 1: Model-Implied Variance Factors.



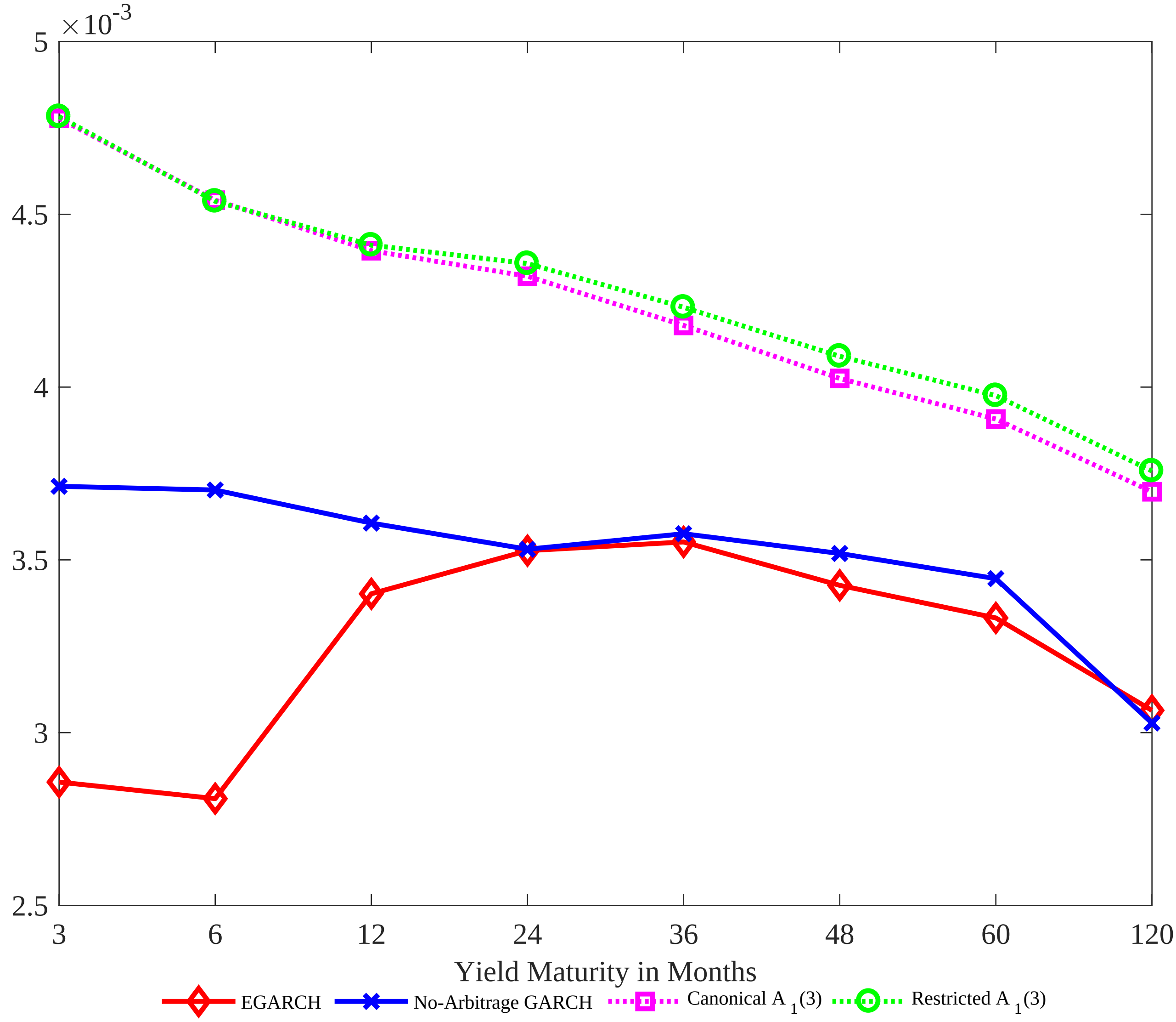
Notes to Figure: We plot the conditional variance factor $\sigma^2_{1,t}$ implied by different models. The dotted line (red) represents EGARCH$(1,1)$ variance estimated from changes in 3-month yields. The estimate is scaled by $1e^4$. The EGARCH$(1,1)$ is estimated assuming that the conditional mean of changes in monthly yields is generated by an AR(1) process. The solid line (blue) represents the variance factor from the no-arbitrage GARCH model. The dash-dot line (magenta) represents the variance factor from the canonical $A_1(3)$ model. The dashed line (green) represents the variance factor from the restricted $A_1(3)$ model. In the restricted $A_1(3)$ model, we set the feedback matrix to be a diagonal matrix. The numbers in the text box show the unconditional correlation between the EGARCH estimates of the 3-month yield and the variance factors from the three models.

Figure 2: Model-Implied Conditional Volatility by Maturity.



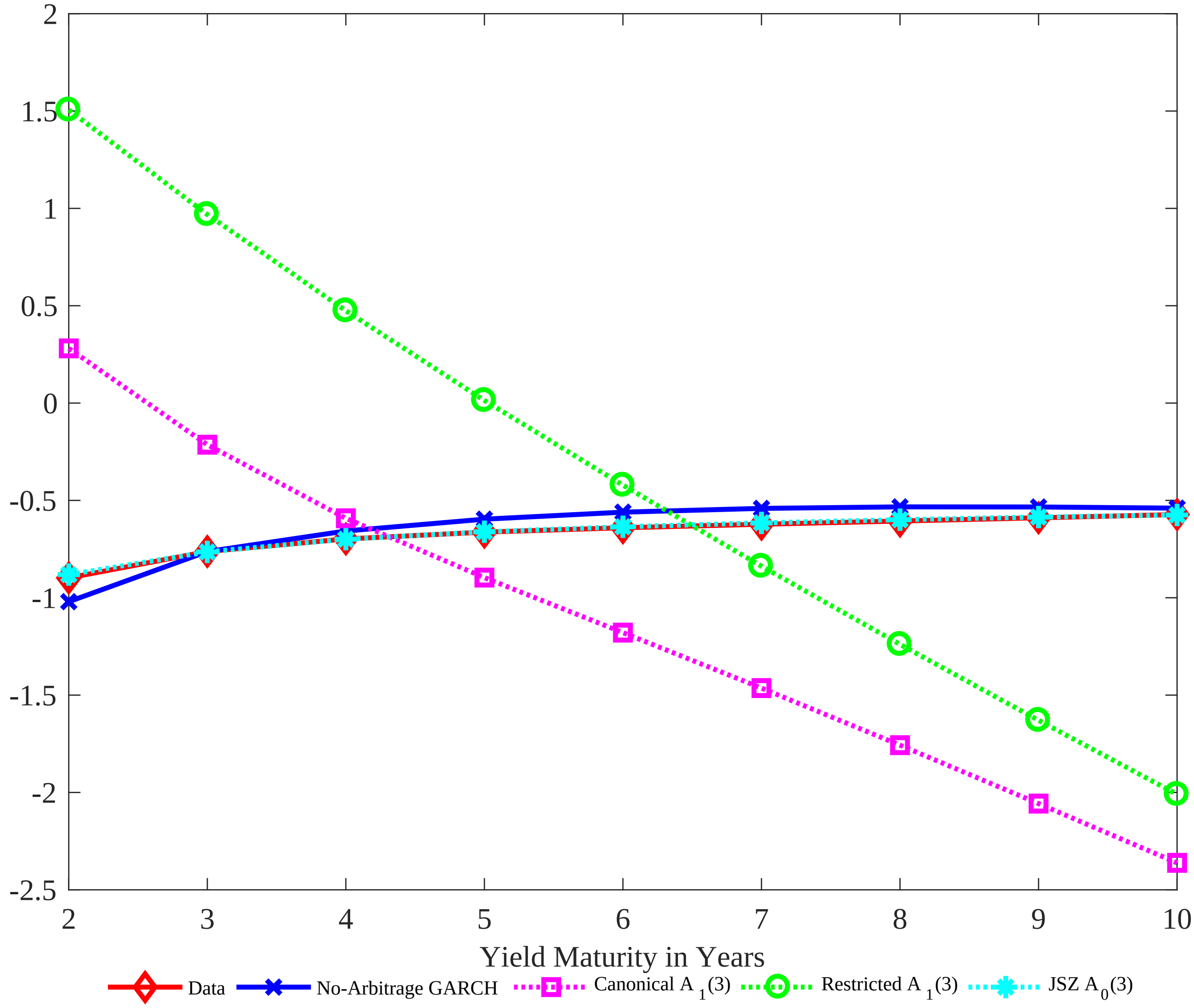
Notes to Figure: We plot the implied one-month conditional yield volatility for different models. For each maturity, the dotted line (red) represents EGARCH$(1,1)$ volatility estimated from changes in 3-month yields. The EGARCH$(1,1)$ is estimated assuming that the conditional mean of changes in monthly yields is generated by an AR(1) process. The solid line (blue) represents the conditional volatility from the no-arbitrage GARCH model. The dash-dot line (magenta) represents the conditional volatility from the canonical $A_1(3)$ model. The dashed line (green) represents the conditional volatility from the restricted $A_1(3)$ model. In the restricted $A_1(3)$ model, we set the feedback matrix to be a diagonal matrix.

Figure 3: The Term Structure of Unconditional Volatility.



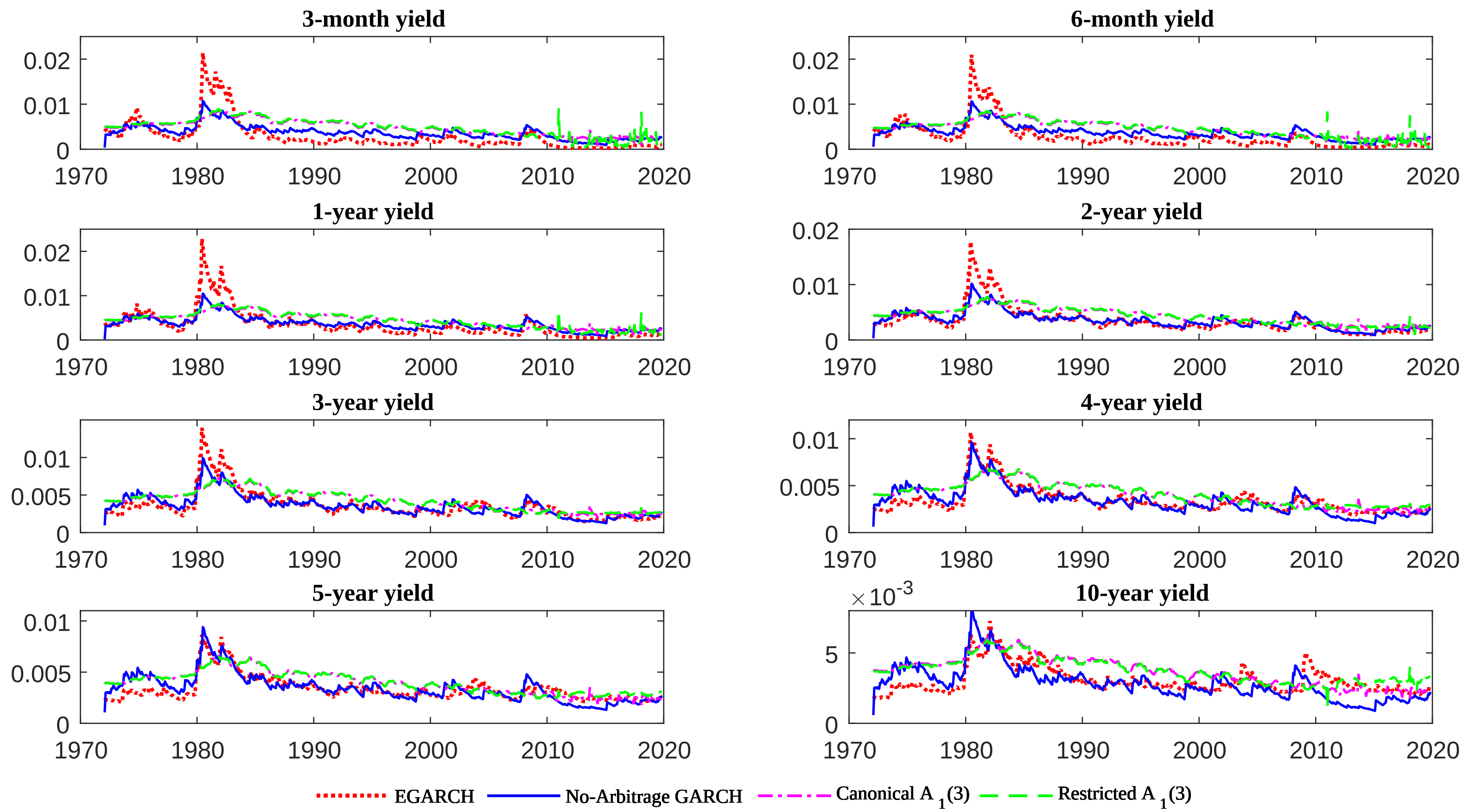
Notes to Figure: We plot the term structure of average yield volatility implied by different models. The diamond line (red) represents the EGARCH(1, 1) term structure. The EGARCH(1, 1) model is estimated assuming that the conditional mean of changes in monthly yields is generated by an AR(1) process. The cross line (blue) represents the term structure of volatility from the no-arbitrage GARCH model. The square line (magenta) represents the term structure of volatility from the canonical $A_1(3)$ model. The circle line (green) represents the term structure of volatility from the restricted $A_1(3)$ model. In the restricted $A_1(3)$ model, we set the feedback matrix to be a diagonal matrix. The x-axis is yield maturity in months.

Figure 4: The Campbell-Shiller Expectations Hypothesis Regression



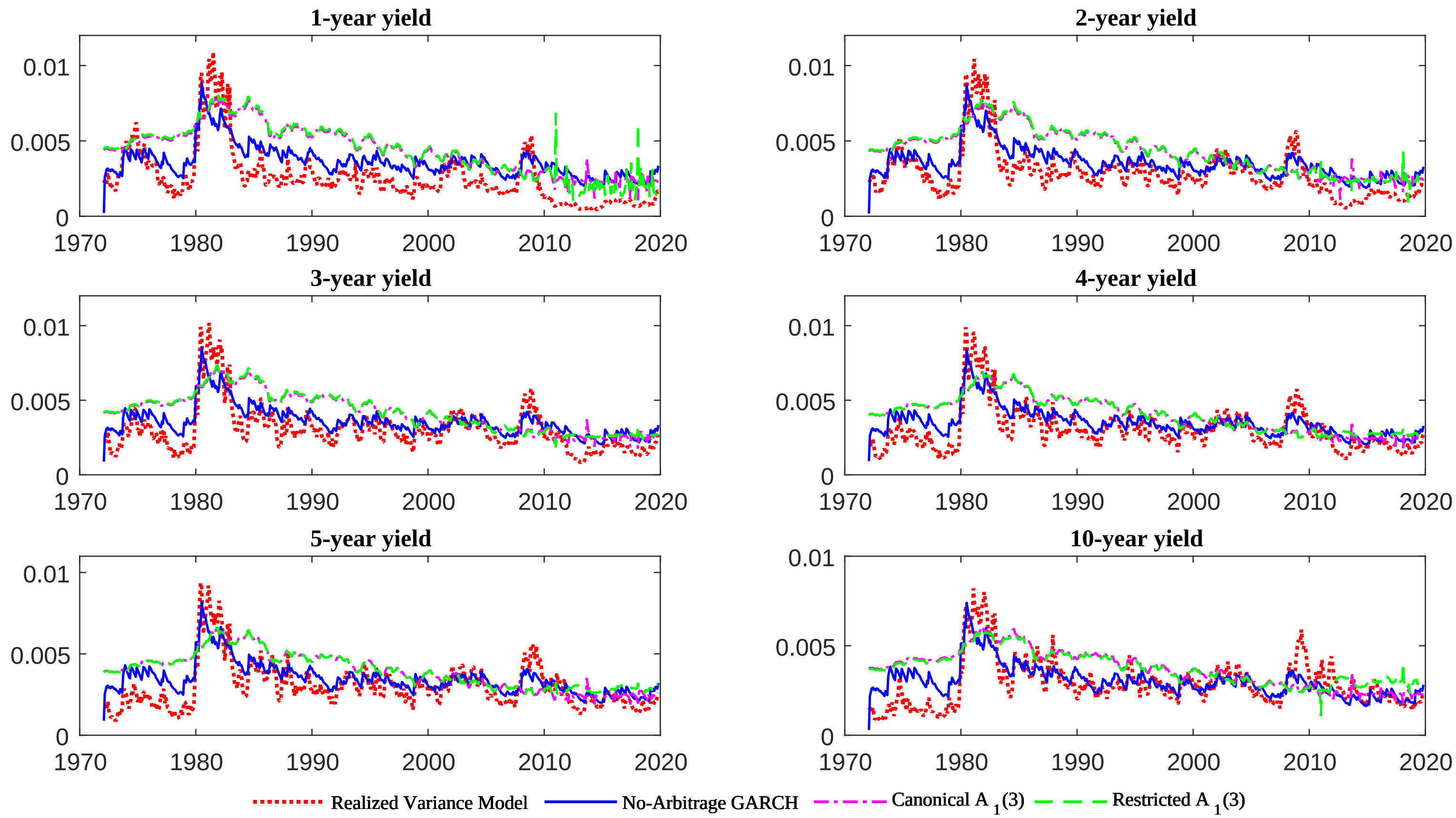
Notes to Figure: We plot the estimated coefficients from the Campbell-Shiller expectation hypothesis regression using yields implied by different models. The diamond line (red) represents the estimates using the yield data. The cross line (blue) represents the estimates from yields implied by the no-arbitrage GARCH model. The square line (magenta) shows the estimates from yields implied by the canonical $A_1(3)$ model. The circle line (green) is for the estimates from yields implied by the restricted $A_1(3)$ model. In the restricted $A_1(3)$ model, we set the feedback matrix to be a diagonal matrix. The asterisk line (cyan) is for the estimates using yields from the Joslin, Singleton, and Zhu (2011) specification of the Gaussian $A_0(3)$ model. The x-axis is yield maturity in years.

Figure 5: Model-Implied Conditional Volatility. Estimation Using Yield Volatility Only.



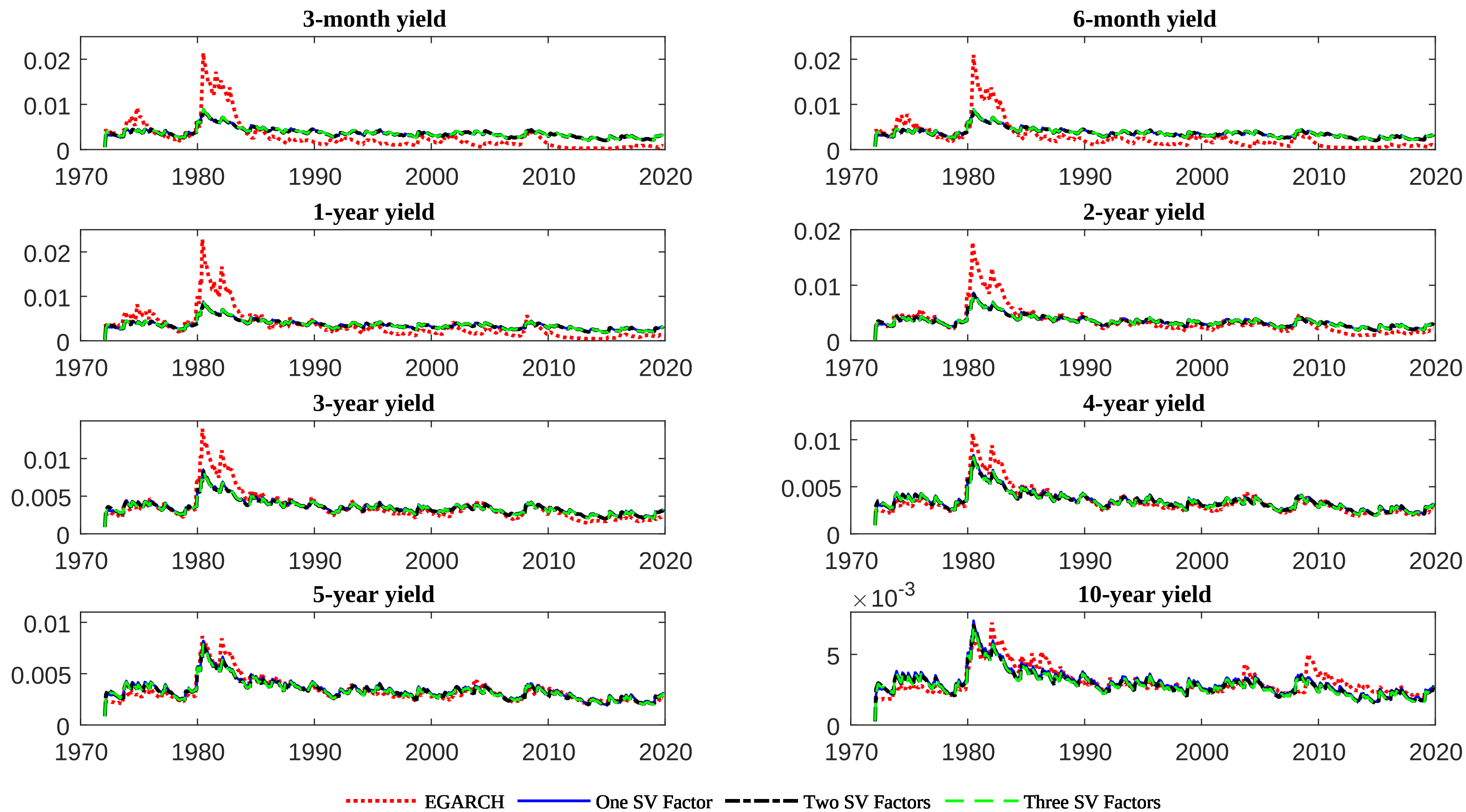
Notes to Figure: We plot the one-month conditional volatility of the yield curve implied by different models estimated by minimizing the sum squared errors of yield volatilities. For each maturity, the dotted line (red) represents the EGARCH$(1,1)$ estimated volatility of changes in yields. The EGARCH$(1,1)$ is estimated assuming that the conditional mean of changes in monthly yields is generated by an AR(1) process. The solid line (blue) represents the volatility from the no-arbitrage GARCH model. The dash-dot line (magenta) represents the volatility from the canonical $A_1(3)$ model. The dashed line (green) represents the volatility from the restricted $A_1(3)$ model. In the restricted $A_1(3)$ model, we set the feedback matrix to be a diagonal matrix.

Figure 6: Model-Implied Conditional Volatility and Realized Volatility.



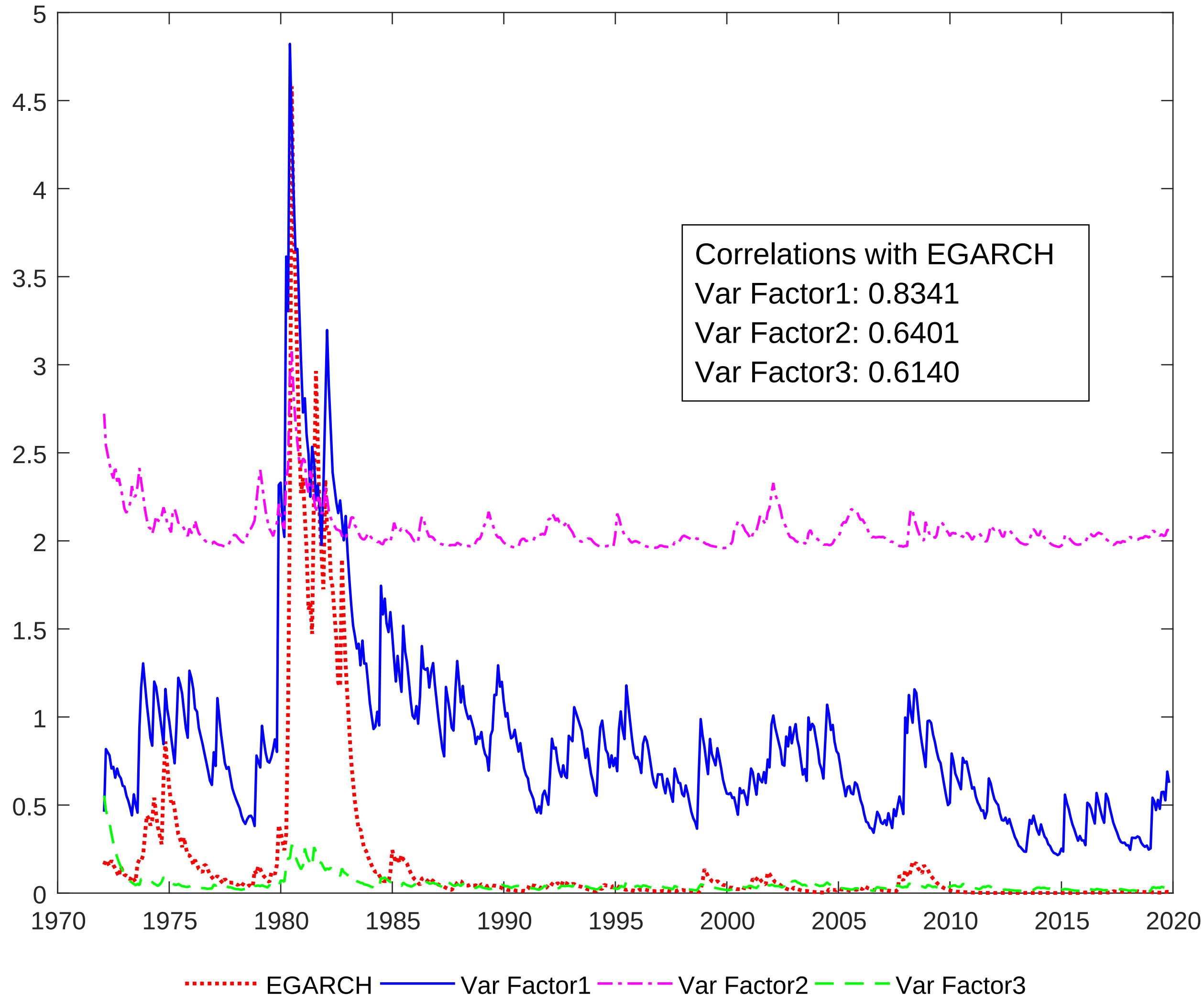
Notes to Figure: We plot the one-month conditional volatility of the yield curve implied by different models. For each maturity, the dotted line (red) represents the conditional volatility implied by the realized variance model. We construct the measure of monthly variance using within-month squared changes in yields. The logarithm of the realized variance follows an ARMA$(1,1)$ process. The solid line (blue) represents the conditional volatility from the no-arbitrage GARCH model. The dash-dot line (magenta) represents the conditional volatility from the canonical $A_1(3)$ model. The dashed line (green) represents the conditional volatility from the restricted $A_1(3)$ model. In the restricted $A_1(3)$ model, we set the feedback matrix to be a diagonal matrix.

Figure 7: Model-Implied Conditional Volatility in the No-Arbitrage GARCH Model with Multiple Volatility Factors.



Notes to Figure: We plot the one-month conditional volatility of the yield curve from the no-arbitrage GARCH model with different numbers of volatility factors. For each maturity, the dotted line (red) represents the EGARCH(1, 1) estimated volatility of changes in yields. The EGARCH(1, 1) is estimated assuming that the conditional mean of changes in monthly yields is generated by an AR(1) process. The solid line (blue) represents the volatility from the no-arbitrage GARCH model with a single volatility factor. The dash-dot line (black) represents the volatility from the model with two volatility factors. The dashed line (green) represents the volatility from the model with three volatility factors.

Figure 8: Variance Factors in the No-Arbitrage GARCH Model with Three Variance Factors.



Notes to Figure: We plot the variance factors $\sigma^2_{i,t}$ from the no-arbitrage GARCH model with three variance factors. For each maturity, the dotted line (red) represents the EGARCH(1, 1) estimated variance of changes in 3-month yield. The estimate is scaled by $1e^4$. The EGARCH(1, 1) is estimated assuming that the conditional mean of changes in monthly yields is generated by an AR(1) process. The solid line (blue) represents the first variance factor. The dash-dot line (magenta) represents the second variance factor, which is scaled by $1e^7$. The dashed line (green) represents the third variance factor, which is scaled by $1e^4$. The numbers in the text box show the unconditional correlation between the EGARCH estimates of the 3-month yield and the model-implied variance factors.

Table 1: Summary Statistics

Panel A: Yields

	Central Moments				Autocorrelation		
	Mean (%)	St.Dev (%)	Skewness	Kurtosis	Lag 1	Lag 12	Lag 30
3 month	4.6508	3.4534	0.6171	3.3040	0.9904	0.8588	0.6391
6 month	4.7758	3.4409	0.5332	3.0506	0.9915	0.8707	0.6632
1 year	5.1349	3.5744	0.4511	2.8107	0.9906	0.8833	0.7019
2 year	5.3581	3.5079	0.3793	2.6595	0.9915	0.8981	0.7500
3 year	5.5477	3.4214	0.3642	2.6022	0.9921	0.9057	0.7775
4 year	5.7130	3.3334	0.3695	2.5853	0.9923	0.9095	0.7938
5 year	5.8587	3.2510	0.3816	2.5896	0.9923	0.9111	0.8032
10 year	6.3640	2.9658	0.4264	2.6891	0.9918	0.9088	0.8110

Panel B: EGARCH Volatilities

	Central Moments				Autocorrelation		
	Mean (bps)	St.Dev (bps)	Skewness	Kurtosis	Lag 1	Lag 12	Lag 30
3 month	28.5701	32.9859	2.7703	11.2903	0.9830	0.7292	0.2958
6 month	28.0909	29.3398	2.7781	12.1228	0.9801	0.6905	0.2998
1 year	34.0188	31.2364	2.6563	11.9225	0.9717	0.6947	0.3562
2 year	35.2671	24.2314	2.6036	11.3548	0.9775	0.6972	0.3708
3 year	35.5186	18.9077	2.4514	10.1627	0.9776	0.6936	0.3635
4 year	34.2632	14.2330	2.2827	8.9541	0.9762	0.6899	0.3539
5 year	33.3196	11.7587	2.0619	7.6196	0.9761	0.7025	0.3603
10 year	30.6445	9.7072	1.5141	4.9637	0.9706	0.6867	0.3639

Notes to Table: We present summary statistics for the data used in estimation. We present the sample mean, standard deviation, skewness, kurtosis, and autocorrelations for each of the yields (Panel A) and yield volatilities (Panel B). The yields are continuously compounded monthly zero-coupon bond yields. The yields volatilities are the EGARCH$(1,1)$ estimates of changes in yields. The EGARCH$(1,1)$ is estimated assuming that the conditional mean of changes in monthly yields is generated by an AR(1) process. The sample period is from 1971:11 to 2019:10.

Table 2: Parameter Estimates

Panel A: The No-Arbitrage GARCH Model								
K_0^P		K_1^P			K_0^Q		K_1^Q	
0.0025	0.9979				0.0017	0.9966		
-0.0039		0.9542			-0.0072		0.9466	
0.0007			0.9449		0.0161			0.7212
ρ_0			ρ_1				Log Likelihood	
-0.0001		0.0003	-0.0338	0.0372			19173.32	
$\omega \times 1e2$	α	β		$\sigma_2 \times 1e4$	$\sigma_3 \times 1e4$			
0.0001	0.0794	0.9064		0.0021	0.0153			
Panel B: The Canonical $A_1(3)$ Model								
K_0^P		K_1^P			K_0^Q		K_1^Q	
0.0222	0.9958				0.0222	1.0011		
-0.4750	0.0892	0.9697	-0.2707		0.0298	0.1246	0.9658	-0.7689
-0.0850	0.0160	0.0016	0.8095		-0.1022	0.0140	0.0021	0.8867
ρ_0			ρ_1				Log Likelihood	
0.0172		0.0015	0.0015	0.0056			15640.90	
	a				b			
0.0000	1.0000	1.0000		1.0000	10.3716	0.1575		
Panel C: The Restricted $A_1(3)$ Model								
K_0^P		K_1^P			K_0^Q		K_1^Q	
0.0226	0.9958				0.0226	1.0010		
0.0000		0.9697			0.0000		0.9657	
0.0000			0.8118		0.0000			0.8890
ρ_0			ρ_1				Log Likelihood	
0.0172		0.0015	0.0015	0.0057			15454.31	
	a				b			
0.0000	1.0000	1.0000		1.0000	10.3615	0.1578		

Notes to Table: We present the estimated parameters and log likelihoods for the no-arbitrage GARCH model (Panel A), the canonical $A_1(3)$ model (Panel B), and the restricted $A_1(3)$ model (Panel C). In the restricted $A_1(3)$ model, we set the feedback matrix to be a diagonal matrix. The estimates are based on the monthly zero-coupon bond yields from 1971:11 to 2019:10. The state variables in the canonical and restricted $A_1(3)$ models follow a first order VAR process when sampled monthly.

Table 3: Unconditional Correlations between the EGARCH(1,1) Estimated Volatility and Conditional Volatility Implied by Different Models

	3 month	6 month	1 year	2 year	3 year	4 year	5 year	10 year	Avg.
No-Arbitrage GARCH	0.85	0.86	0.91	0.94	0.95	0.94	0.92	0.80	0.90
Canonical $A_1(3)$	0.62	0.64	0.69	0.71	0.71	0.70	0.69	0.58	0.67
Restricted $A_1(3)$	0.59	0.61	0.67	0.71	0.71	0.70	0.69	0.59	0.66

Notes to Table: We present the unconditional correlations between the EGARCH$(1,1)$ estimated volatility and the one-month conditional volatility implied by the no-arbitrage GARCH model, the canonical $A_1(3)$ model, and the restricted $A_1(3)$ model. The EGARCH$(1,1)$ is estimated assuming that the conditional mean of changes in monthly yields is generated by an AR(1) process. In the restricted $A_1(3)$ model, we set the feedback matrix to be a diagonal matrix. The last column in the table shows the average across all maturities.

Table 4: Conditional Volatility Fit

	3 month	6 month	1 year	2 year	3 year	4 year	5 year	10 year	Avg.
Panel A: RMSEs of Volatility									
No-Arbitrage GARCH	25.96	22.65	22.09	14.73	9.86	6.10	4.86	5.90	14.02
Canonical $A_1(3)$	32.43	28.60	25.40	18.98	14.68	12.01	10.87	11.09	19.26
Restricted $A_1(3)$	32.89	29.09	25.46	19.07	14.90	12.24	10.98	10.88	19.44
Improve on Canonical $A_1(3)$	19.97%	20.79%	13.04%	22.36%	32.82%	49.21%	55.25%	46.78%	32.53%
Improve on Restricted $A_1(3)$	21.09%	22.14%	13.23%	22.73%	33.84%	50.16%	55.68%	45.75%	33.08%
Panel B: MAPEs of Volatility									
No-Arbitrage GARCH	182.93	139.11	81.36	30.32	17.72	12.82	11.19	15.33	61.35
Canonical $A_1(3)$	224.68	164.69	99.17	50.97	35.18	30.06	28.42	33.14	83.29
Restricted $A_1(3)$	196.67	151.43	92.40	51.17	37.12	32.28	30.31	34.06	78.18
Improve on Canonical $A_1(3)$	18.58%	15.53%	17.96%	40.52%	49.64%	57.36%	60.61%	53.76%	39.25%
Improve on Restricted $A_1(3)$	6.99%	8.14%	11.95%	40.75%	52.27%	60.28%	63.08%	55.00%	37.31%

Notes to Table: We present the volatility fit for the no-arbitrage GARCH model, the canonical $A_1(3)$ model, and the restricted $A_1(3)$ model. In the restricted $A_1(3)$ model, we set the feedback matrix to be a diagonal matrix. Panel A reports the RMSEs in basis points and Panel B reports the MAPEs in percentages. The RMSEs and MAPEs are computed by using an EGARCH(1, 1) model to measure the "true" conditional volatility. The EGARCH$(1, 1)$ is estimated assuming that the conditional mean of changes in monthly yields is generated by an AR(1) process. We also present the percentage RMSE and MAPE improvement of the no-arbitrage GARCH model over the canonical and restricted $A_1(3)$ models. The last column in the table shows the average across all maturities.

Table 5: Yield Fit

Panel A: Stochastic Volatility Models	3 month	6 month	1 year	2 year	3 year	4 year	5 year	10 year	Avg.
No-Arbitrage GARCH	24.22	25.06	4.27	10.44	16.33	18.74	20.10	32.69	18.98
Canonical $A_1(3)$	12.84	17.59	27.12	16.25	9.11	13.36	19.70	36.97	19.12
Restricted $A_1(3)$	18.84	15.09	24.27	15.47	11.53	15.45	21.05	37.92	19.96
Panel B: Gaussian Models									
JSZ	7.95	10.36	12.88	4.65	5.08	5.67	5.95	7.73	7.53
JSZ Restricted_1	25.04	23.13	5.72	5.36	1.31	3.88	5.02	7.71	9.65
JSZ Restricted_2	7.81	14.18	15.71	7.89	6.20	6.75	7.08	11.81	9.68
JSZ Restricted_3	9.33	11.65	14.75	6.62	5.21	6.37	7.42	8.08	8.68

Notes to Table: We present RMSEs based on the fit of yields for the time-varying volatility models (Panel A) and the Gaussian models (Panel B). We reports the results for the no-arbitrage GARCH model, the canonical $A_1(3)$ model, and the restricted $A_1(3)$ model in Panel A. In the restricted $A_1(3)$ model, we set the feedback matrix to be a diagonal matrix. In Panel B, we consider the canonical representation of Joslin, Singleton, and Zhu (2011, referred to as JSZ). We present the results for the maximum flexible specification of the JSZ model and also for three restricted JSZ canonical forms. In the first restricted version, we set the variance-covariance matrix to be a diagonal matrix. In the second restricted version, we also set the $(1, 2)$ and $(1, 3)$ entries of the feedback matrix to be zeros. In the third restricted version, we set the feedback matrix to be a diagonal matrix. The last column in the table shows the average across all maturities. RMSEs are reported in basis points.

Table 6: Estimation Based on Volatilities Only

	3 month	6 month	1 year	2 year	3 year	4 year	5 year	10 year	Avg.
Panel A: Unconditional Correlations									
No-Arbitrage GARCH	0.90	0.91	0.94	0.93	0.90	0.86	0.82	0.66	0.86
Canonical $A_1(3)$	0.62	0.64	0.69	0.71	0.71	0.70	0.69	0.58	0.67
Restricted $A_1(3)$	0.59	0.61	0.67	0.71	0.71	0.70	0.69	0.59	0.66
Panel B: RMSEs of Volatility									
No-Arbitrage GARCH	21.56	18.53	17.47	11.60	8.64	7.53	7.63	9.41	12.80
Canonical $A_1(3)$	32.43	28.60	25.40	18.98	14.68	12.01	10.87	11.09	19.26
Restricted $A_1(3)$	32.89	29.09	25.45	19.07	14.90	12.24	10.98	10.88	19.44
Improve on Canonical $A_1(3)$	33.53%	35.21%	31.25%	38.89%	41.12%	37.33%	29.76%	15.14%	32.78%
Improve on Restricted $A_1(3)$	34.45%	36.30%	31.39%	39.17%	42.01%	38.50%	30.44%	13.49%	33.22%
Panel C: MAPEs of Volatility									
No-Arbitrage GARCH	127.44	100.29	49.37	17.66	16.23	18.31	19.11	26.19	46.83
Canonical $A_1(3)$	224.68	164.69	99.17	50.97	35.18	30.06	28.42	33.14	83.29
Restricted $A_1(3)$	196.67	151.43	92.40	51.17	37.12	32.28	30.31	34.06	78.18
Improve on Canonical $A_1(3)$	43.28%	39.10%	50.22%	65.36%	53.86%	39.08%	32.74%	20.98%	43.08%
Improve on Restricted $A_1(3)$	35.20%	33.77%	46.57%	65.50%	56.27%	43.26%	36.95%	23.10%	42.58%

Notes to Table: We present the unconditional correlations and the volatility fit for models estimated by minimizing the sum squared errors of yield volatilities. Panel A presents the unconditional correlations between the EGARCH$(1,1)$ estimates and the one-month conditional volatility implied from the no-arbitrage GARCH model, the canonical $A_1(3)$ model, and the restricted $A_1(3)$ model. The EGARCH$(1,1)$ is estimated assuming that the conditional mean of changes in monthly yields is generated by an AR(1) process. In the restricted $A_1(3)$ model, we set the feedback matrix to be a diagonal matrix. Panel B reports the RMSEs of volatility in basis points for the three models, and Panel C reports the MAPEs of volatility in percentages for the three models. The RMSEs and MAPEs are computed by using the EGARCH$(1,1)$ model to measure the "true" conditional volatility. We also present the percentage RMSE and MAPE improvement for the no-arbitrage GARCH model over the canonical and restricted $A_1(3)$ models. The last column in the table shows the average across all maturities.

Table 7: Results Based on Realized Volatility

Panel A: Unconditional Correlations

	1 year	2 year	3 year	4 year	5 year	10 year	Avg.
EGARCH$(1,1)$	0.91	0.88	0.87	0.87	0.86	0.85	0.87
No-Arbitrage GARCH	0.86	0.87	0.87	0.85	0.83	0.74	0.84
Canonical $A_1(3)$	0.64	0.59	0.54	0.50	0.47	0.40	0.52
Restricted $A_1(3)$	0.63	0.59	0.53	0.49	0.45	0.40	0.51

Panel B: RMSEs of Volatility

	1 year	2 year	3 year	4 year	5 year	10 year	Avg.
No-Arbitrage GARCH	14.73	10.97	10.05	9.51	9.30	8.96	10.59
Canonical $A_1(3)$	23.17	20.01	17.98	16.79	16.15	15.40	18.25
Restricted $A_1(3)$	23.79	20.45	18.40	17.14	16.40	15.33	18.59
Improve on Canonical $A_1(3)$	36.44%	45.18%	44.11%	43.36%	42.39%	41.78%	42.21%
Improve on Restricted $A_1(3)$	38.10%	46.37%	45.37%	44.51%	43.29%	41.51%	43.19%

Panel C: MAPEs of Volatility

	1 year	2 year	3 year	4 year	5 year	10 year	Avg.
No-Arbitrage GARCH	88.46	47.10	37.02	33.39	32.73	31.75	45.08
Canonical $A_1(3)$	118.10	79.53	63.57	58.22	57.18	61.27	72.98
Restricted $A_1(3)$	113.20	80.16	66.16	61.37	60.30	63.26	74.08
Improve on Canonical $A_1(3)$	25.09%	40.77%	41.77%	42.65%	42.76%	48.17%	40.20%
Improve on Restricted $A_1(3)$	21.85%	41.24%	44.05%	45.60%	45.73%	49.80%	41.38%

Notes to Table: We present the unconditional correlations and the volatility fit for various models when realized volatility instead of EGARCH volatility is used as a measure of model-free yield volatility. We construct the measure of monthly variance using within-month squared changes in yields. The logarithm of the realized variance follows an ARMA$(1,1)$ process. Panel A presents the unconditional correlations between the one-month conditional volatility from the realized variance model and the one-month conditional volatility implied from the EGARCH$(1,1)$ model, the no-arbitrage GARCH model, the canonical $A_1(3)$ model, and the restricted $A_1(3)$ model. The EGARCH$(1,1)$ is estimated assuming that the conditional mean of changes in monthly yields is generated by an AR(1) process. In the restricted $A_1(3)$ model, we set the feedback matrix to be a diagonal matrix. Panel B reports the RMSEs of volatility in basis points for the three stochastic volatility models, and Panel C reports the MAPEs of volatility in percentages for the three stochastic volatility models. We also present the percentage improvement in RMSE and MAPE for the no-arbitrage GARCH model over the canonical and restricted $A_1(3)$ models. The last column in the table shows the average across all maturities.

Table 8: Parameter Estimates for the No-Arbitrage GARCH Model with Multiple Volatility Factors

Panel A: Two Volatility Factors

K_0^P	K_1^P			K_0^Q	K_1^Q		
0.0022	0.9981			0.0011	0.9960		
-0.0048		0.9528		-0.0066		0.9551	
0.0004			0.9428	0.0158			0.7357

ρ_0	ρ_1			Log Likelihood
-0.0001	0.0003	-0.0340	0.0377	19175.59

$\omega \times 1e2$	α	β	$\sigma_3 \times 1e4$
0.0010	0.0720	0.9056	0.0036
0.0000	0.0000	0.8692	

Panel B: Three Volatility Factors

K_0^P	K_1^P			K_0^Q	K_1^Q		
0.0022	0.9985			0.0008	0.9959		
-0.0046		0.9584		-0.0071		0.9544	
0.0007			0.9376	0.0153			0.7313

ρ_0	ρ_1			Log Likelihood
-0.0001	0.0003	-0.0344	0.0349	19176.62

$\omega \times 1e2$	α	β
0.0008	0.0711	0.9021
0.0001	0.0000	0.7314
0.0001	0.0000	0.8431

Panel C: Likelihood Ratio Test

	Three VS One Volatility Factor	Two VS One Volatility Factor
P Value	0.1585	0.1041
Statistics	6.6013	4.5243

Notes to Table: We present the estimated parameters and log likelihoods for the no-arbitrage GARCH model with two volatility factors (Panel A) and the no-arbitrage GARCH model with three volatility factors (Panel B). The estimates are based on monthly zero coupon bond yields from 1971:11 to 2019:10. Panel C presents the p-value and test statistics of the likelihood ratio tests between the no-arbitrage GARCH model with three and one volatility factors, and between the no-arbitrage GARCH model with two and one volatility factors.