Journal of Financial Economics 94 (2009) 47–66

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Can interest rate volatility be extracted from the cross section of bond yields?☆

Pierre Collin-Dufresne [a], Robert S. Goldstein [b], Christopher S. Jones [c,*]

[a] Graduate School of Business, Columbia University, New York, NY 10027, USA and NBER
[b] Carlson School of Management, University of Minnesota, Minneapolis, MN 55455, USA and NBER
[c] Marshall School of Business, University of Southern California, Los Angeles, CA 90089, USA

ARTICLE INFO

Article history:
Received 16 May 2007
Received in revised form
31 January 2008
Accepted 30 June 2008
Available online 30 May 2009

JEL classification:
G12
G13

Keywords:
Term structure of interest rates
Affine models
Stochastic volatility

ABSTRACT

Most affine models of the term structure with stochastic volatility predict that the variance of the short rate should play a 'dual role' in that it should also equal a linear combination of yields. However, we find that estimation of a standard affine three-factor model results in a variance state variable that, while instrumental in explaining the shape of the yield curve, is essentially unrelated to GARCH estimates of the quadratic variation of the spot rate process or to implied variances from options. We then investigate four-factor affine models. Of the models tested, only the model that exhibits 'unspanned stochastic volatility' (USV) generates both realistic short rate volatility estimates and a good cross-sectional fit. Our findings suggest that short rate volatility cannot be extracted from the cross-section of bond prices. In particular, short rate volatility and convexity are only weakly correlated.

1. Introduction

This paper investigates the relation between interest rate volatility and the cross section of bond yields. It is well-established that at least three factors are needed to capture bond yield dynamics: Litterman and Scheinkman (1991) interpret them as level, slope, and curvature. It is also well-established that interest rate volatility is stochastic.[1] As such, this paper focuses on those three- and four-factor models of the term structure that capture stochastic volatility.

By imposing only the condition of no-arbitrage, it can be shown that state variables that drive changes in interest rate volatility generally play a 'dual role' in that they also drive changes in bond yields.[2] For example, theoretical considerations generally imply a strong link between changes in short-rate volatility and changes in curvature. Not surprisingly, this prediction is manifest in the specific models proposed in the literature. For example, affine models of the term structure predict that state variables driving interest rate volatility are also

☆ We thank seminar participants at UCLA, Cornell University, McGill, the University of Minnesota, the University of Arizona, UNC, Barclays Global Investors, Syracuse University, the University of Pennsylvania, the USC Applied Math seminar, the University of Texas at Austin, the CIREQ-CIRANO-MITACS conference on Univariate and Multivariate Models for Asset Pricing, the Econometric Society Meetings in Washington DC, and the Math-Finance Workshop in Frankfurt for suggestions and comments. We also thank an anonymous referee, Luca Benzoni, Michael Brandt, Qiang Dai, Jefferson Duarte, Garland Durham, Bing Han, Philipp Illeditsch, Mike Johannes, Ken Singleton, and especially Suresh Sundaresan and Mike Chernov for many helpful comments.

* Corresponding author.
E-mail address: christopher.jones@marshall.usc.edu (C.S. Jones).

0304-405X/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jfineco.2008.06.007

[1] See, for example, Fama (1976), Engle, Lilien, and Robins (1987), Brenner, Harjes, and Kroner (1996), and Han (2007).

[2] See, for example, Litterman, Scheinkman, and Weiss (1991), and Collin-Dufresne, Goldstein, and Jones (CGJ, 2008, p. 785), and Eq. (11) below.

linear combinations of yields. As a special case, the 'preferred' $A_1(3)$ model of Dai and Singleton (DS, 2000) predicts that the variance of the spot rate is also a linear combination of the level, slope, and curvature of the yield curve.

The first goal of this paper is to investigate whether the variance state variable in the preferred $A_1(3)$ model can simultaneously satisfy its dual roles as the source of time-varying yield volatility and a factor in the yield curve. We choose this model because Duffee (2002) concludes that it offers the best forecasting performance among three-factor models with stochastic volatility, while DS find that it offers the best characterization of unconditional yield volatilities and a sufficiently flexible correlation structure.[3]

On the other hand, there is evidence that the $A_1(3)$ model is misspecified.[4] Therefore, a more general goal of the paper is to investigate the joint dynamics of level, slope, curvature, and volatility factors within a more flexible four-factor affine framework that does not impose at the outset a deterministic link between them.

We also investigate a subset of the affine class that displays 'unspanned stochastic volatility' (USV). These models impose strong parameter constraints in order to generate bond yields that are independent of some of the variables driving volatility. As such, these 'unspanned' variables do not play a dual role and thus are free to more accurately capture the time-series properties of interest rate volatility. This potential benefit is not without costs, however. First, the large number of parameter constraints may prove to be overly restrictive. For example, the $A_1(4)$ USV model has only 11 risk-neutral parameters that affect bond yields, compared to 22 for its unrestricted counterpart. Second, limiting a variance state variable to have only a time-series role means that such a model will be less able to explain cross-sectional yield patterns. For example, the $A_1(4)$ USV model has only three factors to capture the cross section of yields.

Prior empirical work on USV has focused almost exclusively on the spanning relation between interest rate derivatives and bond prices. For example, Collin-Dufresne and Goldstein (CDG, 2002) find changes in swap rates are only weakly correlated with returns on at-the-money cap straddles, while Heidari and Wu (2003) obtain similar results using implied volatilities from swaptions. Li and Zhao (2006) investigate quadratic models of the term structure and find them incapable of explaining the returns on caps of various maturities and strike prices. In contrast, Fan, Gupta, and Ritchken (2003) report that swaption returns are in fact well-spanned by yield changes, while both Bikbov and Chernov (2009) and Joslin (2007) report that the USV restrictions are strongly rejected when bond and option data are used.

Our paper differs from the previous work by focusing on the ability of USV models to explain bond prices themselves. Specifically, we are interested in determining whether USV models are able to simultaneously match both the cross-sectional and time-series properties of bond yields. While it seems clear that the USV model will match some aspects of the time series of volatility (i.e., conditional second moments of yields) relatively well, it is unclear how the restrictions imposed to generate USV will affect the model's ability to capture the cross section and time series of yields (i.e., conditional first moments).

We note that many standard econometric methods used to estimate affine models are unsuitable for investigating models exhibiting USV. Indeed, a consequence of imposing USV restrictions is that the one-to-one mapping assumed by Duffie and Kan (DK, 1996) between yields and factors does not hold. This implies that standard estimation techniques that rely on the 'invertibility' of the term structure (e.g., Chen and Scott, 1993; Pearson and Sun, 1994) with respect to the latent factors cannot be implemented. The Kalman filter-based approaches of Duan and Simonato (1999) and de Jong (2000) are also unsuitable for our purposes because USV restrictions make it impossible for a linear filter to properly update the distribution of the unknown volatility state variable. We therefore write term structure dynamics in a nonlinear state space form and estimate the parameters of the models using Bayesian Markov chain Monte Carlo (MCMC).

The results from estimating the unrestricted essentially affine three-factor model are striking. Most significantly, we obtain the 'self-inconsistent' result that the volatility factor extracted from this model (i.e., the 'term structure-implied volatility') is basically unrelated to volatilities estimated using rolling windows, GARCH volatilities, or implied volatilities from options. Furthermore, the strong in-sample fit of that model breaks down following the end of the estimation period, suggesting deep misspecification.

We interpret these findings as evidence that three-factor models cannot simultaneously describe the yield curve's level, slope, curvature, and volatility. That is, volatility is unable to play the dual role that such a model predicts it does. The estimation of such a model therefore presents a tradeoff between choosing volatility dynamics that are more consistent with one role or the other. For the data set we investigate, and with no parameter

[3] We note that recent empirical literature has reported that the affine class has trouble simultaneously fitting certain cross-sectional and time-series properties of the yield curve (Duffee, 2002; Dai and Singleton, 2002; Duarte, 2004). Indeed, these papers suggest that a more flexible risk premium structure is used to reconcile the time-variation in conditional variances and the forecasting power of the slope of the term structure. Further, Cochrane and Piazzesi (2005) find a state variable that drives risk premia but that is separate from the three factors affecting yield curve shape. The tradeoff we uncover here involves second-order moments, and thus is independent of the risk premia structure. Note that since the volatility structure is invariant under transformation from the historical measure to the risk-neutral measure, proposing a more general risk premia specification will not overcome this problem as it did in Duffee (2002).

[4] For example, Collin-Dufresne, Goldstein, and Jones (2008) propose a representation of affine models in terms of economically meaningful state variables (such as level, slope, and curvature) that can be estimated model-independently. They argue that if a model is well-specified, then state variables 'inverted' from prices using standard econometric techniques should be closely related to the model-independent estimates. They find for the $A_1(3)$ model that there are substantial deviations between the two when volatility is used as an observable state variable.

restrictions imposed, that tradeoff is heavily tilted toward explaining the cross section.[5]

Both the $A_1(3)$ and $A_1(4)$ models exhibiting USV imply realistic behavior for the dynamics of short rate volatility, but the $A_1(3)$ USV model fails on (at least) two dimensions. First, with just two factors affecting the cross section of yields, it cannot provide the same accuracy in fitting the cross section as does the unrestricted three-factor model or the four-factor USV model. Second, the unconditional yield volatilities implied by the model are inconsistent with the data, as they fail to reproduce the 'hump shape' relation between unconditional volatility and maturity found in the data. Therefore, in this paper, we report only the results for the $A_1(4)$ USV model and compare them to the results for the unrestricted $A_1(3)$ and $A_1(4)$ models.[6]

Of the models investigated, only the $A_1(4)$ USV model is able to generate both good cross-sectional and time-series fits of yields. Since it has three factors that affect yields, the model's in-sample fit is very tight. Furthermore, in contrast to the unconstrained $A_1(3)$ model, this model is just as accurate (if not more so) in our three-year 'hold-out' sample. The model is able to simultaneously forecast volatilities of yields at all maturities, both in- and out-of-sample, and it generates the correct hump shape in the term structure of unconditional volatilities. Given our use of a relatively short sample period, we were unable to gauge the model's yield forecasting performance with much accuracy, but we note that the model has at least as much flexibility in its risk premia as the most general models considered by Duffee (2002).

While the unconstrained $A_1(4)$ model performs better than the unconstrained $A_1(3)$ model in that its predicted volatility is at least positively correlated with volatility estimates such as GARCH, it is still grossly inadequate at capturing the volatility of short rates. In particular, for short maturity yields, the USV model outperforms the unconstrained $A_1(4)$ model both in- and out-of-sample when it comes to volatility forecasts. Further, the USV model, fitted to weekly data, delivers forecasts that are equal to or better than GARCH forecasts based on daily data. However, for longer maturities, the USV and non-USV models seem to perform similarly with respect to volatility predictions, and both are inferior to GARCH. These results indicate that within the unconstrained four-factor model there is a tension between capturing the unconditional distribution of yields and the short rate volatility dynamics. These results also suggest that a better model would allow for different drivers for the volatilities of short and long maturity yields. An important implication of our findings is that any strategy that attempts to hedge the volatility risk inherent in fixed income derivatives (if feasible at all) must be substantially more complex than the convexity-based 'butterfly' positions discussed by Litterman, Scheinkman, and Weiss (1991) and implied in Brown and Schaefer (1994a, b). Indeed, our results suggest that implied spot rate volatility measures extracted from the cross section of the yield curve are likely to be bad estimates of actual volatility.

The rest of the paper is as follows. In Section 2, we characterize maximal three- and four-factor models exhibiting USV. In Section 3, we describe an estimation methodology that remains valid under USV, while Section 4 includes all empirical results. We conclude in Section 5.

2. Stochastic volatility affine models of the term structure

Mostly following the notation of DK and DS, the risk-neutral dynamics of a Markov N-dimensional state vector X within an affine framework can be specified as

$$dX_t = \mathcal{K}^{Q}(\Theta^{Q} - X_t)\,dt + \Sigma\sqrt{S_t}\,dZ_t^{Q}, \tag{1}$$

where Z^{Q} is a vector of M $(M \geqslant N)$[7] independent Brownian Motions, $\mathcal{K}^{Q}$ is an $N \times N$ matrix, Σ is $N \times M$, and S is a diagonal $M \times M$ matrix with components

$$S_{ii,t} = \alpha_i + \beta_i^{\top} X_t. \tag{2}$$

The spot rate is an affine function of X:

$$r_t = \delta_0 + \delta_x^{\top} X_t, \tag{3}$$

where δ_x is an N-dimensional vector. The affine class generalizes specifications proposed by Vasicek (1977), Cox, Ingersoll, and Ross (1981, 1985), as well as many others.

Assuming the system is 'admissible,'[8] zero-coupon bond prices take the form:

$$P_t(\tau) = e^{A(\tau) - B(\tau)^{\top} X_t}, \tag{4}$$

where $\tau \equiv (T - t)$ and where $A(\tau)$ and $B(\tau)$ satisfy the ODEs:

$$\frac{dA(\tau)}{d\tau} = -\Theta^{Q\top}\mathcal{K}^{Q\top} B(\tau) + \frac{1}{2}\sum_{i=1}^{M}[\Sigma^{\top} B(\tau)]_i^2 \alpha_i - \delta_0,$$

$$\frac{dB(\tau)}{d\tau} = -\mathcal{K}^{Q\top} B(\tau) - \frac{1}{2}\sum_{i=1}^{M}[\Sigma^{\top} B(\tau)]_i^2 \beta_i + \delta_x,$$

and the initial conditions:

$$A(0) = 0, \quad B(0) = 0.$$

[5] Rather than imposing USV parameter constraints, an alternative method to 'tilt' the estimation procedure toward capturing volatility dynamics is to include data on derivative prices, which tend to be more sensitive to volatility dynamics (e.g., Jagannathan, Kaplin, and Sun, 2003; Bikbov and Chernov, 2009; Almeida, Graveline, and Joslin, 2006). We suspect that, when both options and yield data are used, the variance state variable will be more closely related to interest rate variance and less related to the shape of the yield curve. That does not, however, affect our findings that interest rate volatility is weakly correlated with the level, slope, and curvature of the yield curve.

[6] Results for the $A_1(3)$ USV model are available upon request.

[7] Collin-Dufresne, Goldstein, and Jones (2008) show that if one assumes state vector dynamics as in Eq. (1), then for some cases it is necessary to specify the model with more Brownian motions (M) than state variables (N) to avoid ruling out some identifiable specifications.

[8] That is, assuming that the stochastic differential equation admits a unique strong solution. Sufficient conditions are given in Duffie and Kan (1996).

Defining bond yields $Y_t(\tau)$ via $P_t(\tau) = e^{-\tau Y_t(\tau)}$, we see from Eq. (4) that yields are affine in the state variables:

$$Y_t(\tau) = -\frac{A(\tau)}{\tau} + \frac{B(\tau)^\top}{\tau} X_t. \tag{5}$$

All models we investigate below are special cases of this general affine model. Specifically, using the DS classification scheme, we focus on the $A_1(3)$ and $A_1(4)$ models. These are models with $N = 3$ and 4 state variables, respectively, only one of which drives stochastic volatility.

As is well-known, there are several equivalent ways to represent a given model.[9] Following Collin-Dufresne, Goldstein, and Jones (2008), we choose a representation for the state vector X in which each state variable has a clear economic interpretation. Specifically, we use $X = [r, \mu^Q, V]$ as the state vector for three-factor models and $X = [r, \mu^Q, \theta^Q, V]$ as the state vector for four-factor models, where the state variables are defined as follows:

$$r_t = Y_t(0), \tag{6}$$

$$\mu_t^Q = 2\frac{\partial Y_t(\tau)}{\partial \tau}\bigg|_{\tau=0}, \tag{7}$$

$$\theta_t^Q = 3\frac{\partial^2 Y_t(\tau)}{\partial \tau^2}\bigg|_{\tau=0}, \tag{8}$$

$$V_t = \left(\frac{1}{dt}\right) dr_t^2. \tag{9}$$

In other words, r is the level, μ^Q is two times the slope, and θ^Q is three times the curvature of the yield curve at short maturities. Finally, V is the variance of the short rate. CGJ discuss some of the advantages of this representation and show that[10]

$$\mu_t^Q = \mathrm{E}_t^Q[dr_t]/dt, \tag{10}$$

$$\theta_t^Q = \mathrm{E}_t^Q[d\mu_t^Q]/dt - V_t. \tag{11}$$

Note that Eq. (11) predicts that the short rate variance is intimately linked to curvature. We emphasize that this result assumes only the condition of no-arbitrage.

An interesting feature of this representation is that, since the state variables possess economic interpretations irrespective of the model, their implied time series can be directly compared across models (unlike latent variables). Further, this representation makes it easier to identify the restrictions on parameters necessary to obtain USV. We present these restrictions below.

2.1. The $A_1(3)$ model

Following CGJ, we present the $A_1(3)$ model in terms of the instantaneous risk-neutral means and covariances of the state vector $X = [r, \mu^Q, V]^\top$ as

$$\frac{1}{dt}\mathrm{E}^Q[dX_t] = \begin{bmatrix} \mu_t^Q \\ m_0 + m_r r_t + m_\mu \mu_t^Q + m_V V_t \\ \gamma_V - \kappa_V V_t \end{bmatrix} \tag{12}$$

and

$$\frac{1}{dt}\mathrm{Cov}(dX_t, dX_t^\top) \equiv \Omega_t = \Omega_0 + \Omega_V(V_t - \underline{V}), \tag{13}$$

where the parameter $\underline{V}$ sets the lower bound of the V_t process and where

$$\Omega_0 = \begin{bmatrix} \underline{V} & \underline{c}_{r\mu} & 0 \\ \underline{c}_{r\mu} & \underline{\sigma}_\mu & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \Omega_V = \begin{bmatrix} 1 & c_{r\mu} & c_{rV} \\ c_{r\mu} & \sigma_\mu & c_{\mu V} \\ c_{rV} & c_{\mu V} & \sigma_V \end{bmatrix}. \tag{14}$$

As shown in DS and CGJ, this model is *Q-maximal* in that it has the largest number of parameters that can be identified from the cross section of fixed income securities. For admissibility, we require that the matrix Ω_0 be positive semidefinite and Ω_V positive definite.

2.2. The $A_1(4)$ model

As discussed above, we specify the $A_1(4)$ model by adding an additional state variable to the model. This variable, θ^Q, measures the curvature of the yield curve at short maturities. We present the risk-neutral dynamics of the new state vector $X = [r, \mu^Q, \theta^Q, V]^\top$ in terms of its instantaneous drift and covariance matrix as

$$\frac{1}{dt}\mathrm{E}^Q[dX_t] = \begin{bmatrix} \mu_t^Q \\ \theta_t^Q + V_t \\ a_0 + a_r r_t + a_\mu \mu_t^Q + a_\theta \theta_t^Q + a_V V_t \\ \gamma_V - \kappa_V V_t \end{bmatrix} \tag{15}$$

and

$$\frac{1}{dt}\mathrm{Cov}(dX_t, dX_t^\top) \equiv \Omega_t = \Omega_0 + \Omega_V(V_t - \underline{V}), \tag{16}$$

where

$$\Omega_0 = \begin{bmatrix} \underline{V} & \underline{c}_{r\mu} & \underline{c}_{r\theta} & 0 \\ \underline{c}_{r\mu} & \underline{\sigma}_\mu & \underline{c}_{\mu\theta} & 0 \\ \underline{c}_{r\theta} & \underline{c}_{\mu\theta} & \underline{\sigma}_\theta & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and}$$

$$\Omega_V = \begin{bmatrix} 1 & c_{r\mu} & c_{r\theta} & c_{rV} \\ c_{r\mu} & \sigma_\mu & c_{\mu\theta} & c_{\mu V} \\ c_{r\theta} & c_{\mu\theta} & \sigma_\theta & c_{\theta V} \\ c_{rV} & c_{\mu V} & c_{\theta V} & \sigma_V \end{bmatrix}. \tag{17}$$

This is the *maximal* $A_1(4)$ model, with a total of 22 free risk-neutral parameters. Similarly to the $A_1(3)$ model, we require for admissibility that the matrix Ω_0 be positive semidefinite and Ω_V positive definite.

[9] See, for example, the DS discussion of these 'invariant transformations.'

[10] This can also be verified by direct calculation on the bond price formula presented below.

2.3. The $A_1(4)$ USV model

Following the approach of CDG, we find that the $A_1(4)$ model exhibits USV if the following restrictions are imposed:

$$
\begin{aligned}
a_r &= -2c_{r\mu}^2(3c_{r\mu} - a_\theta),\\
a_\mu &= 7c_{r\mu}^2 - 3c_{r\mu}a_\theta,\\
a_V &= 3c_{r\mu},\\
\sigma_\mu &= c_{r\mu}^2,\\
\sigma_\theta &= c_{r\mu}^4,\\
c_{r\theta} &= c_{r\mu}^2,\\
c_{\mu\theta} &= c_{r\mu}^3.
\end{aligned}
$$

Substituting in the constraints on the risk-neutral drift and Ω_V, we obtain

$$
\begin{aligned}
\frac{1}{dt}\mathrm{E}^Q[d\theta_t^Q] = (a_0 + -2c_{r\mu}^2(3c_{r\mu} - a_\theta)r_t + (7c_{r\mu}^2 - 3c_{r\mu}a_\theta)\mu_t^Q\\
+ a_\theta\theta_t^Q + 3c_{r\mu}V_t)
\end{aligned}
\tag{18}
$$

and

$$
\Omega_V = \begin{bmatrix} 1 & c_{r\mu} & c_{r\mu}^2 & c_{rV} \\ c_{r\mu} & c_{r\mu}^2 & c_{r\mu}^3 & c_{\mu V} \\ c_{r\mu}^2 & c_{r\mu}^3 & c_{r\mu}^4 & c_{\theta V} \\ c_{rV} & c_{\mu V} & c_{\theta V} & \sigma_V \end{bmatrix}. \tag{19}
$$

To ensure that Ω_V is positive semidefinite, we must impose the following constraints on the parameters:

$$c_{\mu V} = c_{rV}c_{r\mu}, \tag{20}$$

$$c_{\theta V} = c_{rV}c_{r\mu}^2. \tag{21}$$

Thus, the $A_1(4)$ model that exhibits USV has a total of 13 risk-neutral parameters, the result of imposing nine restrictions on the 22 parameters of the unrestricted model. The following proposition verifies that the proposed model exhibits USV and provides the closed-form solution for bond prices.

Proposition 1. *If the short rate process follows a four-factor Markov process given by Eqs.* (15) *and* (16) *where the parameters satisfy the admissibility conditions* (18)–(21), *then zero-coupon bond prices are given by*

$$P_t(\tau) = \exp(A(\tau) - B_r(\tau)r_t - B_\mu(\tau)\mu_t^Q - B_\theta(\tau)\theta_t^Q), \tag{22}$$

where the deterministic functions $A(\tau)$, $B_r(\tau)$, $B_\mu(\tau)$, *and* $B_\theta(\tau)$ *are given in Appendix* B.

Proof. See Appendix B.

Clearly, the model displays USV in that a change in volatility does not affect the shape of the yield curve. Interestingly, despite the fact that the short rate has stochastic volatility, the expression obtained for the term structure displays strong similarities to that of a three-factor Gaussian model, as the $B(\cdot)$ functions are linear combinations of exponentially decaying functions of maturity.

From the expression for bond prices in Appendix B, it is clear that only a strict subset of the 13 risk-neutral parameters can be identified from bond prices alone. Time-series information is thus necessary to identify all of the parameters. However, even using both cross-sectional and time-series data on bond prices, we cannot determine the risk-neutral drift parameters (γ_V, κ_V) of V since these parameters affect neither the prices of bonds nor their physical dynamics.[11] Rather, prices of other fixed income derivatives (e.g., caps) must be used to infer these risk-neutral parameters.

Note that bond prices would retain their exponential-affine form in the above model for *any* specification of the process for V_t. Indeed, the proof of Proposition 1 does not depend on the specific process followed by the variance of the short rate.[12] In other words, bond prices can be exponential-affine even if state vector dynamics are not! This could prove helpful in estimating more general models for the volatility dynamics while retaining the analytical tractability of affine models for bond prices.

3. Empirical approach

The primary focus of this paper is to investigate whether standard affine models can simultaneously explain both the cross-sectional and time-series properties of bond prices. In this section, we use data on U.S. Dollar swap and LIBOR rates to estimate the maximal three- and four-factor affine models specified above. We begin by discussing the specification of risk premia and the implied dynamics under the historical measure. We then discuss the data and empirical methodology. Finally, the results are presented.

3.1. Risk premia

In Section 2, we presented the risk-neutral, or Q dynamics for the three- and four-factor models. To complete the model we also need to specify the historical, or P measure dynamics for the state vector. In order to rule out arbitrage opportunities, the two measures must be 'equivalent' in that they must agree on zero-probability events. One implication is that the risk premia specification will change only the drift of the process.[13] For the $A_1(3)$ model, we specify the P drift for the state vector to be

$$
\frac{1}{dt}\mathrm{E}^P[dX] = \begin{bmatrix} \lambda_{r0} + \lambda_{rr}r_t + (1 + \lambda_{r\mu})\mu_t^Q + \lambda_{rV}V_t \\ (m_0 + \lambda_{\mu 0}) + (m_r + \lambda_{\mu r})r_t + (m_\mu + \lambda_{\mu\mu})\mu_t^Q \\ \quad +(m_V + \lambda_{\mu V})V_t \\ (\gamma_V + \lambda_{V0}) - (\kappa_V - \lambda_{VV})V_t \end{bmatrix}. \tag{23}
$$

[11] This statement assumes that the risk premia are general enough so that the risk-neutral parameters (γ_V, κ_V) are distinct from their physical-measure counterparts.

[12] The only condition is that the volatility process be sufficiently regular for the discounted price process to be a martingale.

[13] See Harrison and Kreps (1979) and Harrison and Pliska (1981).

Analogously, for the $A_1(4)$ model, we specify the P drift for the state vector to be

$$\frac{1}{dt}\mathrm{E}^P[dX_t] = \begin{bmatrix} \lambda_{r0} + \lambda_{rr} r_t + (1+\lambda_{r\mu})\mu_t^Q + \lambda_{r\theta}\theta_t^Q + \lambda_{rV} V_t \\ \lambda_{\mu 0} + \lambda_{\mu r} r_t + \lambda_{\mu\mu}\mu_t^Q + (1+\lambda_{\mu\theta})\theta_t^Q \\ \quad +(1+\lambda_{\mu V})V_t \\ (a_0 + \lambda_{\theta 0}) + (a_r + \lambda_{\theta r}) r_t + (a_\mu + \lambda_{\theta\mu})\mu_t^Q \\ \quad +(a_\theta + \lambda_{\theta\theta})\theta_t^Q + (a_V + \lambda_{\theta V})V_t \\ (\gamma_V + \lambda_{V0}) - (\kappa_V - \lambda_{VV})V_t \end{bmatrix}. \tag{24}$$

These 'essentially affine' risk premia structures are the most general structures that maintain admissibility and the affine structure under both the P and Q measures. In order to ensure that the two measures are equivalent, it is sufficient to specify that V_t cannot reach its lower bound $\underline{V}$ under both measures.[14] We therefore impose the Feller condition for both processes as a constraint in our analysis.[15]

With this specification, the unrestricted $A_1(3)$ model has a total of 24 parameters (14 risk-neutral and 10 risk premium parameters). The unrestricted $A_1(4)$ model has a total of 39 parameters (22 risk-neutral and 17 risk premium parameters). The $A_1(4)$ USV model contains 17 risk premia and 13 risk-neutral parameters, but only 11 of the latter are identified from bond prices. For the USV model we can therefore only report the P measure volatility drift parameters, which are each the sum of a risk-neutral drift parameter and the corresponding risk premia.

3.2. Data

We use weekly LIBOR and swap rate data from Datastream from January 6, 1988, to December 29, 2005. On each day in the sample, zero-coupon yield curves are bootstrapped from all available swap rates and the six-month LIBOR rate. For dates before January 1997, when the one-year swap rate first became available, we also use the one-year LIBOR rate. We adjust observed LIBOR quotes to account for the fact that they are recorded approximately six and a half hours prior to the time at which swap rates are quoted. Details of this adjustment are provided by Jones (2003a). Finally, we convert all yields to zero-coupon rates assuming that swaps can be valued as par bond rates.[16] Following Bliss (1997), we use the extended Nelson-Siegel method for bootstrapping.

From the bootstrapped yield curves we extract yields with maturities of 0.5, 1, 2, 3, 4, 5, 7, and 10 years. We choose these eight maturities because on each day in the sample there is some underlying yield quote for each one. We therefore expect the bootstrapped yields to be particularly accurate for these maturities.

3.3. Posterior sampler

We estimate all models using a Bayesian approach that combines elements of Polson, Stroud, and Müller (2001), Lamoureux and Witte (2002), Jones (2003b), Sanford and Martin (2003), and Bester (2004). In each of these papers, data augmentation and a Gibbs-like posterior sampler are used to simplify the computation of posterior distributions of the model parameters. As is common in continuous time finance models, likelihood-based inference is difficult because of imperfectly observed state variables and the fact that transition densities are not known in closed form.

Bayesian data augmentation is attractive because it solves both of these problems. As in Elerian, Chib, and Shephard (2001), Eraker (2001), and Jones (2003b), augmenting with unobservable high frequency data increases the accuracy of the Euler approximation, providing a Gaussian transition density that is easier to work with. Data augmentation also allows us to augment the observed yield data with the term structure factors (i.e., the X's) themselves. While the latter use of data augmentation is critical for our analysis, augmenting with high frequency data turns out to be inconsequential, as a simple Euler approximation applied directly to our weekly data does not appear to inject bias into our results. Nevertheless, our ability to implement more accurate likelihood approximations is still valuable in that it enables us to assess the validity of a simpler approach.

As in Pennacchi (1991), Brandt and He (2006), and Bester (2004), we assume that all data are observed with error. Unlike these papers, we fit our models to the *principal components* of yields rather than the yields themselves. We do so because cross-sectional correlations between model errors in principal components are close to zero, while correlations between yield errors are often above 0.9. This makes our use of a diagonal covariance matrix for the observation errors much more reasonable.[17] The principal component loadings and percentages of variance explained appear in Table 1. As in Litterman and Scheinkman (1991), the first three principal components explain most variation in yields and can roughly be interpreted as level, slope, and curvature.

While the posterior simulator is described in much more detail in Appendix C, we briefly outline our approach here. Letting $\mathscr{P} = \{\mathscr{P}_1, \mathscr{P}_2, \ldots, \mathscr{P}_T\}$ denote the time series of principal component vectors and ϕ the vector of model parameters, we seek to compute $p(\phi|\mathscr{P}) \propto p(\mathscr{P}|\phi)p(\phi)$, where the first term on the right is the likelihood and the

[14] See, for example, Theorem 7.19 in Liptser and Shiryayev (1974, p. 294), and Cheridito, Filipovic, and Kimmel (2007).

[15] The Feller condition for the risk-neutral measure parameters of the process $(V - \underline{V})$ is $2(\gamma_V - \kappa_V \underline{V}) > \sigma_V$. A similar condition applies for the P measure parameters.

[16] If swaps were free of default risk, this would directly follow from absence of arbitrage. In the presence of credit-risk, this assumption is warranted if there is homogeneous credit quality across swap and LIBOR markets. In that case, the zero-coupon curve corresponds to a risk-adjusted corporate curve for issuers with refreshed AA credit quality (see Duffie and Singleton, 1997; Collin-Dufresne and Solnik, 2001; Johannes and Sundaresan, 2006).

[17] An appealing alternative approach used by Brandt and He (2006) is to parameterize the covariances as parsimonious functions of bond maturities and a few free parameters.

Table 1
Principal component loadings.

The table contains the eigenvectors corresponding to the eigenvalues of the covariance matrix of changes in bootstrapped zero-coupon yields from January 1988 to December 2002. They represent the loadings on yields of different maturities used to construct the principal components. The table also reports the percent of the total variance explained by each of the principal components.

	Principal Component					
	1	2	3	4	5	6
6-month	0.08	2.27	1.98	14.63	14.67	103.62
1-year	0.11	1.85	0.07	−17.54	−34.76	−339.62
2-year	0.14	0.95	−1.23	−11.78	18.22	590.44
3-year	0.14	0.19	−1.15	4.07	21.33	−173.48
4-year	0.14	−0.38	−0.73	12.14	1.91	−410.20
5-year	0.14	−0.81	−0.23	13.19	−16.60	−133.93
7-year	0.13	−1.36	0.68	3.98	−26.17	617.28
10-year	0.12	−1.72	1.60	−17.70	22.40	−253.12
% explained	64.82	17.79	7.98	5.32	2.75	1.07
Total % explained by first six principal components: 99.73						

second is the prior. Following earlier approaches, we augment the observable data $\mathscr{P}$ with the term structure factor data $X=\{X_1,X_2,\ldots,X_T\}$. We then integrate out the X's using a Gibbs-like posterior simulator that alternates between performing draws from $p(\phi|\mathscr{P},X)$ and $p(X|\mathscr{P},\phi)$. Under very weak conditions, the resulting sequence of draws of ϕ converges in distribution to our target, the posterior distribution $p(\phi|\mathscr{P})$.

We approximate the true dynamics using the Euler scheme

$$(X_{t+h}-X_t)\sim \mathrm{N}(h(a+bX_t),h\Omega_t), \tag{25}$$

where Ω_t is the instantaneous covariance matrix. In particular, we do not use the true noncentral chi-squared distribution for the square root factors, as in Lamoureux and Witte (2002) and Bester (2004), but instead choose h to be small enough to approximate that distribution.[18] The likelihood function is completed by specifying the relation between the data and the state vector. Given the linearity of bond yields in state variables (5) and the linear relation between principal components and yields,

$$\mathscr{P}_t = \text{PC loadings}\times Y_t,$$

it is easy to see that there is a linear relation between principal components and state variables. Adding a Gaussian error vector $e_t\sim \mathrm{N}(0,\Lambda)$ results in

$$\mathscr{P}_t = K+LX_t+e_t, \tag{26}$$

which is our 'measurement equation.'

Similar to Bester (2004), we find it efficient to further break up the parameter vector into three components, ϕ^{Q}, ϕ^{λ}, and ϕ^{Λ}, where ϕ^{Q} contains all parameters that affect the dynamics of the state vector under the risk-neutral measure. Risk premia parameters comprise ϕ^{λ}, while ϕ^{Λ} includes the measurement error standard deviations. Both of the latter draws are made from closed form densities, with the distribution of ϕ^{λ} following directly from the linear Gaussian structure of the Euler approximation.

Building on Polson, Stroud, and Müller (2001), we decompose the state vector as

$$X_t=\begin{bmatrix}X_t^o\\ V_t\end{bmatrix}, \tag{27}$$

where X^o includes all state variables other than V (i.e., r, μ^{Q}, and, if applicable, θ^{Q}). The reason for doing so is that only V_t affects the factor covariance matrix. As such, once we condition on the entire path of V, we can then write the dynamics of X_t^o and $\mathscr{P}_t$ in linear Gaussian state space form. This enables us to draw the entire multivariate time series X^o at once in closed form using the simulation smoother of de Jong and Shephard (1995). This means that only the draws of V must be made using a relatively inefficient approach involving a separate Metropolis-Hastings draw for each t.

Our use of the Kalman filter is significantly different from that of other studies. Pennacchi (1991), Duan and Simonato (1999), and de Jong (2000) apply the Kalman filter to affine models in a more straightforward manner by including all term structure factors in the state equation, including those that follow square root processes and that impact conditional volatilities. While the Kalman filter is very naturally applied in homoskedastic Gaussian models (e.g., as in Pennacchi, 1991), its validity is not as straightforward when covariances are state-dependent. In short, the problem with conventional linear filters is that filtered estimates of the state variables are simple projections on the observed yields and do not take into account, for example, the quadratic variation in those yields. Thus, when the state vector includes variables that drive yield volatility, a substantial amount of the relevant information in the data is ignored. The result, as Lund (1997) and de Jong (2000) argue, is an incorrect specification of conditional variances, which in turn leads to inconsistent estimates.

Interestingly, however, these studies, as well as that of Duffee and Stanton (2002), have found that methods based on the Kalman filter perform well in simulated samples, with minimal biases and relatively high accuracy. A natural explanation for this result is that the models that they consider are all models with *spanned* stochastic volatility. In that case, the levels of yields may be sufficient to infer all state variables with high accuracy, including those that drive conditional volatilities, and ignoring information in quadratic variation (as well as other nonlinearities) is therefore likely to be innocuous.

In the USV case, however, this result cannot hold since the levels of yields carry no information whatsoever about the volatility state variable, making the inconsistency identified by Lund (1997) and de Jong (2000) particularly severe. Our approach, like that of Polson, Stroud, and Müller (2001), is immune to this criticism because the Kalman filter is only applied as a computational device to evaluate the likelihood *conditional* on a given path of V. This means that the only state variable uncertainty is among 'Gaussian' elements of the state vector (the X_t^o),

[18] Note that the transition density is only known in closed form for one (i.e., volatility) of the three state variables.

which do not impact covariances. This avoids the source of inconsistency for linear filters.

Lastly, we must specify prior distributions. For all cases considered, we specify the priors as either completely flat or exactly proportional to a constant, with three exceptions. First, the prior density for each measurement error standard deviation $\sqrt{\Lambda_{i,i}}$ is proportional to $1/\sqrt{\Lambda_{i,i}}$, as is standard. Second, all regions of the parameter space in which the model is nonstationary, inadmissible, or in violation of the Feller condition are assumed to have zero prior probability. This means that any parameter draw that violates any of these conditions is immediately rejected. Finally, all parameter draws that generate covariance degeneracies are assumed to have zero probability. This means, for instance, that the Ω_t matrix must have full rank for all $V_t > \underline{V}$ and that each $\Lambda_{i,i}$ is positive.

3.4. Point estimates and confidence intervals

For each parameter, we report point estimates along with 95% confidence intervals computed from the 2.5 and 97.5 percentiles of the MCMC draws from the posterior distribution. Let ϕ_i denote the ith MCMC draw of the parameter vector from its posterior distribution and $\tilde{\phi}_i$ be the same vector normalized by the posterior standard deviations. We then find the draw i that minimizes the average value of $|\tilde{\phi}_j - \tilde{\phi}_i|$ over all $j \neq i$. The resulting ϕ_i is the L_1 center of the normalized posterior distribution, a version of the multivariate posterior median, and is taken as our vector of point estimates. Raftery (1996) reports that the unnormalized version often provides an accurate approximation of the posterior mode. Because of the vastly different scales of some of the model parameters, we felt that normalization would be more natural. Using multivariate rather than univariate medians or means is appropriate because univariate medians or means are not necessarily located in regions of high posterior probability.

From MCMC output, it is possible to compute posteriors of functions of parameters in addition to the parameters themselves. We therefore report posterior statistics on restricted parameters, such as the parameter a_V in the $A_1(4)$ USV model, which is restricted to equal $3c_{r\mu}$. As before, confidence intervals are calculated from the 2.5 and 97.5 percentiles of the posterior draws, while the restricted point estimates are computed directly from the point estimates of the unrestricted parameters.

4. Empirical results

We present and compare the results of estimation for the unconstrained $A_1(3)$ and $A_1(4)$ as well as the $A_1(4)$ USV model. The latter is really a three-factor model of the cross section of bond prices that allows for independent variation in short rate (and therefore bond price) volatility. We focus in turn on parameter estimates, in- and out-of-sample yield curve fit, implied time series of the state variables, in- and out-of-sample volatility fit, and volatility forecasting regressions.

4.1. Posterior summaries for parameter estimates

Tables 2A and B report posterior distributions for the parameters of the three models we estimate. All models are estimated using weekly data from January 1988 to December 2002, whereas data from 2003–2005 is saved for out-of-sample analysis. We compute posteriors by setting the discretization parameter h equal to 1 because

Table 2A
Posterior distributions of model parameters.

This table contains parameter point estimates and confidence intervals for three models fitted to weekly bootstrapped yields from January 1988 to December 2002. All parameter values are reported on an annualized basis. Both free parameters and restricted parameters are included, where restricted parameters are functions of the free parameters and are displayed in italics. For free parameters, the point estimates displayed are multivariate posterior medians. Point estimates for restricted parameters are computed as functions of the free parameter point estimates. Confidence interval bounds, in parentheses, are equal to the 2.5 and 97.5 percentiles of the posterior distribution.

	$A_1(3)$	$A_1(4)$ USV	$A_1(4)$
$a_0 \times 10^2$ or m_0	0.211	0.140	5.957
	(0.188, 0.238)	(0.118, 0.163)	(5.284, 8.400)
a_r or m_r	−1.238	*−0.015*	−0.009
	(−1.348, −1.131)	*(−0.019, −0.013)*	(−0.012, −0.001)
a_μ or m_μ	−2.804	*−0.282*	−0.661
	(−2.959, −2.597)	*(−0.324, −0.252)*	(−0.758, −0.587)
a_θ		−1.287	−3.020
		(−1.425, −1.194)	(−3.244, −2.917)
a_V or m_V	−923.1	*−0.260*	−809.5
	(−1003.1, −811.2)	*(−0.272, −0.248)*	(−1104.9, −738.1)
$\gamma_V^P \times 10^4$	0.568	0.803	0.733
	(0.126, 1.018)	(0.469, 1.147)	(0.330, 1.547)
κ_V^P	0.441	1.014	−0.926
	(0.756, 0.056)	(2.116, 0.412)	(−1.933, −0.406)
$\underline{V} \times 10^4$	0.111	0.003	0.245
	(0.046, 0.298)	(0.000, 0.011)	(0.124, 0.440)
$\underline{\sigma}_\mu \times 10^4$	5.809	3.448	10.31
	(3.208, 11.349)	(2.675, 3.999)	(1.91, 23.29)
$\underline{\sigma}_\theta \times 10^4$		5.306	82.4
		(3.711, 7.412)	(16.0, 212.8)
$\underline{c}_{r\mu} \times 10^4$	0.708	−0.102	−0.765
	(0.187, 0.926)	(−0.197, −0.014)	(−2.333, 0.226)
$\underline{c}_{r\theta} \times 10^4$		0.132	2.205
		(0.016, 0.260)	(−0.750, 6.949)
$\underline{c}_{\mu\theta} \times 10^4$		−4.230	−29.12
		(−5.379, −3.112)	(−70.36, −5.56)
σ_μ	17.660	*0.008*	35.37
	(11.327, 24.871)	*(0.007, 0.008)*	(23.05, 50.80)
$\sigma_\theta \times 10^2$		*0.0056*	39009
		(0.0047, 0.0068)	(28037, 57807)
$\sigma_V \times 10^4$	0.025	1.221	0.065
	(0.022, 0.030)	(0.811, 1.968)	(0.044, 0.088)
$c_{r\mu}$	−3.617	−0.087	−4.055
	(−4.526, −2.893)	(−0.091, −0.083)	(−5.714, −2.502)
$c_{r\theta} \times 10^2$		*0.751*	1255
		(0.682, 0.825)	(820, 1834)
$c_{rV} \times 10^3$	−0.376	2.685	0.805
	(−0.504, −0.209)	(0.152, 4.457)	(0.489, 1.346)
$c_{\mu\theta} \times 10$		*−0.0065*	−1165
		(−0.0075, −0.0056)	(−1695, −801)
$c_{\mu V}$	−1.327	*−0.233*	−0.010
	(−2.148, −0.806)	*(−0.388, −0.013)*	(−0.014, −0.007)
$c_{\theta V} \times 10^3$		*0.020*	36.87
		(0.001, 0.034)	(28.16, 50.83)

Table 2B
Posterior distributions of risk premia parameters.

This table contains parameter point estimates and confidence intervals for three models fitted to weekly bootstrapped yields from January 1988 to December 2002. All parameter values are reported on an annualized basis. Point estimates displayed are multivariate posterior medians. Confidence interval bounds, in parentheses, are equal to the 2.5 and 97.5 percentiles of the posterior distribution.

	$A_1(3)$	$A_1(4)$ USV	$A_1(4)$
λ_{r0}	−0.002	−0.021	0.075
	(−0.060, 0.053)	(−0.048, −0.005)	(−0.010, 0.194)
λ_{rr}	−0.043	0.334	0.115
	(−0.323, 0.355)	(0.037, 0.717)	(−0.173, 0.570)
$\lambda_{r\mu}$	−0.408	0.211	1.914
	(−0.601, −0.138)	(−1.367, 1.622)	(−0.290, 4.779)
$\lambda_{r\theta}$		0.332	0.765
		(−1.165, 1.308)	(0.000, 1.722)
λ_{rV}	−46.3	−190.4	−1433.3
	(−388.5, 279.3)	(−277.7, −68.2)	(−3280.1, −53.0)
$\lambda_{\mu 0}$	−0.136	0.088	0.053
	(−0.377, 0.113)	(0.021, 0.152)	(−0.502, 0.551)
$\lambda_{\mu r}$	0.329	−1.537	−0.682
	(−0.890, 1.960)	(−2.396, −0.441)	(−2.558, 1.226)
$\lambda_{\mu\mu}$	0.727	−2.889	0.030
	(−0.198, 1.891)	(−8.153, 1.418)	(−14.792, 11.514)
$\lambda_{\mu\theta}$		−1.691	−0.482
		(−5.885, 2.007)	(−5.293, 3.606)
$\lambda_{\mu V}$	1017.8	345.8	−43.8
	(−769.8, 2193.8)	(78.5, 603.5)	(−8030.1, 8745.3)
$\lambda_{\theta 0}$		−0.084	−0.341
		(−0.166, 0.000)	(−1.771, 1.401)
$\lambda_{\theta r}$		1.548	1.725
		(0.145, 2.631)	(−4.036, 7.235)
$\lambda_{\theta\mu}$		1.915	−5.367
		(−3.572, 8.438)	(−38.655, 39.566)
$\lambda_{\theta\theta}$		0.724	−0.443
		(−4.001, 5.786)	(−12.211, 14.138)
$\lambda_{\theta V}$		−333.7	3349.5
		(−649.0, 17.0)	(−24021.7, 26484.3)
$\lambda_{V0} \times 10^4$	−0.544		0.306
	(−0.994, −0.102)		(−0.113, 1.102)
λ_{VV}	0.475		−0.338
	(0.090, 0.790)		(−1.325, 0.203)

posteriors computed by setting $h = 0.2$ (summarized in Appendix C) resulted in no substantive differences for the $A_1(3)$ USV model but required considerably greater computational effort.

We report parameters for the risk-neutral process in addition to those of the risk premia, so P measure drift parameters are implied. The exceptions are the risk-neutral parameters γ_V and κ_V, which are not reported since they are not identifiable under USV. Instead, we report their P measure counterparts $\gamma_V^P \equiv \gamma_V + \lambda_{V0}$ and $\kappa_V^P \equiv \kappa_V - \lambda_{VV}$.

We draw two interesting observations from Tables 2A and B. First, the parameters of the volatility process are dramatically different for the USV model than for both unrestricted three- and four-factor models. Further, the USV restrictions appear to be clearly rejected by the unrestricted $A_1(4)$ model. These results suggest very different volatility dynamics implied by USV vs. non-USV models. These results also suggest that the USV restrictions may prevent the model from fitting the same moments that the unconstrained model likelihood emphasizes. We will present clear evidence of this tradeoff between fitting volatility vs. fitting cross section of yields below.

Second, many of the risk premia parameters seem to be imprecisely estimated. Their estimates, magnitudes, and signs seem to vary dramatically between USV and non-USV specifications. Furthermore, most of the additional risk premia parameters introduced in the four-factor model are indistinguishable from zero, indicating that the generalized essentially affine risk premia are likely overparameterized for this model. We present further evidence of this when looking at out-of-sample fits below.

4.2. In- and out-of-sample yield fit

Table 3 reports the models' in- and out-of-sample fits of the yield curve in terms of bias and root mean squared errors (RMSE) for each maturity. We first describe the methodology used to obtain these results and then discuss the findings.

4.2.1. Methodology

Throughout the rest of the paper, we evaluate model performance using the point estimates from Tables 2A and B. While it would be preferable to integrate over the entire posterior distribution, this turns out to be very computationally demanding for most of the analysis we perform. For computing fitted yields, we therefore rerun our posterior sampler only where the state variables are sampled and the parameter values are held fixed.[19]

We report the average error and RMSE of each model for each maturity, where errors are defined as actual yields minus fitted yields and fitted yields are computed via Eq. (5). We also compare the RMSEs of the $A_1(4)$ USV model to those of the two unrestricted three- and four-factor models using the method of Diebold and Mariano (1995). For each model, we compute forecast errors, say $\hat{e}_{1,t}$ and $\hat{e}_{2,t}$, and calculate t-statistics for the difference in squared forecast errors

$$\hat{e}_{1,t}^2 - \hat{e}_{2,t}^2. \tag{28}$$

In this case, a significantly positive mean would indicate the superiority of model 2 over model 1.

For all statistics, standard errors are calculated using the method of Newey and West (1987). For ease of comparison, all standard errors in a given panel are calculated using the same lag length. These lag lengths were chosen by calculating the optimal lag length for each series individually using the method of Newey and West (1994) and then averaging those optimal lags across series. The same procedure is used throughout the paper. If two RMSEs are significantly different, they are separated by an inequality sign signifying the direction of the rejection of the null, along with either one (for 5%) or two (for 1%) stars signifying the level of significance.

[19] In a few places, such as Table 3, where it is feasible to compute results both by integrating over the posterior and by conditioning on the multivariate medians, we find extremely minor differences.

Table 3
In-sample and out-of-sample yield fits.

This table contains statistics for both the in-sample and out-of-sample fits of zero-coupon yields (Y). For each model, fitted yields ($\hat{Y}_t$) are calculated for 0.5, 1, 2, 3, 4, 5, 7, and 10-year maturities. The table examines the bias and root mean squared errors in $\hat{e}_t = Y_t - \hat{Y}_t$, where $\hat{Y}_t$ denotes the model fitted value. * and ** denote statistical significance at the 5% and 1% levels, respectively, where standard errors are calculated using the method of Newey and West (1987) with 21 lags for the first two panels, 9 for the third panel, and 8 for the bottom panel. For biases, statistical significance relates to the null hypothesis that the bias is zero. For RMSE, the statistical significance of the pairwise comparison of two models is reported, along with an inequality sign that reflects the direction of the rejection. The sample size is 782 weeks for in-sample results and 156 weeks for out-of-sample results.

	$A_1(3)$		$A_1(4)$ USV		$A_1(4)$
			In-sample mean $\hat{e}$ (basis points)		
6-month	−0.11		0.69		−0.21*
1-year	0.14		−0.70		0.29
2-year	0.06		−0.79		0.10
3-year	−0.03		0.08		−0.09
4-year	−0.08		0.69		−0.16**
5-year	−0.08		0.86		−0.13
7-year	−0.04		0.26		0.00
10-year	0.13		−1.04		0.18
			In-sample RMSE (basis points)		
6-month	3.39	<**	4.85	>**	0.94
1-year	4.48	<**	6.05	>**	1.94
2-year	2.52	<**	3.81	>**	1.13
3-year	1.42	<*	1.63	>**	1.08
4-year	2.56	<**	3.86	>**	0.54
5-year	2.86	<**	4.34	>**	0.95
7-year	1.57	<**	1.97	>**	1.31
10-year	3.95	<**	5.92	>**	1.23
			Out-of-sample mean $\hat{e}$ (basis points)		
6-month	−4.13**		−0.04		−0.12
1-year	4.73**		0.40		0.11
2-year	3.74**		−0.40		0.23
3-year	−0.93**		−0.58*		−0.06
4-year	−3.63**		−0.07		−0.23**
5-year	−4.15**		0.62		−0.19*
7-year	−1.31**		1.23**		0.18
10-year	5.41**		−1.16		0.07
			Out-of-sample RMSE (basis points)		
6-month	5.13	>**	2.99	>**	0.52
1-year	5.77	>*	3.98	>**	0.86
2-year	4.98	>**	1.90	>*	0.69
3-year	1.12		1.27	>**	0.39
4-year	4.61	>**	2.23	>**	0.48
5-year	5.44	>**	2.28	>**	0.57
7-year	1.97	>*	1.36	>**	0.49
10-year	7.09	>**	3.20	>**	0.65

4.2.2. Results

The top panel of Table 3 shows that all models imply reasonably unbiased in-sample fits of individual yields, with no rejections of zero mean errors. However, not reported in the table is the fact that the errors from all three models are highly autocorrelated, often with first-order autocorrelations above 0.8, an indication that all the models are misspecified.[20]

Given that it nests the other two models, it is not surprising that the root mean squared error criterion favors the unrestricted $A_1(4)$ model. Interestingly, between the unrestricted $A_1(3)$ and the $A_1(4)$ USV model, both with three factors explaining the cross section of bond prices, the former offers modest but significant reductions in RMSE in sample.

However, the situation reverses for out-of-sample yield fits using data from January 2003 to December 2005. While both of the $A_1(3)$ and $A_1(4)$ USV models display significant biases, as seen in the third panel, they are extremely large for the $A_1(3)$ specification. In addition, the errors are much larger in terms of RMSE for $A_1(3)$ than they are for $A_1(4)$ USV. Overall, the dramatic out-of-sample breakdown of the unrestricted $A_1(3)$ model is the most notable feature of Table 3. It is suggestive of some type of serious misspecification, the form of which we identify below.[21]

4.3. Properties of model-implied time series

We now examine some properties of the model-implied state variables that are estimated by running the posterior sampler with parameters held fixed at the point estimates from Tables 2A and B. The resulting draws of the state variables are then averaged to get smoothed estimates.

Table 4 reports a variety of correlations between observed time series and related model-implied variables. Using data from the estimation sample (1988–2002), we see that every model matches both the average yield (defined as the average of the 0.5, 1, 2, 5, 7, and 10-year yields), the slope of the yield curve (defined as $Y_{10y} - Y_{6m}$), and curvature (defined as $Y_{10y} - 2Y_{3y} + Y_{6m}$). The correlation between model-implied and observed state variables is greater than 0.997 for all models. This is not surprising since all models have at least three variables entering the cross section of bond prices, and level, slope, and curvature are known to be the foremost variables explaining the shape of the yield curve. What is more surprising is the tremendous disparity in the time series of volatility implied by the three models. Only for the USV model do we see a large positive correlation (0.78) between model-implied volatilities and model-free volatilities computed using a 30-day rolling window estimator applied to daily changes in the six-month yield. In contrast, the unrestricted models display either low correlation (0.15) for the $A_1(4)$ or even negative (−0.60!) correlation for the $A_1(3)$ model. Very similar results are obtained when we use Bollerslev's (1986) GARCH(1,1) model instead of the rolling window estimator.

The panels on the left-hand-side of Fig. 1 illustrate the inability of the unrestricted three- and four-factor models to capture short rate volatility. In particular, they show the

[20] These results are available upon request.

[21] In unreported results, available upon request, we also estimated an unrestricted two-factor model and a three-factor USV model. The results clearly showed that at least three factors driving the cross section of yields are necessary to obtain a decent in- and out-of-sample yield fit. This is not surprising given the principal component analysis of Litterman and Scheinkman (1991).

Table 4
Correlations of observed and model-implied time series.

This table reports correlations between actual and model-implied series computed over the 1988–2002 sample period. Average yield is simply the average of the 0.5, 1, 2, 3, 4, 5, 7, and 10-year zero yields. Slope is defined as the 10-year yield minus the 6-month yield. Curvature is defined using the 3-year yield in addition. Rolling 30-day window and GARCH(1,1) volatilities are calculated from demeaned daily changes in the 6-month rate. The implied volatility series is obtained from short-term Eurodollar futures options. These data are only available starting in 1991, so correlations involving implied volatilities are calculated over the 1991–2002 sample period.

	$A_1(3)$	$A_1(4)$ USV	$A_1(4)$	Daily GARCH	Eurodollar implied vol
Actual vs. model average yield	1.000	1.000	1.000		
Actual vs. model slope	0.998	0.998	1.000		
Actual vs. model curvature	0.998	0.997	1.000		
Rolling vs. model volatility	−0.600	0.783	0.156	0.957	0.676
GARCH vs. model volatility	−0.587	0.786	0.140	1.000	0.693
Eurodollar implied vs. model volatility	−0.498	0.605	0.377	0.693	1.000
Actual curvature vs. model volatility	0.275	−0.087	−0.065	−0.052	−0.103
Actual curvature vs. model variance	0.285	−0.072	−0.073	−0.020	−0.124

resulting time series of $\mathrm{E}[\sqrt{V_t}|\mathscr{P},\hat{\phi}]$ for each specification along with the same 30-day trailing-window volatilities used above. A vertical dotted line denotes the end of the estimation period. For the USV model, implied and trailing-window volatilities track each other closely. In contrast, both unrestricted $A_1(3)$ and $A_1(4)$ models imply very little variation in short rate volatility. Indeed, the standard deviations of the time series of short rate volatilities for the (i) rolling window volatilities, (ii) $A_1(4)$ USV model, (iii) $A_1(3)$ model, and (iv) unrestricted $A_1(4)$ model are 0.36%, 0.32%, 0.04%, and 0.05%, respectively. Moreover, the variation that does exist in the unconstrained models appears to be mostly unrelated to the variation in the rolling window volatilities.

We also investigate volatilities for the 10-year yields. Unfortunately, all models do rather poorly. Below, we provide additional evidence suggesting that more than one state variable driving volatility is probably necessary in order to capture the term structure of volatilities. In particular, we find that the standard deviations of the time series of volatilities for the (i) rolling window volatilities, (ii) $A_1(4)$ USV model, (iii) $A_1(3)$ model, and (iv) unrestricted $A_1(4)$ model are 0.30%, 0.12%, 0.10%, and 0.02%, respectively. Note in particular that the unrestricted $A_1(4)$ model predicts 10-year yield volatilities that are 18 times less volatile than the rolling window volatility estimates!

As a robustness check with respect to our estimates of 'realized' volatility, we also report in Table 4 the correlations of model volatilities with implied volatilities from short-maturity Eurodollar futures options. The Eurodollar series was constructed from CBOE data beginning in 1991. Note that implied volatilities are quoted using the model of Black (1976), implying that they should be interpreted as the volatility of proportional changes in (or logarithms of) Eurodollar futures prices. As such, we report correlations with the product of Eurodollar implied volatility with the Eurodollar futures price. Under reasonable assumptions, this should approximate the *level* of short-term interest rate volatility, making it comparable to the other volatility proxies we consider.

We find that the USV model-implied volatility series is highly positively correlated (0.6) with our implied volatility measure. (We also find that the implied volatilities are closely related to the trailing window and GARCH series.) On the other hand, Table 4 shows that, for the non-USV models, these correlations are significantly lower (0.38 for $A_1(4)$) or even negative (−0.5 for $A_1(3)$).

Table 4 also reports correlations between various (model-implied and realized) volatility measures and the actual curvature in yields. In general, curvature is weakly related to volatility estimates in both $A_1(4)$ models, GARCH, and implied volatilities. In contrast, curvature is much more strongly related to volatility estimates in the $A_1(3)$ model.[22]

These results highlight the tension between the dual roles that volatility plays in an unrestricted affine model, as it affects both the cross section of bond prices as well as the time-series properties of the short rate. The estimation of such a model therefore presents a tradeoff between choosing volatility dynamics that are more consistent with either role, and in the present data set it seems that the tradeoff is heavily tilted toward explaining the cross section. The result is that volatilities imputed from the unrestricted $A_1(3)$ models are essentially nonsensical, being unrelated to all other volatility proxies. Instead, the model appears to use the variance process to provide a better fit of the cross section, as evidenced by a relation between V_t and curvature that holds only for this model. Instead, the $A_1(4)$ USV model has enough flexibility to both fit the yield curve and generate realistic volatility dynamics.

We interpret these findings as evidence that three state variables cannot simultaneously describe yield curve level, slope, curvature, and volatility.[23] That is, volatility is unable to play the dual role that the unrestricted $A_1(3)$ model predicts that it does. Less formally, volatility cannot

[22] Note that the sign of this correlation contradicts the theoretical predictions of Litterman, Scheinkman, and Weiss (1991) and Brown and Schaefer (1994a, b). However, as can be seen from Eq. (11), this correlation should depend on how the risk-neutral expected change in slope is correlated with volatility.

[23] This is equivalent to saying that three state variables cannot simultaneously capture volatilities and Q measure drifts. Our result therefore contrasts with Duarte (2004), who finds that three-factor models are unable to simultaneously capture term structure volatilities and P measure drifts (or, equivalently, risk premia).

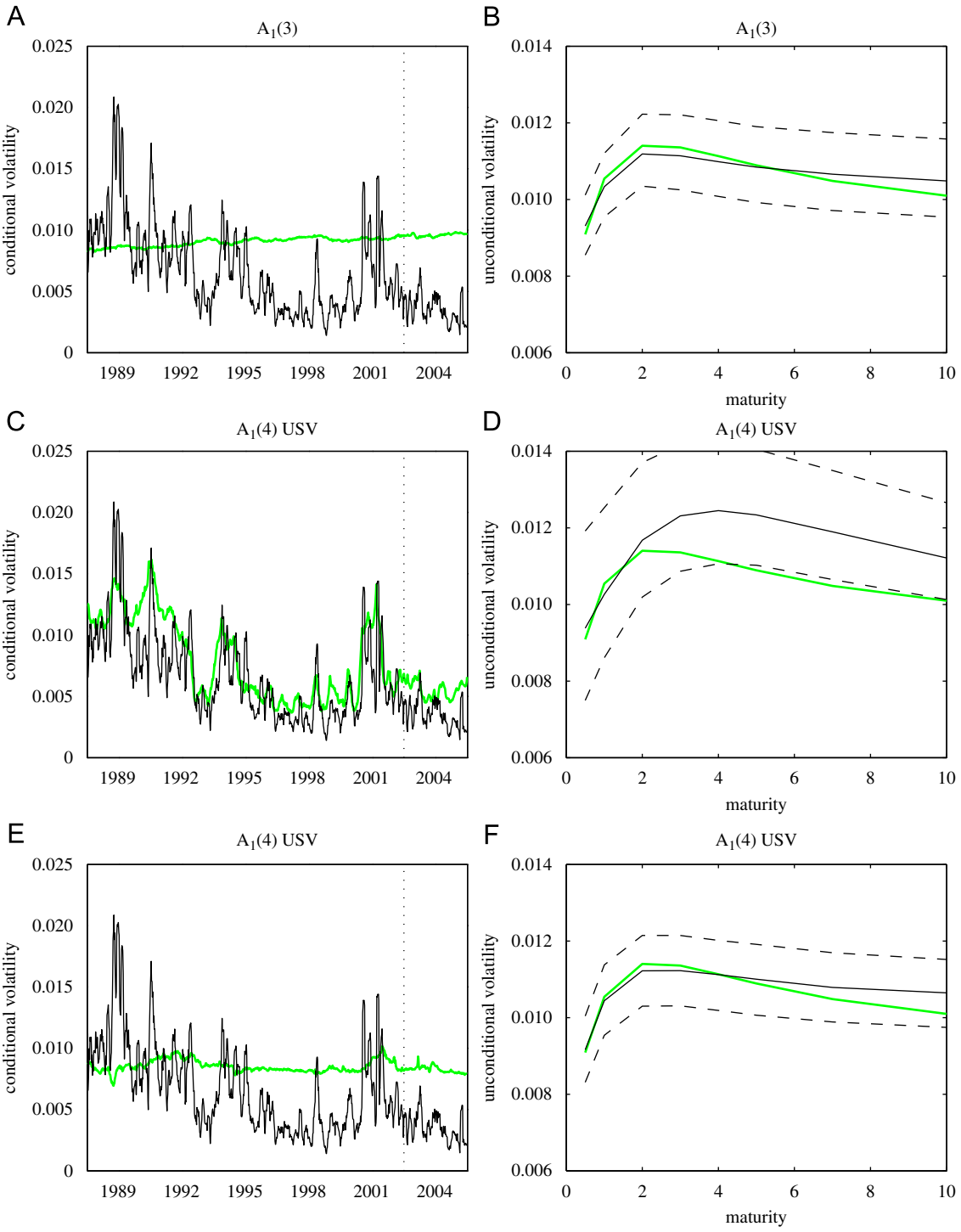

Fig. 1. In the panels on the left-hand side of the figure, the black line depicts the fitted path of the volatility of the six-month yield that constructed from rolling 30-day windows. The gray lines correspond to smoothed estimates of instantaneous volatility implied by each of the affine specifications considered. The vertical dotted lines denote the end of the estimation period. In the panels on the right-hand side of the figure, the thick gray line depicts the sample standard deviation of monthly changes in yields as a function of maturity computed over the 1988–2005 sample period. Distributions of model-implied sample standard deviations were calculated by simulation under the parameter values given in Tables 2A and B. The means and 95% confidence intervals of these distributions are depicted by solid and dashed black lines, respectively.

reasonably be 'inverted' from the yield curve, at least for the models we consider. Conversely, our results suggest that the dynamics of stochastic volatility, as proxied, say, by a GARCH estimate using the implied short rate series, are not able to capture adequately movements in the third principal component of yields.

Interestingly, the unrestricted $A_1(4)$ model, which has enough degrees of freedom to fit the three main principal components as well as volatility, does not perform well at capturing volatility. We interpret this finding as implying that the parameters that match the fourth principal component of yields results in higher values of the likelihood function relative to those that match short rate volatility.

4.4. In- and out-of-sample volatility forecasts

Here we investigate the volatility forecasting performance of the different models. Table 5 reports the models' in- and out-of-sample volatility forecasts. We first discuss methodology, then the results.

4.4.1. Methodology

Because our sample size is relatively short, we focus on short horizon (one-week) forecasts of two different volatility proxies. All forecasts are constructed using the parameter estimates reported in Tables 2A and B, so the bulk of our forecasts are in-sample. After using two years of data to initialize the forecasts, we are left with a 677-week in-sample period. With our hold-out sample from 2003 and 2005, we perform a 156-week out-of-sample validation of those results.

To construct a forecast, we first estimate the value of the current state variables. These are computed identically to the previous section, except that only data observed up until time t are used to infer state variables at t (though for in-sample forecasting the parameter estimates are based on data subsequent to t as well). Given estimates of the current values of the state variables, we simulate 10,000 paths of the model and form a forecast distribution of the state variables one week ahead (time $t+1$), from which we then compute a distribution for each yield.[24]

For volatility forecasting, we consider two alternative proxies for realized volatility. The first proxy is the absolute one-week change in the yield of each maturity. The second proxy is a volatility measure constructed using daily data, which we note were not used for estimation. For a given week with N days (typically, $N=5$), this is defined as

$$\hat{\sigma}_{t,\tau} \equiv \sqrt{\sum_{i=1}^{N} \Delta Y(t,i,\tau)^2}, \tag{29}$$

where $Y(t,i,\tau)$ is the τ-maturity yield observed on the ith day following observation t. The forecast of each volatility proxy is constructed simply by averaging over the Monte Carlo simulations of that proxy. Thus, under the null hypothesis that the model and parameter values are correct, every forecast should be unbiased.

[24] Perhaps because of our short sample, results for forecasting changes in yields were not informative about the relative performance of the three- and four-factor models. In sample, there were no significant differences between models in terms of bias and RMSE across maturities. Therefore, we do not report these results.

Table 5
In-sample and out-of-sample volatility forecasts.

This table contains statistics on in-sample one-week forecasts of two different volatility proxies. For each model, expected absolute yield changes ($E[|\Delta Y|]$) and expected 'realized volatility' ($E[\hat{\sigma}]$) are calculated for 0.5, 1, 2, 3, 4, 5, 7, and 10-year maturities. Realized volatility is defined by $\hat{\sigma}^2_{t,\tau} = \sum_{i=1}^{5} \Delta Y(t,i,\tau)^2$ and is calculated using daily data. The table examines the forecast bias (actual minus forecast) and root mean squared error of $|\Delta Y|$ and $\hat{\sigma}$, where all yields are expressed in basis points. * and ** denote statistical significance at the 5% and 1% levels, where standard errors are calculated using the method of Newey and West (1987) with 12, 17, 6, and 7 lags, respectively, for the four panels of the table. For biases, statistical significance relates to the null hypothesis that the bias is zero. For RMSE, the statistical significance of the pairwise comparison of two models is reported, along with an inequality sign that reflects the direction of the rejection. Forecasts begin in January 1990, allowing for a sample size of 677 weeks for in-sample results and 156 weeks for out-of-sample results.

	$A_1(3)$		$A_1(4)$ USV		$A_1(4)$
			In-sample RMSE of weekly $\|\Delta Y\|$		
6-month	8.82	>**	7.88	<**	8.47
1-year	9.23	>**	8.71	<**	9.04
2-year	9.76	>**	9.45		9.57
3-year	9.71		9.57		9.55
4-year	9.61		9.59		9.49
5-year	9.52		9.55		9.42
7-year	9.38		9.41		9.29
10-year	9.15		9.11		9.06
			In-sample RMSE of $\hat{\sigma}$		
6-month	6.95	>**	5.26	<**	6.29
1-year	6.79	>**	5.87	<**	6.54
2-year	7.19		6.87		6.97
3-year	7.03		7.09		6.86
4-year	6.79		7.05		6.68
5-year	6.60		6.91		6.51
7-year	6.34		6.54		6.24
10-year	6.17		6.12		6.04
			Out-of-sample RMSE of weekly $\|\Delta Y\|$		
6-month	7.17	>**	4.23	<**	5.87
1-year	7.11	>**	5.22	<**	6.37
2-year	8.00	>**	7.18	<*	7.40
3-year	8.45	>**	8.07		8.00
4-year	8.71	>*	8.48	>**	8.37
5-year	8.86		8.67	>*	8.57
7-year	8.88		8.67		8.62
10-year	8.71		8.42		8.42
			Out-of-sample RMSE of $\hat{\sigma}$		
6-month	8.66	>**	4.69	<**	6.95
1-year	7.40	>**	4.54	<**	6.35
2-year	7.01	>**	5.64	<**	6.00
3-year	6.97	>**	6.27		6.18
4-year	7.00	>**	6.53	>**	6.37
5-year	7.02	>**	6.62	>*	6.48
7-year	6.92	>*	6.50		6.42
10-year	6.60	>*	6.08		6.05

4.4.2. Results

In-sample results on forecasted volatility are reported in Table 5. The results show that the $A_1(4)$ USV model

outperforms the unrestricted $A_1(3)$ model in terms of RMSE both in- and out-of-sample. The comparison of the USV model with the unrestricted $A_1(4)$ model is more interesting. It shows that the USV model generally performs better at the short-end (both in- and especially out-of-sample), but at the longer end (for maturities greater than four years) the unrestricted model seems to have a comparable or even superior performance than the USV model. So the unambiguously better fit of short rate volatility dynamics in the USV model does not necessarily translate into better volatility forecasting ability at longer maturities.

Moreover, the unrestricted models do a better job matching unconditional volatilities than does the $A_1(4)$ USV model. The panels on the right-hand side of Fig. 1 display the relation between maturity and the unconditional volatility of four-week yield changes. Results from actual data over the 1988–2005 sample are displayed as a thick gray line. Means and 95% confidence intervals of model-implied sampling distributions are depicted by solid and dashed black lines, respectively. As discussed by Dai and Singleton (2003), the hump-shaped plot of the unconditional volatility as a function of maturity is an empirical regularity that many affine models have difficulty matching. This figure shows that both the unrestricted $A_1(3)$ and $A_1(4)$ models match the unconditional maturity-volatility relation better than the model with USV.[25] The $A_1(4)$ USV model generates a clear hump, but it produces volatilities that are too high for maturities above two years.[26] This indicates that a good fit ofthe short-maturity volatility (which only the USV model provides) is not necessary for obtaining a good unconditional volatility structure.

4.5. Volatility forecasting regressions

To better understand the role of term structure factors and volatility in the various models, we perform an additional forecasting experiment. In particular, we run regressions where we forecast future realized daily volatilities of six-month and ten-year yields. As independent variables we use one of the model forecasts (GARCH, $A_1(3)$, $A_1(4)$, or $A_1(4)$ USV) and three term structure factors (the first three principal components, representing level, slope, and curvature) that capture the shape of the yield curve. The GARCH forecasts of six-month yield volatility is identical to those used in Table 4, while the GARCH forecasts of ten-year yield volatility are based on the identical model fitted to demeaned daily changes in the ten-year yield.

Examining the incremental explanatory power of each model's volatility forecast over that of the term structure shape factors provides a direct test as to whether yield volatility is spanned by those factors, as common models imply. More generally, these regressions should reveal whether it is information from the cross section or the time series that is more useful in predicting volatility.

Table 6 for contains the results of these various forecasting regressions. We note several interesting results:

- Term structure factors (slope and curvature) have some predictive power for future volatility, but more so at the short end (R^2 of 0.15 in regression 1 vs. 0.03 in regression 1′). However, the GARCH(1,1) estimate drives out most of the economic significance of these factors even though slope and curvature remain statistically significant at the short end.
- Of the three models we investigate, only the $A_1(4)$ USV matches the GARCH(1,1) performance at the short end. This is actually quite striking given that the $A_1(4)$ USV model was estimated and its forecasts constructed from weekly data, while the GARCH model has the advantage of being fitted from daily data. Furthermore, the incremental information contained in the cross section of the yield curve, i.e., in the three principal components, is lowest for the $A_1(4)$ USV model (measured by how much the R^2 increases upon including the term structure factors). On the other hand, even for this model the slope coefficient in the predictive regression is just 0.7 and significantly different from one, which indicates that the forecast is not unbiased.[27]
- For longer maturity yield volatilities, the performance of all the forecasts worsens, with the best-performing forecast (GARCH) having an R^2 of only 0.09. Of all the term structure factors, only the slope seems to affect long yield volatilities, but this effect is driven out by including the GARCH or any other forecast.
- For longer maturities, $A_1(4)$ USV is the only affine model whose forecast is reliably positively related to future realized volatility. However, with an R^2 of 0.01, it appears clearly dominated by the GARCH forecast.
- For both long and short maturity volatility forecasts, the unconstrained $A_1(4)$ and $A_1(3)$ volatility forecasts

[25] In fact, the good fit of the $A_1(3)$ model was pointed out by Dai and Singleton (2000) as one reason they advocated this model.

[26] The upward bias in long-maturity volatilities for the $A_1(4)$ USV model seems to be related to our requirement that the Feller condition hold, which we ensured by rejecting all parameter draws that did not satisfy $\gamma_V > \kappa_V \underline{V} + \sigma_V/2$. Thus, imposing this condition clearly increases prior weight on larger values of γ_V, which has the effect of raising the long-run mean of the variance process, thereby raising unconditional volatilities of every maturity. We note that the Feller condition can be relaxed if more restrictive risk premia are assumed for the V process. Furthermore, because the parameters of the risk premia of V are not identifiable from bond prices under USV, more restrictive risk premia would not affect the ability of the two USV specifications to fit bond prices. When we eliminate the requirement that the Feller condition must hold, the $A_1(4)$ USV model produces a volatility hump with confidence intervals that include the observed curve.

[27] We note that there are at least two possibilities for this. First, as shown in Ahn et al. (2003) multiple *volatility* factors might actually be necessary for explaining different dynamics in short and long maturity yield volatilities. Second, the forecasts from the $A_1(4)$ USV model were constructed somewhat sub-optimally. Specifically, we constructed forecasts by conditioning on the point estimates in Tables 2A and B rather than integrating across the entire posterior distribution. Unfortunately, the alternative is computationally unfeasible, and we see no way to measure the potential impact of this simplification.

Table 6
Short and long maturity volatility forecast regressions.

This table contains coefficients and standard errors from regressions of daily realized volatilities ($\tilde{\sigma}$), as defined in Table 5, on different forecasting variables. Newey-West standard errors are calculated using 12 lags. GARCH volatilities in the top panel are the same fitted values used in Table 4, while the bottom panel uses a GARCH(1,1) fitted to demeaned daily 10-year yield changes. Other model forecasts are the same as those used in Table 5, and principal components are identical to those computed in Table 1. Regressions are estimated using data from 1990 to 2002, allowing for a sample size of 677 weeks.

Specification number	Model Name	Intercept*	Model Volatility	1st PC*	2nd PC*	3rd PC*	Adjusted R-squared
			6-month yield volatilities				
1		−0.054		0.839	−0.437	1.228	0.155
		(0.022)		(0.512)	(0.077)	(0.545)	
2	GARCH(1,1)	0.000	0.697				0.246
		(0.000)	(0.077)				
3	GARCH(1,1)	−0.036	0.565	0.227	−0.175	0.770	0.268
		(0.012)	(0.084)	(0.327)	(0.065)	(0.308)	
4	$A_1(3)$	0.007	−5.046				0.135
		(0.001)	(0.897)				
5	$A_1(3)$	7.695	−53.503	−14.607	2.986	−4.242	0.157
		(6.413)	(44.276)	(12.814)	(2.849)	(4.540)	
6	$A_1(4)$ USV	0.000	0.698				0.238
		(0.000)	(0.073)				
7	$A_1(4)$ USV	−0.007	0.672	−0.525	−0.032	0.809	0.245
		(0.016)	(0.103)	(0.466)	(0.094)	(0.391)	
8	$A_1(4)$	−0.002	2.472				0.089
		(0.001)	(0.538)				
9	$A_1(4)$	−0.218	1.695	0.491	−0.204	1.149	0.164
		(0.110)	(1.123)	(0.699)	(0.187)	(0.531)	
			10-year yield volatilities				
1′		0.128		−0.208	−0.215	0.225	0.031
		(0.020)		(0.444)	(0.086)	(0.411)	
2′	GARCH(1,1)	0.000	0.777				0.090
		(0.000)	(0.088)				
3′	GARCH(1,1)	0.023	0.707	−0.317	−0.088	0.397	0.093
		(0.018)	(0.089)	(0.298)	(0.058)	(0.272)	
4′	$A_1(3)$	0.001	−0.057				−0.001
		(0.000)	(0.323)				
5′	$A_1(3)$	0.606	−2.332	−1.943	0.172	−0.420	0.030
		(1.793)	(8.733)	(6.544)	(1.459)	(2.486)	
6′	$A_1(4)$ USV	0.001	0.473				0.010
		(0.000)	(0.246)				
7′	$A_1(4)$ USV	0.074	0.667	−0.697	−0.075	0.071	0.040
		(0.028)	(0.347)	(0.491)	(0.112)	(0.407)	
8′	$A_1(4)$	−0.003	3.155				0.013
		(0.002)	(1.631)				
9′	$A_1(4)$	0.401	−2.038	−0.060	−0.313	0.255	0.032
		(0.454)	(3.454)	(0.508)	(0.193)	(0.434)	

*denotes a coefficient that has been multiplied by 100

are driven out by the slope and curvature factors. In other words, we cannot reject that the volatility forecasts of these unconstrained models are spanned by yield curve factors. Further, their forecasts are generally less accurate than those based on the GARCH or $A_1(4)$ USV models, both of which have volatilities that are not spanned by the term structure factors.

In conclusion, these results show that while cross-sectional term structure factors have some explanatory power for forecasting future yield volatility, the bulk of the explanatory power is provided by time-series forecasts. The latter is not subsumed by the cross-sectional factors.[28] Lastly, long and short end volatilities seem to behave quite differently.

4.6. Some robustness results

4.6.1. Risk premia

As we discussed previously, a number of the risk premia parameters are not significant. Because of concerns about overparameterization, we repeated some of our in/out-of-sample yield fit and volatility forecasting

[28] We note that if we include six principal components instead of three, or use raw yields instead of PCs, the same result obtains.

exercises after re-estimating the models with all risk premium coefficients set to zero. The results show that the particular specification of risk premium structure is of second-order importance when it comes to the performance of the model at fitting yields or forecasting volatility. If anything, the out-of-sample performance of the models is slightly enhanced when the risk premia are more parsimonious, though these differences are too small to be driving any of our conclusions.

This result seems to arise from the relative invariance of the parameters that affect the risk-neutral distribution to the specification of risk premia. Since these parameters are almost solely responsible for determining yield curve fit, those results are unaffected by changing risk premia. Our volatility forecasts are unaffected, for the most part, because they are short-term in nature, and are therefore driven more by the ability of the model to fit current volatilities rather than predict whether these volatilities will drift up or down over time. Of course, none of this suggests that the specification of risk premia is not important, and with a longer sample we would expect differences in specification to be of primary importance for forecasting yield levels or changes, particularly over longer horizons, as argued by Duffee (2002).

4.6.2. *Curvature-volatility relation for Treasury yields*

Given our use of swap and LIBOR rates throughout the paper, it is possible that the small but significant level of default risk that is priced into those rates might be driving these results. Alternatively, these results might be attributed to our bootstrapping method or a sample size that is not as long (1988–2002) as some used elsewhere in the literature. To alleviate these concerns, we replicate some of the results (available upon request) with two alternative data sets constructed from Treasury yields. The first is constructed from all available Treasury bills and strips. These data are available daily from 1990–2002, and we form constant maturity yields by linear interpolation of all available yields. On each day we compute yields with maturities identical to those used previously, except that we replace the six-month rate with a three-month rate because of better data availability. The second data set results from merging the monthly Fama-Bliss data, with evenly-spaced maturities between one and five years, with a daily three-month Treasury bill rate from the Federal Reserve. We collect these data for the 1960–2002 period.

In general, the patterns that emerge from these results are the same as those observed for our LIBOR/swap sample. One exception is that over the 1960–2002 sample the first principal component appears to have a very significant role in predicting future volatility. This appears to be mostly due to the inclusion of the 1979–1982 period, which some authors consider a period of structural change in monetary policy (e.g., Huizinga and Mishkin, 1984; Campbell, 1987). It is well-known from Chan et al. (1992) that this period included historical highs in both the level of nominal yields and their volatility, and it is for this reason we find the first principal component to be a useful predictor of volatility. Noticeably, the curvature of the yield curve, which both Litterman, Scheinkman, and Weiss (1991) and Brown and Schaefer (1994a, b) argue should be most related to volatility, does not seem important at all during this period. Nevertheless, the more important result from these data is that volatility continues to be much more accurately predicted by the GARCH measure. Controlling for the shape of the yield curve, the incremental benefit of adding GARCH volatilities is large and highly significant. The effect of adding PCs to a regression that already includes a GARCH volatility is, on the other hand, very minor.

These results confirm that time series-based volatility proxies contain the vast majority of information relevant for predicting future volatilities and that the information contained in the yield curve alone is insufficient for producing accurate forecasts. Since this result obtains across sample periods and several data sets, it seems therefore unambiguous that short rate volatility is not 'spanned' by the yield curve.

5. Conclusion

We investigate several affine models of the term structure that generate stochastic volatility in yields. We find that the unrestricted $A_1(3)$ model implies a volatility time series that is essentially unrelated to the actual volatility of the short rate process. This surprising result is a consequence of the dual role played by the volatility state variable in the unrestricted affine model: it is both a linear combination of yields (i.e., it affects the cross section of the term structure) and the quadratic variation of the short rate (i.e., it impacts the time series of the term structure). Bayesian estimation results in more weight being placed on the first role at the expense of the second. While the in-sample fit of yields is excellent, a clear out-of-sample breakdown casts doubt on the model's adequacy in this role as well.

We then investigate an $A_1(4)$ model exhibiting USV. The USV specification allows the model to fit level, slope, and curvature while simultaneously producing a volatility process that is highly correlated with both GARCH and option-implied volatility series. It does so by explicitly introducing variation in curvature that is unrelated to volatility, a straightforward generalization within the representations introduced by CGJ. This model is also (mostly) successful in replicating the 'hump shape' in term structure volatility, and it performs as well out-of-sample as it does in-sample.

We also consider an unrestricted $A_1(4)$ model. In principle, this model could fit level, slope, curvature, as well as volatility perfectly. However, results indicate that the likelihood criterion is strongly tilted toward fitting the cross section of yields at the expense of the predicted volatility series, which, even in this unconstrained four-factor model, bears little resemblance to actual volatility time series estimated via GARCH or from implied option volatilities. However, the model performs well in fitting the unconditional volatility structure and longer-maturity yield conditional volatilities. We conclude that fitting short maturity volatilities is not necessary to obtain a good fit of yields and longer maturity conditional

volatilities, as short- and long-dated volatilities appear to be driven by different components.

While our results confirm the findings of Litterman and Scheinkman (1991) that at least three factors are needed to explain the cross sectional features of the yield curve, it further demonstrates that these factors are an inadequate description of the state space, as they are incapable of replicating observed patterns of conditional volatility.

Appendix A. SDE representations

There are several ways to obtain an SDE representation of the form (1)–(3) from the mean/covariance forms used in the paper. For the risk-neutral drift, if

$$\frac{1}{dt}\mathrm{E}^{\mathrm{Q}}[dX_t] = a^{\mathrm{Q}} + b^{\mathrm{Q}}X_t, \tag{A.1}$$

then simply set

$$\mathscr{K}^{\mathrm{Q}} = -b^{\mathrm{Q}} \quad \text{and} \quad \Theta^{\mathrm{Q}} = -(b^{\mathrm{Q}})^{-1}a^{\mathrm{Q}}. \tag{A.2}$$

Define the P drift parameters similarly.

Given a covariance matrix of the form

$$\frac{1}{dt}\mathrm{Cov}(dX_t, dX_t^{\top}) = \Omega_0 + \Omega_V(V_t - \underline{V}) \tag{A.3}$$

that is at least positive semidefinite for every V, it is always possible to find matrices $\Omega_0^{1/2}$ and $\Omega_V^{1/2}$ such that $\Omega_0^{1/2}\Omega_0^{1/2\top} = \Omega_0$ and $\Omega_V^{1/2}\Omega_V^{1/2\top} = \Omega_V$. When Ω_V is positive definite, $\Omega_V^{1/2}$ can be chosen to be its Cholesky decomposition.

Now assume that V_t is the last element of X_t and let N denote the total number of state variables that comprise X. Then define

$$\Sigma = \begin{bmatrix} \Omega_0^{1/2} \\ \Omega_V^{1/2} \end{bmatrix}, \quad \alpha = \begin{bmatrix} 1_N \\ 0_N - \underline{V} \end{bmatrix}, \quad \text{and} \quad \beta = \begin{bmatrix} 0_N & 0_N & 0_N \\ 0_N & 0_N & 1_N \end{bmatrix}, \tag{A.4}$$

where 1_N is a $N \times 1$ vector of ones and 0_N is a vector of zeros. It is easy to verify that these choices, within the specification (1)–(3), imply the same covariance matrix as (A.3).

These manipulations put the model into standard affine form, but since Σ has $2N$ rows the model is written with twice as many Brownian motions as state variables. To write the model with as many Brownian motions as state variables, it is always possible to write the model in SDE form as

$$dX_t = \mathscr{K}^{\mathrm{Q}}(\Theta^{\mathrm{Q}} - X_t)\,dt + \Omega_t^{1/2}\,dZ_t^{\mathrm{Q}}, \tag{A.5}$$

where $\Omega_t^{1/2}$ is some matrix (e.g., the Cholesky decomposition) such that $\Omega_t^{1/2}\Omega_t^{1/2\top} = \Omega_t$. However, as CGJ demonstrate, it is *not* always the case that a formulation of this type is of the standard affine form given in (1)–(3). Thus, it is only guaranteed that a fully general N-factor model may be written in standard affine form if we allow the number of Brownian motions to exceed the number of state variables.

Appendix B. Proof of Proposition 1

We claim that bond prices are of the form (22), with coefficients given by

$$B_r(\tau) = \frac{e^{c_{r\mu}\tau}(6c_{r\mu} - 2a_\theta)}{4c_{r\mu}^2 - c_{r\mu}a_\theta} + \frac{e^{2c_{r\mu}\tau}(3c_{r\mu} - a_\theta)}{-10c_{r\mu}^2 + 2c_{r\mu}a_\theta} + \frac{7c_{r\mu} - 3a_\theta}{-6c_{r\mu}^2 + 2c_{r\mu}a_\theta} - \frac{2c_{r\mu}^2 e^{(-3c_{r\mu}+a_\theta)\tau}}{\Gamma},$$

$$B_\mu(\tau) = \frac{a_\theta}{2c_{r\mu}^2(-3c_{r\mu} + a_\theta)} + \frac{e^{2c_{r\mu}\tau}(2c_{r\mu} - a_\theta)}{10c_{r\mu}^3 - 2c_{r\mu}^2 a_\theta} + \frac{e^{c_{r\mu}\tau}(c_{r\mu} - a_\theta)}{-4c_{r\mu}^3 + c_{r\mu}^2 a_\theta} + \frac{3c_{r\mu}e^{(-3c_{r\mu}+a_\theta)\tau}}{\Gamma},$$

$$B_\theta(\tau) = \frac{e^{c_{r\mu}\tau}}{c_{r\mu}^2(-4c_{r\mu} + a_\theta)} + \frac{1}{6c_{r\mu}^3 - 2c_{r\mu}^2 a_\theta} + \frac{e^{2c_{r\mu}\tau}}{10c_{r\mu}^3 - 2c_{r\mu}^2 a_\theta} - \frac{e^{(-3c_{r\mu}+a_\theta)\tau}}{\Gamma}$$

$$A(\tau) = \int_0^\tau \Big[\frac{1}{2}B_\mu^2(\underline{\sigma}_\mu - \sigma_\mu \underline{V}) + \frac{1}{2}B_\theta^2(\underline{\sigma}_\theta - \sigma_\theta \underline{V}) + B_r B_\mu(\underline{c}_{r\mu} - c_{r\mu}\underline{V}) + B_r B_\theta(\underline{c}_{r\theta} - c_{r\theta}\underline{V}) + B_\mu B_\theta(\underline{c}_{\mu\theta} - c_{\mu\theta}\underline{V}) - a_0 B_\theta\Big], \tag{B.1}$$

where

$$\Gamma = (3c_{r\mu} - a_\theta)(4c_{r\mu} - a_\theta)(5c_{r\mu} - a_\theta). \tag{B.2}$$

We note that the integral in (B.1) has an analytic expression, but to simplify notation we leave it in integral form.

To prove this claim it is sufficient to show that $e^{-\int_0^t r_s\,ds}P(t,T)$ is a Q-martingale for P as defined in Eq. (22). Indeed, in that case we have $e^{-\int_0^t r_s ds}P(t,T) = \mathrm{E}_t^Q[e^{-\int_0^T r_s ds}P(T,T)]$, which implies

$$P(t,T) = \mathrm{E}_t^Q[e^{-\int_0^T r_s\,ds}],$$

since Eq. (22)–(B.1) imply $P(T,T) = 1$. To show that $e^{-\int_0^t r_s\,ds}P(t,T)$ is a Q-martingale, we apply Itô's lemma to Eq. (22). Using the fact that the functions $A(\cdot)$, $B_r(\cdot)$, and $B_\mu(\cdot)$ satisfy the system of ODE:

$$B_r' = a_r B_\theta + 1, \tag{B.3}$$

$$B_\mu' = B_r + a_\mu B_\theta, \tag{B.4}$$

$$B_\theta' = B_\mu + a_\theta B_\theta, \tag{B.5}$$

$$A' = \tfrac{1}{2}B_\mu^2(\underline{\sigma}_\mu - \sigma_\mu \underline{V}) + \tfrac{1}{2}B_\theta^2(\underline{\sigma}_\theta - \sigma_\theta \underline{V}) + B_r B_\mu(\underline{c}_{r\mu} - c_{r\mu}\underline{V}) + B_r B_\theta(\underline{c}_{r\theta} - c_{r\theta}\underline{V}) + B_\mu B_\theta(\underline{c}_{\mu\theta} - c_{\mu\theta}\underline{V}) - a_0 B_\theta \tag{B.6}$$

$$0 = -3c_{r\mu}B_\theta + \tfrac{1}{2}B_r^2 + \tfrac{1}{2}B_\mu^2 c_{r\mu}^2 + \tfrac{1}{2}B_\theta^2 c_{r\mu}^4 + B_r B_\mu c_{r\mu} + B_r B_\theta c_{r\mu}^2 + B_\mu B_\theta c_{r\mu}^3, \tag{B.7}$$

and that, in particular, because of the following restrictions on a_r and a_μ:

$$a_r = -2c_{r\mu}^2(3c_{r\mu} - a_\theta),$$
$$a_\mu = 7c_{r\mu}^2 - 3c_{r\mu}a_\theta,$$

we have

$$B_r = -c_{r\mu}(B_\mu + c_{r\mu}B_\theta) + \sqrt{2B_\mu + 6c_{r\mu}B_\theta}. \tag{B.8}$$

We therefore find that

$$\mathrm{E}^Q[dP(t,T) - r_t P(t,T)] = 0. \tag{B.9}$$

It then follows straightforwardly that $e^{-\int_0^t r_s\,ds}P(t,T)$ is indeed a Q-martingale.

Appendix C. Details of the MCMC procedure

Our MCMC algorithm alternates between drawing different blocks of unobservable state variables and model parameters. To do so, it is convenient to define $U_t = V_t - \underline{V}$, which is the deviation of V_t from its lower bound. In addition, we parameterize the Ω_0 and Ω_V matrices in terms of their lower diagonal decompositions $\Omega_0^{1/2}$ and $\Omega_V^{1/2}$, though we report the parameters of the original matrices. In most cases, these matrices are not full rank, but the decompositions are easily derived, often as the Cholesky decompositions of the nonzero part of the matrix.

We decompose the parameter vector into three blocks, ϕ^λ, ϕ^Λ, and ϕ^Q, where ϕ^λ includes all risk premia parameters, ϕ^Λ includes measurement error standard deviations, and ϕ^Q includes all parameters that drive factor dynamics under the Q measure. The block for ϕ^Q will also include the draw of X^o, the state variables other than U. Finally, the draws of U_t are performed, as in Jones (2003b) or Jacquier, Polson, and Rossi (1994), separately for each $t \in \{1, 1+h, 1+2h, \ldots, T\}$, a set of $(T-1)/h+1$ blocks. Thus, there are a total of $(T-1)/h+4$ separate blocks:

- $p(U_t|\mathscr{P}, U_1, \ldots, U_{t-h}, U_{t+h}, \ldots, U_T, X^o, \phi)$ for each $t \in \{1, 1+h, 1+2h, \ldots, T\}$.
- $p(\phi^\Lambda|\mathscr{P}, U, X^o, \phi^\lambda, \phi^Q)$.
- $p(\phi^\lambda|\mathscr{P}, U, X^o, \phi^\Lambda, \phi^Q)$.
- $p(\phi^Q, X^o|\mathscr{P}, U, \phi^\Lambda, \phi^\lambda)$.

Draws are rejected if they violate stationarity, admissibility, or Feller conditions, or if they imply near-degenerate covariance matrices. This involves restrictions on parameters that keep variances bounded away from zero. In practice, none of these restrictions (except for the Feller condition) appears to bind, but they nevertheless serve to eliminate the possibility of improper posteriors, discussed by Hobert and Casella (1996) and Johannes and Polson (2009).

C.1. The linear state space representation

In all of the $A_1(N)$ the models we consider, only U_t (or V_t) enters the conditional variance of the state vector X_{t+h}. Conditional on the entire path of U, the remaining state variables X^o_{t+h} have a mean that is linear in X^o_t and a variance that does not depend on X^o_t. At the same time, the observed data also have a mean that is linear in the unobserved vector X^o_t, with a covariance matrix Λ that we have assumed constant. We can obtain this relation by rewriting (26) as

$$\mathscr{P}_t - L^v(U_t + \underline{V}) = K + L^o X^o_t + e_t, \tag{C.1}$$

where L^v denotes the last column of L and L^o the remaining columns.

Thus, conditional on the path of U, we have state and measurement equations that are Gaussian and linear in the state variable X^o_t. This enables the use of the standard Kalman filter to compute $p(\mathscr{P}|\phi, U)$. One small complication is that our state equation defines transitions over a unit of time of length h, while the measurement equation is only applicable for observation times $t \in \{1, 2, \ldots, T\}$. To resolve this asymmetry, consider the equivalent situation where yields were instead observed at every length-h interval, but that the measurement error variance for non-integer t was infinitely large. For non-integer t, the Kalman 'gain' matrix is then zero, meaning that the observed data have no effect on the conditional distribution of the state vector. Thus, we can apply the Kalman filter in its textbook form simply by zeroing out the Kalman gain matrix when $t \notin \{1, 2, \ldots, T\}$.

C.2. Drawing U_t

For $t \in \{1, 2, \ldots, T\}$, the target density can be decomposed using the model's Markov structure as

$$p(U_t|\mathscr{P}, U_1, \ldots, U_{t-h}, U_{t+h}, \ldots, U_T, X^o, \phi)$$
$$\propto p(U_{t+h}, X^o_{t+h}|U_t, X^o_t, \phi)p(U_t|\mathscr{P}_t, U_{t-h}, X^o_{t-h}, X^o_t, \phi),$$

which follows from the Markov property and Bayes rule. Note that $p(U_t|\mathscr{P}_t, U_{t-h}, X^o_{t-h}, X^o_t, \phi)$ is a Gaussian density for U_t. This is because it is a conditional density derived from the multivariate Gaussian density of $\{U_t, X^o_t, \mathscr{P}_t\}$ given $\{U_{t-h}, X^o_{t-h}\}$. We use this density as the candidate generator for a Metropolis-Hastings draw, forming the acceptance factor from the omitted component of the target density. We therefore replace the current draw U_t with the candidate U^*_t with probability

$$\min\left\{\frac{p(U_{t+h}, X^o_{t+h}|U^*_t, X^o_t, \phi)}{p(U_{t+h}, X^o_{t+h}|U_t, X^o_t, \phi)}, 1\right\},$$

also rejecting all draws of negative U_t.

For $t \notin \{1, 2, \ldots, T\}$, the draw is similar, with the target density instead proportional to

$$p(U_{t+h}, X^o_{t+h}|U_t, X^o_t, \phi)p(U_t|U_{t-h}, X^o_{t-h}, X^o_t, \phi).$$

The second component is a slightly different Gaussian candidate generating density for U_t, and the Metropolis-Hastings acceptance probability is unchanged from before.

C.3. Drawing ϕ^{Λ}

Given ϕ^{Q}, X^{o}, and U, we may compute fitted principal components as $K+LX_t$ and construct a time series of measurement errors $e_t=\mathscr{P}_t-K-LX_t$, where it was assumed that $e_t\sim N(0,\Lambda)$. Since Λ was assumed diagonal (measurement errors are cross-sectionally uncorrelated), we may consider the error for each principal component separately. With a flat prior on each measurement error standard deviation (i.e., $p(\sqrt{\Lambda_{i,i}})\propto 1/\sqrt{\Lambda_{i,i}}$), we have the standard result that $\sqrt{\Lambda_{i,i}}$ has an inverted gamma distribution with T degrees of freedom and a location parameter equal to the root mean squared measurement error of the ith principal component.

C.4. Drawing ϕ^{λ}

To draw ϕ^{λ} we write the Euler approximation (25) as

$$X_{t+h}-X_t=h(a+bX_t)+\sqrt{h}\Omega_t^{1/2}\varepsilon_{t+h}.$$

Since the drift is linear and the residual covariance matrix is known (as a function of U and ϕ^{Q}), we can directly apply the seeming unrelated regression approach of Chib and Greenberg (1996) to draw the a vector and the b matrix. Since we are using the 'generalized essentially affine' risk premia of Cheridito, Filipovic, and Kimmel (2007), each nonzero element of a and b is effectively a free parameter (subject to stationarity and admissibility conditions). Thus, no linkage between the P and Q drift parameters need be imposed and we can simply back out risk premia according to

$$\lambda_0+\lambda_X X_t=a+bX_t-a^{Q}-b^{Q}X_t,$$

where a^{Q} and b^{Q} are Q measure parameters analogous to a and b.

C.5. Drawing ϕ^{Q} and X^{o}

In this block we seek a draw from $p(\phi^{Q},X^{o}|\mathscr{P},U,\phi^{\Lambda},\phi^{\lambda})$, which we decompose as

$$p(\phi^{Q}|\mathscr{P},U,\phi^{\Lambda},\phi^{\lambda})p(X^{o}|\mathscr{P},U,\phi^{Q},\phi^{\Lambda},\phi^{\lambda}).$$

Because our prior on ϕ^{Q} is completely flat and independent of ϕ^{Λ} and ϕ^{λ}, we have

$$\begin{aligned}p(\phi^{Q}|\mathscr{P},U,\phi^{\Lambda},\phi^{\lambda})&\propto p(\mathscr{P},U|\phi^{Q},\phi^{\Lambda},\phi^{\lambda})\\&\propto p(\mathscr{P}|U,\phi^{Q},\phi^{\Lambda},\phi^{\lambda})p(U|\phi^{Q},\phi^{\Lambda},\phi^{\lambda}).\end{aligned}$$

The second term, $p(U|\phi^{Q},\phi^{\Lambda},\phi^{\lambda})$, is easily evaluated because U is a univariate Markov process whose dynamics are fully described by the Euler approximation. The first term, $p(\mathscr{P}|U,\phi^{Q},\phi^{\Lambda},\phi^{\lambda})$, is evaluated using the Kalman filter. As noted above, once we condition on the entire path of U, we may write the dynamics of X_t^{o} and $\mathscr{P}_t$ in linear Gaussian state space form.

We use a random walk Metropolis chain to draw a candidate value ϕ^{Q*} for replacing the current value ϕ^{Q}. The acceptance factor, in this case, is just the ratio of the target densities, so that we accept ϕ^{Q*} over ϕ^{Q} with probability

$$\min\left\{\frac{p(\mathscr{P}|U,\phi^{Q*},\phi^{\Lambda},\phi^{\lambda})p(U|\phi^{Q*},\phi^{\Lambda},\phi^{\lambda})}{p(\mathscr{P}|U,\phi^{Q},\phi^{\Lambda},\phi^{\lambda})p(U|\phi^{Q},\phi^{\Lambda},\phi^{\lambda})},1\right\},$$

which we are now able to compute. Given the resulting draw of ϕ^{Q}, we may invoke the simulation smoother of de Jong and Shephard (1995) to draw the entire multivariate time series X^{o} all at once from the density $p(X^{o}|\mathscr{P},U,\phi)$.

Following Bester (2004), we alternate between usually drawing the entire ϕ^{Q} vector at once using a multivariate candidate generator and occasionally (once every 10 iterations) drawing each element of ϕ^{Q} individually. The covariance matrix of the candidate generator is chosen by running a long preliminary chain and computing the sample covariance matrix of the draws of ϕ^{Q} from the chain. A second chain is run in which a scaling parameter is chosen adaptively to set the Metropolis acceptance rate approximately equal to 0.4. A final third chain is run to generate the posteriors reported.

C.6. Sensitivity to the choice of h

The results of Jones (2003b) and Eraker (2001) suggest that even nonlinear term structure models do not suffer from appreciable discretization bias when the discretization interval is set equal to one day. Given our use of weekly data, this suggests that at a minimum we should investigate values of h as small as 0.2 (one fifth of the weekly observation interval). In this section we compare some results presented in the paper, which were computed using $h=1$, to the case in which $h=0.2$. To reduce computation time, we calculated results only for the $A_1(3)$ USV model.

Parameter posterior distributions from the two cases turn out to be extremely similar, with differences only noticeable for the parameters that drive the V process (γ_V^{p}, κ_V^{p}, and σ_V). Even these differences are still small relative to the posterior standard deviations. Furthermore, these differences are not large enough to generate any observable differences in the estimated r and μ^{Q} state variables, and only extremely minor differences are observed for estimated short rate variances. Overall, differences are very small and appear clearly insufficient to change any of the conclusions of the paper.

References

Ahn, D., Dittmar, R., Gallant, A., Gao, B., 2003. Purebred or hybrid?: reproducing the volatility in term structure dynamics. Journal of Econometrics 116, 147–180.

Almeida, C., Graveline, J., Joslin, S., 2006. Do options contain information about excess bond returns? Working paper, University of Minnesota and MIT.

Bester, C., 2004. Random field and affine models for interest rates: an empirical comparison. Working paper, University of Chicago.

Bikbov, R., Chernov, M., 2009. Unspanned stochastic volatility in affine models: evidence from Eurodollar futures and options. Management Science, forthcoming.

Black, F., 1976. The pricing of commodity contracts. Journal of Financial Economics 3, 167–179.

Bliss, R., 1997. Testing term structure estimation methods. Advances in Futures and Options Research 9, 197–231.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Brandt, M., He, P., 2006. Simulated maximum likelihood estimation of affine models. Working paper, Duke University.
Brenner, R.J., Harjes, R.H., Kroner, K.F., 1996. Another look at models of the short-term interest rate. Journal of Financial and Quantitative Analysis 31, 85–107.
Brown, R., Schaefer, S., 1994a. Interest rate volatility and the shape of the term structure. Philosophical Transactions of the Royal Society: Physical Sciences and Engineering 347, 449–598.
Brown, R., Schaefer, S., 1994b. The term structure of real interest rates and the CIR theory of the term structure of interest rates. Journal of Financial Economics 35, 3–42.
Campbell, J.Y., 1987. Stock returns and the term structure. Journal of Financial Economics 18, 373–399.
Chan, K., Karolyi, G., Longstaff, F., Sanders, A., 1992. An empirical comparison of alternative models of the short-term interest rate. Journal of Finance 47, 1209–1227.
Chen, R., Scott, L., 1993. Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates. Journal of Fixed Income 3, 14–31.
Cheridito, P., Filipovic, D., Kimmel, R., 2007. Market price of risk specifications for affine models: theory and evidence. Journal of Financial Economics 83, 123–170.
Chib, S., Greenberg, E., 1996. Markov chain Monte Carlo simulation methods in econometrics. Econometric Theory 12, 409–431.
Cochrane, J., Piazzesi, M., 2005. Bond risk premia. American Economic Review 95 (1), 138–160.
Collin-Dufresne, P., Goldstein, R., 2002. Do bonds span the fixed income markets? Theory and evidence for unspanned stochastic volatility. Journal of Finance 57, 1685–1730.
Collin-Dufresne, P., Goldstein, R., Jones, C., 2008. Identification of maximal affine term structure models. Journal of Finance 63, 743–795.
Collin-Dufresne, P., Solnik, B., 2001. On the term structure of default premia in the swap and LIBOR markets. Journal of Finance 56, 1095–1115.
Cox, J., Ingersoll Jr., J., Ross, S., 1981. A reexamination of the traditional hypotheses about the term structure of interest rates. Journal of Finance 36, 769–799.
Cox, J., Ingersoll Jr., J., Ross, S., 1985. A theory of the term structure of interest rates. Econometrica 53, 385–407.
Dai, Q., Singleton, K., 2000. Specification analysis of affine term structure models. Journal of Finance 55, 1943–1978.
Dai, Q., Singleton, K., 2002. Expectations puzzle, time-varying risk premia, and affine models of the term structure. Journal of Financial Economics 63, 415–441.
Dai, Q., Singleton, K., 2003. Term structure dynamics in theory and reality. The Review of Financial Studies 16, 631–678.
de Jong, F., 2000. Time series and cross section information in affine term structure models. Journal of Business and Economic Statistics 18, 300–314.
de Jong, P., Shephard, N., 1995. The simulation smoother for time series models. Biometrika 82, 339–350.
Diebold, F., Mariano, R., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–263.
Duan, J.-C., Simonato, J.-G., 1999. Estimating and testing exponential-affine term structure models by Kalman filter. Review of Quantitative Finance and Accounting 13, 111–135.
Duarte, J., 2004. Evaluating an alternative risk preference in affine term structure models. The Review of Financial Studies 17, 370–404.
Duffee, G., 2002. Term premia and interest rate forecasts in affine models. Journal of Finance 57, 405–443.
Duffee, G., Stanton, R., 2002. Estimation of dynamic term structure models. Working paper, UC Berkeley.
Duffie, D., Kan, R., 1996. A yield-factor model of interest rates. Mathematical Finance 6, 379–406.
Duffie, D., Singleton, K., 1997. An econometric model of the term structure of interest rate swap yields. Journal of Finance 52, 1287–1381.
Elerian, O., Chib, S., Shephard, N., 2001. Likelihood inference for discretely observed nonlinear diffusions. Econometrica 69 (4), 959–993.
Engle, R., Lilien, D., Robins, R., 1987. Estimating time varying risk premia in the term structure: the ARCH-M model. Econometrica 55, 391–407.
Eraker, B., 2001. MCMC analysis of diffusion models with application to finance. Journal of Business and Economic Statistics 19 (2), 177–191.
Fama, E., 1976. Inflation uncertainty and expected returns on Treasury bills. Journal of Political Economy 84, 427–448.
Fan, R., Gupta, A., Ritchken, P., 2003. Hedging in the possible presence of unspanned stochastic volatility: evidence from swaption markets. Journal of Finance 58, 2219–2248.
Han, B., 2007. Stochastic volatilities and correlations of bond yields. JF 62, 1491–1524.
Harrison, M., Kreps, D., 1979. Martingales and multiperiod securities markets. Journal of Economic Theory 20, 381–408.
Harrison, M., Pliska, S., 1981. Martingales and stochastic integrals in the theory of continuous trading. Stochastic Processes and their Applications 11, 313–316.
Heidari, M., Wu, L., 2003. Are interest rate derivatives spanned by the term structure of interest rates? Journal of Fixed Income 13, 75–86.
Hobert, J., Casella, G., 1996. The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. Journal of the American Statistical Association 91, 1461–1473.
Huizinga, J., Mishkin, F., 1984. Inflation and real interest rates on assets with different risk characteristics. Journal of Finance 39, 699–712.
Jacquier, E., Polson, N., Rossi, P., 1994. Bayesian analysis of stochastic volatility models. Journal of Business and Economic Statistics 12, 413–417.
Jagannathan, R., Kaplin, A., Sun, S., 2003. An evaluation of multi-factor CIR models using LIBOR, swap rates and cap and swaption prices. Journal of Econometrics 116, 113–146.
Johannes, M., Polson, N., 2009. MCMC methods for continuous-time financial econometrics. In: Ait-Sahalia, Y., Hansen, L. (Eds.), Handbook of Financial Econometrics. Elsevier forthcoming.
Johannes, M., Sundaresan, S., 2006. The impact of collateralization on swap rates. Journal of Finance 62, 383–410.
Jones, C., 2003a. Bootstrapping zero coupon yields from asynchronous LIBOR and swap quotes. Working paper, University of Southern California.
Jones, C., 2003b. Nonlinear mean reversion in the short-term interest rate. The Review of Financial Studies 16, 793–843.
Joslin, S., 2007. Pricing and hedging volatility risk in fixed income markets. Working paper, Stanford University.
Lamoureux, C., Witte, H., 2002. Empirical analysis of the yield curve: the information in the data viewed through the window of Cox, Ingersoll, and Ross. Journal of Finance 57, 1479–1520.
Li, H., Zhao, F., 2006. Unspanned stochastic volatility: evidence from hedging interest rate derivatives. Journal of Finance 61, 341–378.
Liptser, R., Shiryayev, A., 1974. Statistics of Random Processes I, II. Springer, Berlin.
Litterman, R., Scheinkman, J., 1991. Common factors affecting bond returns. Journal of Fixed Income 1, 54–61.
Litterman, R., Scheinkman, J., Weiss, L., 1991. Volatility and the yield curve. Journal of Fixed Income 1, 49–53.
Lund, J., 1997. Econometric analysis of continuous-time arbitrage-free models of the term structure of interest rates. Working paper.
Newey, W., West, K., 1987. A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.
Newey, W., West, K., 1994. Automatic lag selection in covariance matrix estimation. The Review of Economic Studies 61, 631–653.
Pearson, N., Sun, T., 1994. Exploiting the conditional density in estimating the term structure: an application to the Cox, Ingersoll, and Ross model. Journal of Finance 49, 1279–1304.
Pennacchi, G., 1991. Identifying the dynamics of real interest rates and inflation evidence using survey data. Review of Financial Studies 4, 53–86.
Polson, N., Stroud, J., Müller, P., 2001. Affine state-dependent variance models. Working paper, University of Chicago and Wharton.
Raftery, A., 1996. Hypothesis testing and model selection. In: Gilks, W., Spiegelhalter, D., Richardson, S. (Eds.), Markov Chain Monte Carlo in Practice. Chapman & Hall, London, pp. 163–188.
Sanford, A., Martin, G., 2003. Simulation-based Bayesian estimation of affine term structure models. Working paper, Monash University.
Vasicek, O., 1977. An equilibrium characterization of the term structure. Journal of Financial Economics 5, 177–188.