

Can Interest Rate Volatility be Extracted From the Yield Curve? A Test of Unspanned stochastic Volatility*

Pierre Collin-Dufresne[†]

Robert S. Goldstein[‡]

Christopher S. Jones[§]

First Version: January 25, 2002

This Version: April 25, 2004

*We thank seminar participants at UCLA, Cornell University, McGill, the University of Minnesota, UNC, 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 helpful comments. We would like to thank Luca Benzoni, Michael Brandt, Mike Chernov, Qiang Dai, Jefferson Duarte, Garland Durham, Bing Han, Mike Johannes, and Ken Singleton for many helpful comments.

[†] Associate Professor at the Graduate School of Industrial Administration, Carnegie Mellon University, GSIA 315A, 5000 Forbes ave, Pittsburgh PA 15213, dufresne@andrew.cmu.edu.

[‡] Associate Professor at the Olin School of Business, Washington University in St. Louis, Campus Box 1133, One Brookings Drive, St. Louis, MO 63130, goldstein@olin.wustl.edu

[§] Assistant Professor at the Marshall School of Business, University of Southern California, 701 Hoffman Hall, Los Angeles, CA 90089, christopher.jones@marshall.usc.edu.

Can Interest Rate Volatility be Extracted From the Yield Curve?

A Test of Unspanned stochastic Volatility

Abstract

Most affine models of the term structure with stochastic volatility predict that the variance of the short rate is simultaneously a linear combination of yields *and* the quadratic variation of the spot rate. However, we find that the $A_1(3)$ model generates a time series for the variance state variable that is strongly *negatively correlated* with a GARCH estimate of the quadratic variation of the spot rate process. We then investigate affine models that exhibit ‘unspanned stochastic volatility (USV)’. We find that the $A_1(4)$ USV model generates both realistic volatility estimates and a good cross-sectional fit. Separately, we identify a transformation of the original model that allows us to replace the ‘latent’ state vector with an observable one. With this representation: (i) the state variables have simple physical interpretations such as level, slope and curvature, (ii) their dynamics remain affine and tractable, (iii) the model is by construction ‘maximal’ (i.e., it is the most general model that is econometrically identifiable), (iv) model-insensitive estimates of the state vector process implied from the term structure are readily available, and (v) those parameters that are not identifiable from bond prices alone if the model is restricted to exhibit ‘unspanned stochastic volatility’ (USV) are isolated.

Key words. Term Structure of Interest rates, Affine Models

JEL Classification Numbers. G12, G13.

1 Introduction

The affine class of term structure models as characterized by Duffie and Kan (DK, 1996) has become the dominant class of models because of its analytical tractability.¹ In particular, the affine class possesses closed-form solutions for both bond- and bond-option prices (Duffie, Pan, and Singleton (2000)), efficient approximation methods for pricing swaptions (Collin-Dufresne and Goldstein (2002b), Singleton and Umansev (2002)), and closed-form moment conditions for empirical analysis (Singleton (2001), Pan (2002)). As such, it has generated much attention both theoretically and empirically.²

However, recent papers have reported that standard affine models have trouble simultaneously fitting some cross-sectional and time-series properties of the yield curve (Duffee (2002), Dai and Singleton (2002b)). For example, Duffee (2002) reports that standard three-factor affine models cannot match the observed relationship between expected returns on bonds and the slope of the term structure. Duffee improves on this shortcoming by proposing a more flexible ‘essentially affine’ specification of the risk-premia. This added flexibility significantly reduces the tension between fitting expected returns, which are tied to physical measure dynamics, and fitting the cross-section of bonds, which are determined by the risk-neutral distribution.³ However, both Duffee and Duarte (2003) find that three factor affine models, even with generalized risk premia, cannot simultaneously capture both the time-variation in conditional variances and the forecasting power of the slope of the term-structure. Further, Duffee reports that adding a fourth factor would make his investigation impractical.

In this paper we also report a trade-off between capturing cross-sectional and time-series properties of the term structure. Here, however, the trade-off involves the second-order moments.⁴ Specifically, most affine models with stochastic volatility predict that the variance of the short rate is simultaneously a linear combination of yields *and* the quadratic variation of the spot rate. Yet, when we estimate the unrestricted essentially affine $A_1(3)$ model of the term structure, we obtain the ‘self-inconsistent’ result that the factors that explain the term structure are essentially unrelated to actual term structure volatility. In particular, the volatility factor extracted from this model using maximum likelihood estimation is strongly *negatively* correlated with volatilities estimated using standard GARCH or EGARCH models applied to the time series of the 6-month rate.

¹The affine class essentially includes all multi-factor extensions of the models of Vasicek (1977), and Cox, Ingersoll and Ross (1985).

²See the recent survey by Dai and Singleton (2003) and the references therein.

³See also Chacko (1997).

⁴Note 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).

We interpret these findings as evidence that the $A_1(3)$ model cannot simultaneously describe the yield curve level, slope, curvature, and volatility. That is, volatility is unable to play the dual role that the $A_1(3)$ model predicts that 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 restrictions imposed, that tradeoff is heavily tilted towards explaining the cross section.

We emphasize that our findings may have implications beyond the affine class of models. Indeed, using model-insensitive proxies for interest rate level (Y_{3Y}), slope ($Y_{10Y} - Y_{3Y}$), curvature ($Y_{10Y} - 2Y_{3Y} + Y_{6M}$) and a GARCH estimate for volatility, we find that these four features are (unconditionally, anyway) weakly correlated, suggesting that there may be no 3-factor model that can simultaneously capture these four features of the term structure.⁵

Given that standard affine models fail at producing a time series for the variance state variable that has anything to do with the quadratic variation of the spot rate, we also empirically investigate three and four-factor models that exhibit unspanned stochastic volatility. (Collin Dufresne and Goldstein (CDG 2002)). These models are constructed to break the tension between the time series and cross sectional features that most stochastic volatility affine models possess. In particular, these models impose parameter constraints so that the variance can not be determined from a linear combination of yields. We find that of the models investigated, only the $A_1(4)$ USV model is able to generate both a good cross sectional and time series fit of yields. An implication 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 (1993). Further, given the sensitivity of option prices to the specification of volatility dynamics, realistically captured only by the USV models, we speculate that explicitly imposing USV conditions may be useful for pricing such derivatives.

In order to facilitate empirical estimation of these models, we identify a transformation of the ‘canonical representation’ of Dai and Singleton (DS 2000) that we find to be very useful. Recall that the affine framework is defined by a finite set of N latent state variables $\{X_i(t)\}$ possessing joint Markov dynamics such that the spot rate, the risk-neutral drift, and covariance matrix are all affine functions of the state vector. Here, the state variables $\{X_i\}$ are referred to as ‘latent’ in that their physical meaning is inherently tied to the parameter values of the model. That is, the physical definition of these latent state variables changes each time a new trial parameter vector is considered. This poses both theoretical and practical econometric problems. For example, as stressed by DS, an important econometric issue in dealing with latent variables is that some of the

⁵We note that Brandt and Chapman (2004) report an estimate of a three-factor quadratic Gaussian model that appears to perform extremely well with respect to the moments they choose to capture. However, they do not attempt to fit the correlation between variance and curvature.

model's parameters may not be identifiable. Indeed, DS identify a set of transformations which can be performed on either the state vector, or its dynamics, that leave security prices unchanged.⁶ DS define a model to be 'maximal' if it has the maximum number of identifiable parameters and if it generates well-defined (i.e., admissible) dynamics for the latent state vector. Given a specific risk-premium specification, DS identify a 'canonical representation' for maximal affine three-factor models in terms of the physical measure dynamics of a latent state vector.

In contrast to the 'invariant transformation' approach of DS, below we identify a set of rotations which allows us to transform an underidentified affine model with latent state variables into a maximal affine model with observable state variables. (i.e., state variables that have physical meaning independent of the parameter values of the model.) Specifically, we rotate so that our state vector is composed of two types of variables: i) the first few components in the Taylor series expansion of the yield curve, and ii) their quadratic covariation. This representation has several advantages over those previously proposed in the literature. First, this representation permits model-insensitive estimates of the state variables to be estimated directly from the yield curve, which in turn can be used (e.g., in a simple OLS estimation procedure) to obtain an excellent first-guess for the parameter vector before embarking upon a more formal estimation procedure.⁷ This significantly reduces the probability that the 'maximum likelihood' found was in fact just a local maximum whose parameter vector generates poor properties for features of the term structure that are not directly investigated. Second, within this representation all risk-neutral parameters can be identified from the cross-section of fixed income security prices alone, even before the risk-premia structure has been specified. The risk-premia structure can then be specified with as much flexibility as required to fit the time series properties of the state variables. We use this fact to introduce the notion of Q-maximality, which is the maximum number of parameters that can be identified from cross-sectional information only. In contrast to the Q-maximal number of parameters, which has always remained the same, the total maximal number of parameters that can be identified has changed as researchers have moved from the 'completely affine' (e.g., DS) to the essentially affine (e.g., Duffee (2002)) to the 'generalized essentially affine' (e.g., Cheridito et al (2003)) risk premia structure.

Finally, this representation is very helpful in identifying and estimating maximal models that exhibit USV. As demonstrated by CDG, if a model displays USV, then there are parameters that are not identifiable from bond prices alone, even if the model is maximal in the sense of DS 2000. This is because the notion of maximality introduced by DS implicitly assumes that, in addition to

⁶DS identify three such types of 'invariant transformations': (i) rotation of the state vector \mathcal{T}_A , (ii) Diffusion rescaling \mathcal{T}_D , (iii) Brownian motion rotation \mathcal{T}_O .

⁷In contrast, when the state vector is latent, the parameter vector and state vector must be determined simultaneously, making it very difficult to come up with a reasonable parameter vector a priori.

bonds, other fixed income derivatives are observed.⁸ Our representation isolates these parameters, permitting the remaining parameters of the model to be estimated from bond price data alone. Furthermore, our representation simplifies the form of the parameter constraints imposed by USV, in turn facilitating empirical investigation.

Below we provide a full characterization of the maximal $A_1(3)$ and $A_1(4)$ models exhibiting USV. In addition, we propose an empirical approach for estimating USV models. Note that an immediate consequence of these models is that the one-to-one mapping assumed by DK (1996) between yields and factors does not hold. This in turn implies that standard estimation techniques, which rely on the ‘invertibility’ of the term structure with respect to the latent factors, cannot be implemented. Instead, we use a simulated likelihood-based approach based on the importance sampler of Richard and Zhang (1996, 1997).

The rest of the paper is as follows. In Section 2 we provide a general approach for deriving maximal affine models with observable state variables. In Section 3 we characterize the maximal $A_1(3)$ and $A_1(4)$ models exhibiting USV. In Section 4 we describe the estimation technique used to account for USV and to test the USV restrictions. We conclude in Section 5.

2 Identifying Maximal Affine Models with Observable State Vectors

Mostly following the notation of Dai and Singleton (2000), the risk-neutral dynamics of a Markov state vector X within an affine framework can be specified by as:

$$dX(t) = \mathcal{K}^Q \left(\theta^Q - X(t) \right) dt + \Sigma \sqrt{S(t)} dW^Q(t), \quad (1)$$

where \mathcal{K}^Q and Σ are $(N \times N)$ matrices, and S is a diagonal matrix with components

$$S_{ii}(t) = \alpha_i + \beta_i^\top X(t). \quad (2)$$

The spot rate is an affine function of X :

$$r(t) = \delta_0 + \delta_x^\top X(t), \quad (3)$$

where δ_x is an N dimensional vector. Assuming the system is admissible (that is, anything showing up under a square root is guaranteed to remain non-negative), then bond prices take the form:

$$P(t, \tau) = e^{A(\tau) - B(\tau)^\top X(t)}, \quad (4)$$

⁸Effectively, the invariant transformations that DS identify are those that leave the characteristic function of the state vector unchanged. However, there are many other invariant transformations that leave bond prices unchanged if the model displays USV. For example, arbitrary scaling of the volatility process parameters will not affect bond prices in the USV case.

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

$$\frac{dA(\tau)}{d\tau} = -\theta^{Q^\top} \mathcal{K}^{Q^\top} B(\tau) + \frac{1}{2} \sum_{i=1}^N \left[\Sigma^\top B(\tau) \right]_i^2 \alpha_i - \delta_0 \quad (5)$$

$$\frac{dB(\tau)}{d\tau} = -\mathcal{K}^{Q^\top} B(\tau) - \frac{1}{2} \sum_{i=1}^N \left[\Sigma^\top B(\tau) \right]_i^2 \beta_i + \delta_x, \quad (6)$$

and the initial conditions:

$$A(0) = 0, \quad B(0) = 0. \quad (7)$$

Defining bond yields $Y(t, \tau)$ via $P(t, \tau) = e^{-\tau Y(t, \tau)}$, we see from equation (4) that yields are affine in the state variables:

$$Y(t, \tau) = -\frac{A(\tau)}{\tau} + \frac{B(\tau)^\top}{\tau} X(t). \quad (8)$$

It is convenient to perform a Taylor series expansion on the yield curve and the time-dependent coefficients $A(\tau)$ and $B(\tau)$. To simplify notations we define $\partial_{\tau=c}^n f(t, \tau) := \frac{\partial^n}{\partial \tau^n} f(t, \tau) \Big|_{\tau=c}$. We find:

$$Y(t, \tau) = Y(t, 0) + \tau \partial_{\tau=0} Y(t, \tau) + \left(\frac{\tau^2}{2!} \right) \partial_{\tau=0}^2 Y(t, \tau) + \dots \quad (9)$$

$$A(\tau) = A(0) + \tau \partial_{\tau=0} A(\tau) + \left(\frac{\tau^2}{2!} \right) \partial_{\tau=0}^2 A(\tau) + \dots \quad (10)$$

$$B(\tau) = B(0) + \tau \partial_{\tau=0} B(\tau) + \left(\frac{\tau^2}{2!} \right) \partial_{\tau=0}^2 B(\tau) + \dots \quad (11)$$

Using the initial conditions in equation (7), and collecting terms of the same order τ , we find from equation (8) the following relation between the terms of the expansion:

$$\begin{aligned} Y_n(t) &\equiv \partial_{\tau=0}^n Y(t, \tau) \\ &= \frac{1}{n+1} \left(-\partial_{\tau=0}^{n+1} A(\tau) + \sum_{i=1}^N \partial_{\tau=0}^{n+1} B_i(\tau) X_i(t) \right) \quad \forall n = 0, 1, 2, \dots \end{aligned} \quad (12)$$

Equation (12) implies that the $\{Y_n\}$ variables are linear in the original latent state vector $\{X\}$, and hence can be chosen to be the state vector. That is, equation (12) implies that we can transform the original model, which is written in terms of latent variables $\{X\}$, into a model whose state vector is composed of some subset of the Taylor series components $\mathbf{Y} = \{Y_0, Y_1, \dots\}$ of the (observable) yield curve, and hence are themselves observable. As is well-known, any linear rotation of an affine model will generate a model that maintains an affine structure, implying that the instantaneous covariance matrix of the dynamics of the state vector will be affine in the $\{Y_n\}$. As such, in those cases where the covariance matrix is not constant, it will sometimes be

convenient to further rotate the state vector to include some subset of the quadratic (co-)variations of elements of \mathbf{Y} , which we call \mathbf{V} .

We emphasize that the state variables $\{Y_0, Y_1, Y_2\}$ of the representation have natural, physical interpretations as level, slope and curvature. Furthermore, by successive differentiation of the system of ODE's given in equations (5) and (6), and making use of the boundary conditions in equation (7), we can explicitly solve for the loadings in the definition of $\{Y_0, Y_1, Y_2\}$:

$$\partial_{\tau=0} A(\tau) = -\delta_0 \quad (13)$$

$$\partial_{\tau=0} B(\tau) = \delta_x \quad (14)$$

$$\partial_{\tau=0}^2 A(\tau) = -\theta^{Q\top} \mathcal{K}^{Q\top} \delta_x \quad (15)$$

$$\partial_{\tau=0}^2 B(\tau) = -\mathcal{K}^{Q\top} \delta_x \quad (16)$$

$$\partial_{\tau=0}^3 A(\tau) = \theta^{Q\top} \mathcal{K}^{Q\top} \mathcal{K}^{Q\top} \delta_x + \sum_{i=1}^N [\Sigma^\top \delta_x]_i^2 \alpha_i \quad (17)$$

$$\partial_{\tau=0}^3 B(\tau) = \mathcal{K}^{Q\top} \mathcal{K}^{Q\top} \delta_x - \sum_{i=1}^N [\Sigma^\top \delta_x]_i^2 \beta_i. \quad (18)$$

Plugging these into equation 12 and identifying the terms we find:

$$\begin{aligned} Y_0(t) &= \delta_0 + \delta_x^\top X(t) \\ &\equiv r(t) \end{aligned} \quad (19)$$

$$\begin{aligned} Y_1(t) &= \frac{1}{2} \delta_x^\top \mathcal{K}^Q (\theta^Q - X(t)) \\ &= \frac{1}{2} \frac{d}{dt} E_t^Q [dr(t)] \\ &\equiv \frac{1}{2} \mu^Q(t) \end{aligned} \quad (20)$$

$$\begin{aligned} Y_2(t) &= \frac{1}{3} \left(-\delta_x^\top \mathcal{K}^Q \mathcal{K}^Q (\theta^Q - X(t)) - \sum_{i=1}^N [\Sigma^\top \delta_x]_i^2 (\alpha_i + \beta_i^\top X(t)) \right) \\ &= \frac{1}{3} \frac{d}{dt} \left(E_t^Q [d\mu^Q(t)] - (dr(t))^2 \right) \\ &\equiv \frac{1}{3} \left(\frac{1}{dt} E_t^Q [d\mu^Q(t)] - V(t) \right). \end{aligned} \quad (21)$$

Hence, the level ($Y(t, 0)$), slope ($\partial_{\tau=0} Y(t, \tau)$) and curvature ($\partial_{\tau=0}^2 Y(t, \tau)$) are intimately related to the short rate, its risk-neutral drift, and the expected change in the drift minus the short rate's variance. In Appendix A1, we show that this relationship holds even outside of the affine framework.

Note that in the $A_0(3)$ sub-family of models, the covariance matrix of state vector dynamics is constant. As such, for our proposed rotation, all state variables should come from the Taylor series expansion of the yield curve. For example, an appropriate state vector for this class of models would consist of $(Y_0(t), Y_1(t), Y_2(t))$, or equivalently, $(r, \mu^Q, Y_2(t))$. However, for the $A_1(3)$ sub-family of models, where one state variable shows up linearly in the covariance matrix, it is sometimes convenient to rotate the state vector to (r, μ^Q, V) , where V is the quadratic variation process, defined by $dV_t = \frac{1}{dt}(dr_t)^2$. Note that the variance state variable is observable as well in that it has a physical interpretation independent of the model's parameter values. Below, we investigate the $A_1(3)$ model.

Conceptually, this rotation of the state vector to observables is similar to the original idea of DK (1996) who proposed to view affine models as ‘yield-factor’ models, effectively rotating from the latent state vector to an ‘observable’ state vector defined in terms of yields with constant time to maturities. However, their approach turns out to be intractable, in general, because the maximality restrictions take the form of restrictions on the solution of the Riccati equations, which are not generally known in closed-form.⁹ In contrast, our approach only involves the solution of these Riccati equations and their higher order derivatives at zero, all of which are known functions of the parameters.

2.1 Maximality of Proposed Representation

As discussed previously, the concept of maximality refers to the maximum number of parameters that are identifiable given a time series of prices for *all* fixed income securities (i.e., not just bonds). Here we demonstrate that, in addition to generating a state vector that is observable, rotating to our proposed representation automatically generates a model where all of the (remaining) parameters are identifiable econometrically. Furthermore, it identifies the maximum number of parameters that are identifiable from *cross sectional* information only. We term this property Q-maximality. It is useful to distinguish between maximality and Q-maximality because maximality, which refers to the total number of risk-neutral parameters and risk-premia that are identifiable, has changed as researchers have moved from the ‘completely affine’ (e.g., DS) to the essentially affine (e.g., Duffee (2002)) to the ‘generalized essentially affine’ (e.g., Cheridito et al (2003)) risk

⁹To provide some intuition, consider a two factor model where the two affine state variables are taken to be the one and two year-to-maturity yields. The arbitrage free yields of all maturities will be affine functions of the one and two year to maturity yields, with loadings on these two ‘yield state variables’ solving Riccati equations similar to those derived above. However, in addition to the standard boundary conditions at maturity, the solution to these Riccati equation must also satisfy the consistency conditions that the one and two year to maturity yields must be priced correctly. This implies an additional six restrictions on the loadings. The practical issue is that since the Riccati equations cannot be solved in closed-form, one can only use a numerical approach to constrain parameters of the ‘yield state vector’ to approximately satisfy these consistency conditions.

premia structure. However, the number of parameters that are identifiable from just cross sectional information has remained the same. Furthermore, as we demonstrate below, models which exhibit USV can generate bond prices possessing an exponential-affine structure even if the underlying dynamics are not affine. Under such circumstances, researchers might choose a risk premium structure that falls outside of even the ‘generalized essentially affine’ class. Once again, however, our rotation identifies the number of parameters that can be identified cross sectionally.

It is straightforward to demonstrate that all parameters that show up in the risk neutral dynamics under our representation are econometrically identifiable because we can readily identify fixed income securities whose prices allow us to determine their values. We demonstrate this claim with a simple example.

2.2 The $A_1(3)$ Model

Consider the $A_1(3)$ model of Dai and Singleton (2000). It is defined by 3 state variables, one of which follows a square-root process. One of its representations under the risk-neutral measure can be written with 18 parameters:

$$dv = (\gamma_V - \kappa_V v) dt + \sigma_V \sqrt{v} dz_v^Q \quad (22)$$

$$d\theta = [\gamma_\theta - \kappa_\theta \theta - \kappa_{\theta r} r - \kappa_{\theta v} v] dt + \sigma_{\theta r} \sqrt{\alpha_r + v} dz_r^Q + \sqrt{\xi^2 + \beta v} dz_\theta^Q + \sigma_{\theta v} \sqrt{v} dz_v^Q \quad (23)$$

$$dr = [\gamma_r - \kappa_r r - \kappa_{r\theta} \theta - \kappa_{rv} v] dt + \sqrt{\alpha_r + v} dz_r^Q + \sigma_{r\theta} \sqrt{\xi^2 + \beta v} dz_\theta^Q + \sigma_{rv} \sqrt{v} dz_v^Q. \quad (24)$$

DS demonstrate that this model is underidentified, and thus econometric analysis cannot determine all of the parameters. Following the approach proposed in the previous section, we rotate the $A_1(3)$ model from an unobservable latent state vector (r, θ, v) to the ‘observable’ state vector (r, μ^Q, V) . This rotation takes a maximum model which has 18 parameters, not all of which are identifiable, to a maximal model with 14 identifiable parameters inherent in its dynamics:

$$dV_t = (\gamma_V - \kappa_V V_t) dt + \sigma_V \sqrt{V_t - \psi_1} dZ_1^Q(t) \quad (25)$$

$$dr_t = \mu_t^Q dt + \sigma_1 \sqrt{V_t - \psi_1} dZ_1^Q(t) + \sqrt{\sigma_2^2 V_t - \psi_2} dZ_2^Q(t) + \sqrt{\sigma_3^2 V_t - \psi_3} dZ_3^Q(t) \quad (26)$$

$$d\mu_t^Q = (m_0 + m_r r_t + m_\mu \mu_t^Q + m_V V_t) dt + \nu_1 \sqrt{V_t - \psi_1} dZ_1^Q(t) + \nu_2 \sqrt{\sigma_2^2 V_t - \psi_2} dZ_2^Q(t) + \nu_3 \sqrt{\sigma_3^2 V_t - \psi_3} dZ_3^Q(t), \quad (27)$$

where by definition of V_t we have:

$$\sigma_1^2 + \sigma_2^2 + \sigma_3^2 = 1 \quad (28)$$

$$\sigma_1^2 \psi_1 + \psi_2 + \psi_3 = 0. \quad (29)$$

The model is admissible if¹⁰

$$\gamma_V \geq \kappa_V \psi_1 \quad (30)$$

$$\psi_1 \geq \max\left(\frac{\psi_2}{\sigma_2^2}, \frac{\psi_3}{\sigma_3^2}\right). \quad (31)$$

We claim that, since the state variables $\{r, \mu^Q, V\}$ are observable, all of the risk-neutral parameters in this model are identifiable. In particular, it is straightforward to select a set of fixed income securities whose prices will identify the values of any of these parameters. For example, consider the four parameters that show up in the risk-neutral drift of $\mu^Q(t)$, namely $\{m_0, m_r, m_\mu, m_V\}$. We claim that their values can be identified from four securities whose date-T cash flows are:

1. $e^{\int_t^T ds r(s)} \left(\mu^Q(T) - \mu^Q(t) \right)$
2. $e^{\int_t^T ds r(s)} \left(\mu^Q(T) - \mu^Q(t) + \int_t^T r(s) ds \right)$
3. $e^{\int_t^T ds r(s)} \left(\mu^Q(T) - \mu^Q(t) + \int_t^T \mu^Q(s) ds \right)$
4. $e^{\int_t^T ds r(s)} \left(\mu^Q(T) - \mu^Q(t) + \int_t^T V(s) ds \right)$

Indeed, the prices P of these four securities can be written as:

$$\begin{aligned} P(1) &= \mathbf{E}_t^Q \left[e^{-\int_t^T ds r(s)} e^{\int_t^T ds r(s)} (\mu^Q(T) - \mu^Q(t)) \right] \\ &= \mathbf{E}_t^Q \left[\int_t^T ds (m_0 + m_r r(s) + m_\mu \mu^Q(s) + m_V V(s)) \right] \\ P(2) &= \mathbf{E}_t^Q \left[e^{-\int_t^T ds r(s)} e^{\int_t^T ds r(s)} \left(\mu^Q(T) - \mu^Q(t) + \int_t^T r(s) ds \right) \right] \\ &= \mathbf{E}_t^Q \left[\int_t^T ds \left(m_0 + (m_r + 1) r(s) + m_\mu \mu^Q(s) + m_V V(s) \right) \right] \\ P(3) &= \mathbf{E}_t^Q \left[e^{-\int_t^T ds r(s)} e^{\int_t^T ds r(s)} \left(\mu^Q(T) - \mu^Q(t) + \int_t^T \mu^Q(s) ds \right) \right] \\ &= \mathbf{E}_t^Q \left[\int_t^T ds \left(m_0 + m_r r(s) + (m_\mu + 1) \mu^Q(s) + m_V V(s) \right) \right]. \\ P(4) &= \mathbf{E}_t^Q \left[e^{-\int_t^T ds r(s)} e^{\int_t^T ds r(s)} \left(\mu^Q(T) - \mu^Q(t) + \int_t^T V(s) ds \right) \right] \end{aligned}$$

¹⁰Note that as a practical matter it may be simpler to verify admissibility by using $v \equiv (V - \psi_1)$ as a state variable, since in this case zero is a natural lower boundary.

$$= \mathbf{E}_t^Q \left[\int_t^T ds \left(m_0 + m_r r(s) + m_\mu \mu^Q(s) + (m_V + 1) V(s) \right) \right].$$

It is easiest to prove our claim by considering securities that are about to mature. Indeed, in the limit $T \Rightarrow t + \Delta t$, we find

$$P(1) = \Delta t \left[m_0 + m_r r(t) + m_\mu \mu^Q(t) + m_V V(t) \right] \quad (32)$$

$$P(2) = \Delta t \left[m_0 + (m_r + 1) r(t) + m_\mu \mu^Q(t) + m_V V(t) \right] \quad (33)$$

$$P(3) = \Delta t \left[m_0 + m_r r(t) + (m_\mu + 1) \mu^Q(t) + m_V V(t) \right]. \quad (34)$$

$$P(4) = \Delta t \left[m_0 + m_r r(t) + m_\mu \mu^Q(t) + (m_V + 1) V(t) \right]. \quad (35)$$

Clearly, the coefficients $\{m_0, m_r, m_\mu, m_V\}$ can be determined from the prices $\{P(1), P(2), P(3), P(4)\}$, and the observable state variables $\{r(t), \mu^Q(t), V(t)\}$.

Just as intuitively, *given* a time series of the state vector, the historical measure parameters (or equivalently, the risk premia) can be identified via time series analysis. Indeed, it is clear that even from an OLS perspective, the coefficients $\{\beta\}$ in the regression

$$\left(r_{t+\Delta t} - r_t \right) = \beta_0 + \beta_1 r_t + \beta_2 \mu_t^Q + \beta_3 V_t + \epsilon \quad (36)$$

are identifiable.¹¹

In addition to the advantages mentioned above, our proposed rotation is especially invaluable for affine models that exhibit unspanned stochastic volatility (USV), because it isolates those parameters which are not identifiable from bond prices alone. Furthermore, this rotation allows us to express the parameter restrictions needed to generate USV in a much simpler form, in turn facilitating empirical investigation. DO WE WANT TO CONTRAST THIS WITH CDG 2002 AND CHERNOV?

The $A_1(3)$ model is identified above in equations (25)-(27). Alternatively, we can express the restrictions imposed by the maximality condition on the drift vector and instantaneous covariance matrix in the following form:

$$\begin{bmatrix} \frac{1}{dt} \mathbf{E}^Q [dr] \\ \frac{1}{dt} \mathbf{E}^Q [d\mu^Q] \\ \frac{1}{dt} \mathbf{E}^Q [dV] \end{bmatrix} = \begin{bmatrix} \mu^Q \\ m_0 + m_r r + m_\mu \mu^Q + m_V V \\ \gamma_V + \kappa_V V \end{bmatrix} \quad (37)$$

¹¹Admissibility might restrict some of the $\{\beta\}$ to have a particular relationship with their risk-neutral counterparts. That fact has no effect on the argument given here.

$$\Sigma^2 = \begin{bmatrix} V & c_0 + c_V V & \sigma_1 (\sigma_V V - \psi_1) \\ c_0 + c_V V & \sigma_0^\mu + \sigma_V^\mu V & \nu_1 (\sigma_V V - \psi_1) \\ \sigma_1 (\sigma_V V - \psi_1) & \nu_1 (\sigma_V V - \psi_1) & \sigma_V (\sigma_V V - \psi_1) \end{bmatrix}, \quad (38)$$

where, by definition

$$c_V = \sigma_1 \nu_1 + \sigma_2^2 \nu_2 + \sigma_3^2 \nu_3 \quad (39)$$

$$\sigma_V^\mu = \nu_1^2 + \nu_2^2 \sigma_2^2 + \nu_3^2 \sigma_3^2 \quad (40)$$

$$\sigma_0^\mu = -(\nu_1^2 \psi_1 + \nu_2^2 \psi_2 + \nu_3^2 \psi_3) \quad (41)$$

$$c_0 = -(\sigma_1 \nu_1 \psi_1 + \nu_2 \psi_2 + \nu_3 \psi_3). \quad (42)$$

Again we see that 14 risk-neutral parameters can be identified: six from the drift and eight from the variance-covariance matrix. Below, we will use both of these representations to simplify the notation.

2.3 Model-insensitive estimation of the state variables

When a model is specified in terms of latent state variables, estimates of the state vector depend on the assumed values of the parameters, which are not initially available. In contrast, as demonstrated above, the two state variables (r, μ^Q) in our representation are proportional to the level and slope of the term structure at zero. In theory, this suggests that it should be possible to obtain model-insensitive estimates of these state variables simply by observing the yield curve. Such estimates can be quite valuable. For example, they can be used to obtain reasonable estimates of the parameters, which in turn can be used as first guesses for a full-fledged estimation. This should be especially useful for multi-factor models with more than three factors.

In practice, however, we rarely observe the entire (continuous) term structure of zero-coupon yields. Rather, we only observe discrete points along the curve. Further, there may be some noise resulting from, e.g., bid-ask spread and non-synchronous trading. To investigate how this would affect the model-independent recovery of the state variables, we perform the following experiment. We simulate a two factor $A_2(2)$ model using the estimates of Duffie and Singleton (1997). We sample 10 years of weekly data and use a set of maturities typical of those used in the term structure literature, namely $\{0.5, 1, 2, 5, 7, 10\}$ years. Then we add i.i.d. noise with either 2b or 5bp standard deviations to account for potential ‘measurement errors.’ We estimate the level and slope at zero of the term structure by using two types of polynomials (quadratic and cubic). From our previous results the two state variables r and μ^Q can be estimated as, respectively, the level and twice the first derivative at zero. We then regress the estimates obtained from the polynomial

fits on the true value of the simulation, i.e., we perform the following regressions:

$$\begin{aligned}\text{true } r_t &= \alpha^r + \beta^r \times \text{estimated } r_t + \epsilon_t^r \\ \text{true } \mu_t^Q &= \alpha^\mu + \beta^\mu \times \text{estimated } \mu_t^Q + \epsilon_t^\mu,\end{aligned}$$

where r_t is the instantaneous short rate and μ_t^Q is its drift under the risk-neutral measure. If the model-independent estimates are unbiased and accurate, we expect to find coefficients β^r and β^μ close to one, along with high R^2 values. The results reported in Table 1 are encouraging. They show that the estimate of r is unbiased and accurate even given a high level of noise. Further, the estimate of r is insensitive to the type of polynomial used. The results for μ^Q are also quite good, but accuracy tends to diminish faster as noise increases. The R^2 drops as low as 89% in the high noise case for the less efficient cubic polynomial. Further, the order of the polynomial seems to matter for the estimate of the first derivative. For example, the quadratic spline seems to systematically bias the estimate ($\beta^\mu \approx 1.6$) of the second derivative. However, it is extremely highly correlated with the state variable ($R^2 \approx 0.98$).

We emphasize that we have made no particular effort to find an appropriate interpolation procedure. Rather, we have used the simplest available procedures, and did not try any others. These first results thus seem very promising. The first state variable can be recovered very accurately without much effort from available data. The second state variable can be recovered quite accurately with an appropriate interpolation/extrapolation procedure.¹² Below we demonstrate that similar accuracy is apparently obtained using actual data, since we find our model-insensitive estimates to be extremely highly correlated with estimates from full-fledged estimation procedures.

3 Identification of $A_1(3)$ and $A_1(4)$ Models

In this section we propose a full characterization of the ‘maximal’ $A_1(3)$ model exhibiting USV. Recall that by definition a model exhibits USV if state variables driving volatility risk cannot be hedged by trading in bond prices alone.

3.1 Maximal three factor model exhibiting USV

Collin-Dufresne and Goldstein (2002a) (proposition 6) provide six necessary and sufficient conditions for a three-factor affine model to exhibit USV. In applying these conditions to the particular

¹²We conjecture that a more sophisticated procedure based on either a term structure model (such as a two-factor Gaussian model) or a Nelson-Siegel-type spline would provide a more robust method for recovering r and μ^Q , even in the presence of substantial noise.

$A_1(3)$ framework, however, only one state variable enters the conditional covariance matrix, forcing two of the conditions to be satisfied automatically. The remaining conditions are:

$$m_r = -2c_V^2 \quad (43)$$

$$m_\mu = 3c_V \quad (44)$$

$$m_V = 1 \quad (45)$$

$$\sigma_V^\mu = c_V^2. \quad (46)$$

Interestingly, note that our representation leads naturally to the condition in equation (45). Indeed, equation (21) shows that $m_V = 1$ is a necessary condition for Y^2 to be independent of V , which in turn is a necessary condition for the entire yield curve to be independent of V .

Since the maximal $A_1(3)$ model has 14 risk-neutral parameters, and USV imposes 4 restrictions, the $A_1(3)$ -USV model has at most ten risk-neutral parameters (3 from the drift and 7 from the variance-covariance matrix) to estimate. However, we demonstrate below that once admissibility is enforced, the number of risk-neutral parameters gets reduced further to nine. Indeed, admissibility requires that the model satisfy both the USV conditions given in equations (43)-(46) and the admissibility conditions given in equations (28)-(31).

Combining the USV conditions (44) and (46) we see that to obtain stationary model under the Q measure, the parameters must satisfy:

$$c_V = -\sqrt{\sigma_V^\mu}.$$

Hence, from the definitions in equations (39) and (40), it follows that the parameters must satisfy the following system of equations:

$$\begin{cases} \sigma_1 \nu_1 + \sigma_2^2 \nu_2 + \sigma_3^2 \nu_3 & = -\sqrt{\nu_1^2 + \nu_2^2 \sigma_2^2 + \nu_3^2 \sigma_3^2} \\ \sigma_1^2 + \sigma_2^2 + \sigma_3^2 & = 1. \end{cases} \quad (47)$$

If we can find parameters that satisfy these equations, then the three USV conditions (43)-(45) can be satisfied by appropriately choosing the parameter values for m_r , m_μ , m_V . Further, the admissibility conditions (29)-(31) can be satisfied by appropriately choosing the parameter values for $\{\psi_1, \psi_2, \psi_3\}$.

To show there exists a solution to the system in equations (47), note that if we define the two vectors $\{u, v\}$ in \mathfrak{R}^3 by their coordinates $u = [\sigma_1, \sigma_2, \sigma_3]$ and $v = [\nu_1, \nu_2 \sigma_2, \nu_3 \sigma_3]$, then the system can be rewritten as:

$$\begin{cases} \|u\| & = 1 \\ u \cdot \frac{v}{\|v\|} & = -1. \end{cases} \quad (48)$$

The geometric interpretation is straightforward: The solution must satisfy $u = -\frac{v}{\|v\|}$, or equivalently:

$$\sigma_1 = -\frac{\nu_1}{\sqrt{\nu_1^2 + \nu_2^2 \sigma_2^2 + \nu_3^2 \sigma_3^2}} \quad (49)$$

$$\sigma_2 = -\frac{\nu_2 \sigma_2}{\sqrt{\nu_1^2 + \nu_2^2 \sigma_2^2 + \nu_3^2 \sigma_3^2}} \quad (50)$$

$$\sigma_3 = -\frac{\nu_3 \sigma_3}{\sqrt{\nu_1^2 + \nu_2^2 \sigma_2^2 + \nu_3^2 \sigma_3^2}}. \quad (51)$$

There are three possible solutions to this system:

- Case 1: $\sigma_2, \sigma_3 \neq 0$
- Case 2: $\sigma_2 = 0$
- Case 3: $\sigma_2 = \sigma_3 = 0$.

It can be shown that Case 1 is a degenerate case that reduces to a two-factor model, and that Case 3 is a special case of Case 2.¹³ Hence, we focus our attention on Case 2.

In this case equation (50) holds for any value of ν_2 . Further equation (49) and (51) imply that

$$\nu_1 = \sigma_1 c_V \quad (52)$$

$$\nu_3 = c_V. \quad (53)$$

Thus the system of equations becomes:

$$dV_t = (\gamma_V - \kappa_V V_t)dt + \sigma_V \sqrt{V_t - \psi_1} dZ_1^Q(t) \quad (54)$$

$$dr_t = \mu_t^Q dt + \sigma_1 \sqrt{V_t - \psi_1} dZ_1^Q(t) + \sqrt{(1 - \sigma_1^2)V_t + \sigma_1^2 \psi_1 + \psi_2} dZ_3^Q(t) + \sqrt{-\psi_2} dZ_2^Q(t) \quad (55)$$

$$d\mu_t^Q = (m_0 - 2c_V^2 r_t + 3c_V \mu_t^Q + V_t) dt + c_V \left(\sigma_1 \sqrt{V_t - \psi_1} dZ_1^Q(t) + \sqrt{(1 - \sigma_1^2)V_t + \sigma_1^2 \psi_1 + \psi_2} dZ_3^Q(t) \right) + \nu_2 \sqrt{-\psi_2} dZ_2^Q(t), \quad (56)$$

with the following conditions:

$$\kappa_V > 0 \quad \text{for stationarity} \quad (57)$$

$$c_V < 0 \quad \text{for stationarity} \quad (58)$$

$$\gamma_V - \kappa_V \psi_1 > 0 \quad \text{for admissibility} \quad (59)$$

$$-\psi_2 > 0 \quad \text{for admissibility} \quad (60)$$

$$1 > \sigma_1^2 \quad \text{for admissibility} \quad (61)$$

$$\psi_1 + \psi_2 > 0 \quad \text{for admissibility.} \quad (62)$$

¹³The proofs are available upon request.

Thus the model has 9 parameters under the Q measure:

$$\gamma_V, \kappa_V, \sigma_V, \psi_1, \sigma_1, c_V, \psi_2, m_0, \nu_2. \quad (63)$$

Finally, we verify that the short rate process given by equations (54)-(56) above exhibits USV in that the zero-coupon bond price is not a function of the volatility state variable V_t :

Proposition 1 *If the short rate process follows a three-factor Markov process given by equations (54), (55) and (56) where the parameters satisfy the admissibility conditions (57)-(63) then zero-coupon bond prices are given by:*

$$P(t, T) = \exp \left(A(T-t) - B_r(T-t)r_t - B_\mu(T-t)\mu_t^Q \right), \quad (64)$$

where the deterministic functions $A(\cdot)$, $B_r(\cdot)$ and $B_\mu(\cdot)$ are given by:

$$B_r(\tau) = \frac{-3 + 4e^{c_V\tau} - e^{2c_V\tau}}{2c_V} \quad (65)$$

$$B_\mu(\tau) = \frac{(1 - e^{c_V\tau})^2}{2c_V^2} \quad (66)$$

$$\begin{aligned} A(\tau) = & \frac{1}{96c_V^5} \left(-3e^{4c_V\tau}(2c_Vc_0 - \sigma_0^\mu) + 16e^{3c_V\tau}(3c_Vc_0 - \sigma_0^\mu) + 25\sigma_0^\mu \right. \\ & - 48e^{c_V\tau}(-5c_Vc_0 - 2c_V^2m_0 + \sigma_0^\mu) + 12e^{2c_V\tau}(2c_V(-6c_0 - c_Vm_0) + 3\sigma_0^\mu) \\ & \left. + 6c_V(-23c_0 + 12c_Vm_0 + 2(-6c_Vc_0 - 4c_V^2m_0 + \sigma_0^\mu)\tau) \right), \end{aligned} \quad (67)$$

and where the parameters c_0 and σ_0^μ can be written as

$$c_0 = \psi_2(c_V - \nu_2) \quad (68)$$

$$\sigma_0^\mu = \psi_2 \left[(c_V)^2 - (\nu_2)^2 \right]. \quad (69)$$

Proof: See Appendix A

From equations (65)-(67) it is clear that only four parameters, $\{m_0, \psi_2, \nu_2, c_V\}$, can be identified from the cross-section of bond prices. Further, from observing a time series of bond prices we can determine both the volatility state variable and the three diffusion parameters $\{\sigma_V, \sigma_1, \psi_1\}$. However, using only panel data on bond prices, we cannot determine the risk-neutral drift parameters (γ_V, κ_V) of V .¹⁴ Rather, prices of other fixed income derivatives (e.g., caps) must be used to infer these risk-neutral parameters.

¹⁴This 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.

Finally, we note that bond prices would retain their exponential-affine form in the above model for *any* specification of the process for V_t . Indeed, note that the proof of Proposition 1 does not depend in any way on the process followed by the variance of the short rate.¹⁵ In other words, bond prices can be exponential-affine even if the state vector is 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. On a similar note, going through the proof it is clear that the bond prices that obtain in our model are identical to bond prices that would obtain in a two-factor Gaussian model. Intuitively, since the volatility process does not affect bond prices in the USV case, our model makes the same cross-sectional predictions as a two-factor Gaussian model.¹⁶

3.2 $A_1(4)$ USV model

In this section we identify an $A_1(4)$ model which exhibits USV. For identification purposes, we see from equation (21) that is natural to use $\{Y_0, Y_1, Y_2, V\}$, or equivalently $\{r_t, \mu_t^Q, \theta_t^Q, V_t\}$ as the state vector, where θ_t^Q is defined by $\theta_t^Q \equiv 3Y_{2,t} = \left(\frac{1}{dt} E_t^Q [d\mu_t^Q] - V_t\right)$. The maximal $A_1(4)$ model is given by:

$$dV_t = (\gamma_V - \kappa_V V_t) dt + \sigma_V \sqrt{V_t - \psi_1} dZ_1^Q(t) \quad (72)$$

$$dr_t = \mu_t^Q dt + \sigma_1 \sqrt{V_t - \psi_1} dZ_1^Q(t) + \sum_{i=2}^4 \sqrt{\sigma_i^2 V_t - \psi_i} dZ_i^Q(t) \quad (73)$$

$$d\mu_t^Q = (\theta_t^Q + V_t) dt + \nu_1 \sqrt{V_t - \psi_1} dZ_1^Q(t) + \sum_{i=2}^4 \nu_i \sqrt{\sigma_i^2 V_t - \psi_i} dZ_i^Q(t) \quad (74)$$

$$d\theta_t^Q = (a_0 + a_r r_t + a_\mu \mu_t^Q + a_\theta \theta_t^Q + a_V V_t) dt + \eta_1 \sqrt{V_t - \psi_1} dZ_1^Q(t) + \sum_{i=2}^4 \eta_i \sqrt{\sigma_i^2 V_t - \psi_i} dZ_i^Q(t), \quad (75)$$

where by definition of V_t we have:

$$\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2 = 1 \quad (76)$$

¹⁵The only condition is that the volatility process be sufficiently regular for the stochastic integral in equation (A.20) to be a martingale.

¹⁶In fact, one can easily show that the cross-section of bond prices in the USV model is identical to that of a two-factor Gaussian model with the following dynamics:

$$dr_t = \mu_t^Q dt + \sigma_r dB_r(t) \quad (70)$$

$$d\mu_t^Q = (m_0 - 2c_V^2 r_t + 3c_V \mu_t^Q) dt + \sigma_\mu dB_\mu(t), \quad (71)$$

where B_r, B_μ are two correlated Brownian motions, and $dB_r(t) dB_\mu(t) = \rho dt$. Note that it has one less degree of freedom than the ‘maximal’ two-factor Gaussian model. In that sense, the maximal three factor USV model predicts a term structure of bond prices which is identical to a restricted two-factor Gaussian model. Of course, the time series properties of the USV model will be radically different (conditional variances are stochastic)!

$$\sigma_1^2 \psi_1 + \psi_2 + \psi_3 + \psi_4 = 0. \quad (77)$$

The model is admissible if

$$\gamma_V \geq \kappa_V \psi_1 \quad (78)$$

$$\psi_1 \geq \max \frac{\psi_i}{\sigma_i^2} \quad \forall i = 2, 3, 4 \text{ s.t. } \sigma_i \neq 0. \quad (79)$$

Note that the maximal unrestricted $A_1(4)$ model has a total of 22 free risk-neutral parameters $(\gamma_V, \kappa_V, \sigma_V, \{\psi_i, \nu_i, \eta_i, \sigma_i\}_{i=1}^4, a_0, a_r, a_\theta, a_\mu, a_V)$, and two restrictions from equations (76), (77).

For the $A_1(4)$ model to display USV, the model must satisfy certain restrictions. To identify these restrictions, it is convenient to define the vectors:

$$\sigma \equiv (\sigma_1, \sigma_2, \sigma_3, \sigma_4) \quad (80)$$

$$\nu \equiv (\nu_1, \nu_2 \sigma_2, \nu_3 \sigma_3, \nu_4 \sigma_4) \quad (81)$$

$$\eta \equiv (\eta_1, \eta_2 \sigma_2, \eta_3 \sigma_3, \eta_4 \sigma_4). \quad (82)$$

As for the $A_1(3)$ USV model, it is convenient to introduce a representation for the instantaneous variance covariance matrix of the state variables (r_t, μ_t^Q, θ_t) :

$$\Sigma^2 = \begin{bmatrix} V & c_{r\mu} V + c_{r\mu}^0 & c_{r\theta} V + c_{r\theta}^0 \\ c_{r\mu} V + c_{r\mu}^0 & (\sigma_\mu)^2 V + \sigma_\mu^0 & c_{\mu\theta} V + c_{\mu\theta}^0 \\ c_{r\theta} V + c_{r\theta}^0 & c_{\mu\theta} V + c_{\mu\theta}^0 & (\sigma_\theta)^2 V + \sigma_\theta^0 \end{bmatrix}, \quad (83)$$

where, by definition

$$c_{r\mu} \equiv \sigma \cdot \nu \quad (84)$$

$$c_{r\theta} \equiv \sigma \cdot \eta \quad (85)$$

$$c_{\mu\theta} \equiv \nu \cdot \eta \quad (86)$$

$$\sigma_\theta^2 \equiv \|\eta\|^2 \quad (87)$$

$$\sigma_\mu^2 \equiv \|\nu\|^2. \quad (88)$$

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

$$a_r = -2c_{r\mu}^2 (3c_{r\mu} - a_\theta) \quad (89)$$

$$a_\mu = 7c_{r\mu}^2 - 3c_{r\mu} a_\theta \quad (90)$$

$$a_V = 3c_{r\mu} \quad (91)$$

$$\sigma_\mu = -c_{r\mu} \quad (92)$$

$$\sigma_\theta = c_{r\theta} \quad (93)$$

$$c_{r\theta} = c_{r\mu}^2 \quad (94)$$

$$c_{\mu\theta} = c_{r\mu}^3 \quad (95)$$

As for the $A_1(3)$ USV model, there is a natural geometric interpretation for the restrictions on the variance covariance matrix. For example, equations (76), (84), (88), and (92) imply that the vectors σ and η are collinear but pointing in opposite directions. Similarly, equations (76), (85), (87) and (93) imply that the vectors σ and η are collinear and pointing in the same direction. Combining these results with the implications from equations (84), (85) and (94) we conclude that

$$\sigma = \frac{\eta}{\|\eta\|} = -\frac{\nu}{\|\nu\|} \quad (96)$$

$$\|\eta\| = \|\nu\|^2. \quad (97)$$

In order to identify the set of parameters that satisfy these restrictions, we investigate a few distinct cases.

- Case 1: $\sigma_2, \sigma_3, \sigma_4 \neq 0$. We claim that this case reduces to a two-factor model. Indeed, equation (96) implies that $\eta_2 = \eta_3 = \eta_4$ and $\nu_2 = \nu_3 = \nu_4$. Therefore, we can define the Brownian motion B_t^Q by $\sqrt{\sigma^2 V - \bar{\psi}} dB^Q(t) \equiv \sum_{i=2}^4 \sqrt{\sigma_i^2 V - \psi_i} dZ_i^Q(t)$, where $\bar{\sigma}^2 = \sum_{i=2}^4 \sigma_i^2$ and $\bar{\psi} = \sum_{i=2}^4 \psi_i$. It follows that the dynamics of the state vector is then adapted to the natural filtration generated by the two Brownian motions (Z_1^Q, B^Q) . That is, this case reduces to a two-factor model as claimed.
- Case 2: $\sigma_i = 0$ for some $i \in [2, 3, 4]$ and $\sigma_j \neq 0$ for all $j \in [2, 3, 4]$ such that $j \neq i$. Analogous to the previous case, we can show that this case reduces to a three factor model.
- Case 3: $\sigma_j = \sigma_i = 0$ for some $i \neq j \in [2, 3, 4]$. Without loss of generality, assume $\sigma_3 = \sigma_4 = 0$. Then equations (96) and (97) imply:

$$\nu_1 = c_{r\mu} \sigma_1 \quad (98)$$

$$\nu_2 = c_{r\mu} \quad (99)$$

$$\eta_1 = c_{r\mu}^2 \sigma_1 \quad (100)$$

$$\eta_2 = c_{r\mu}^2. \quad (101)$$

Further, from equations (76) and (77) we have:

$$\sigma_1^2 + \sigma_2^2 = 1 \quad (102)$$

$$\sigma_1^2 \psi_1 + \psi_2 + \psi_3 + \psi_4 = 0. \quad (103)$$

Combining all of these results, we obtain the following representation for the $A_1(4)$ USV model.

$$dV_t = (\gamma_V - \kappa_V V_t)dt + \sigma_V \sqrt{V_t - \psi_1} dZ_1^Q(t) \quad (104)$$

$$\begin{aligned} dr_t = & \mu_t^Q dt + \sigma_1 \sqrt{V_t - \psi_1} dZ_1^Q(t) + \sqrt{(1 - \sigma_1^2)V_t + \sigma_1^2 \psi_1 + \psi_3 + \psi_4} dZ_2^Q(t) \\ & + \sqrt{-\psi_3} dZ_3^Q(t) + \sqrt{-\psi_4} dZ_4^Q(t) \end{aligned} \quad (105)$$

$$\begin{aligned} d\mu_t^Q = & (\theta_t^Q + V_t) dt + c_{r\mu} \sigma_1 \sqrt{V_t - \psi_1} dZ_1^Q(t) + c_{r\mu} \sqrt{(1 - \sigma_1^2)V_t + \sigma_1^2 \psi_1 + \psi_3 + \psi_4} dZ_2^Q(t) \\ & + \nu_3 \sqrt{-\psi_3} dZ_3^Q(t) + \nu_4 \sqrt{-\psi_4} dZ_4^Q(t) \end{aligned} \quad (106)$$

$$\begin{aligned} d\theta_t^Q = & \left(a_0 - 2c_{r\mu}^2 (3c_{r\mu} - a_\theta) r_t + (7c_{r\mu}^2 - 3c_{r\mu} a_\theta) \mu_t^Q + a_\theta \theta_t^Q + 3c_{r\mu} V_t \right) dt \\ & + c_{r\mu}^2 \sigma_1 \sqrt{V_t - \psi_1} dZ_1^Q(t) + c_{r\mu}^2 \sqrt{(1 - \sigma_1^2)V_t + \sigma_1^2 \psi_1 + \psi_3 + \psi_4} dZ_2^Q(t) \\ & + \eta_3 \sqrt{-\psi_3} dZ_3^Q(t) + \eta_4 \sqrt{-\psi_4} dZ_4^Q(t). \end{aligned} \quad (107)$$

Note that the $A_1(4)$ model exhibiting USV has a total of 14 risk-neutral parameters ($\gamma_V, \kappa_V, \sigma_V, \psi_1, \psi_3, \psi_4, \nu_3, \nu_4, \eta_3, \eta_4, \sigma_1, a_0, c_{r\mu}, a_\theta$), as opposed to 22 for the unrestricted model.¹⁷

The admissibility restrictions are:

$$\kappa_V > 0 \quad \text{for stationarity} \quad (108)$$

$$c_{r\mu} < 0 \quad \text{for stationarity: see equation (92)} \quad (109)$$

$$a_\theta - 3c_{r\mu} < 0 \quad \text{for stationarity: see, for example, equation (116)} \quad (110)$$

$$\gamma_V - \kappa_V \psi_1 > 0 \quad \text{for admissibility} \quad (111)$$

$$\psi_3, \psi_4 < 0 \quad \text{for admissibility} \quad (112)$$

$$1 > \sigma_1^2 \quad \text{for admissibility: see equation (102)} \quad (113)$$

$$\psi_1 + \psi_3 + \psi_4 > 0 \quad \text{for admissibility.} \quad (114)$$

Note that this model nests the $A_1(3)$ USV model which may be obtained by setting $\psi_4 = \nu_4 = \eta_4 = 0$ and $a_\theta = 3c_{r\mu}$ and $\eta_3 = -2c_{r\mu}^2 + 3c_{r\mu}\nu_3$ (this can be readily verified by an appropriate change of variable in the previous model).

The following proposition verifies that the proposed model exhibits USV and provides the closed-form solution for bond prices.

¹⁷Note that the two restrictions $\sigma_3 = \sigma_4 = 0$ makes one of the seven restrictions from equations (89)-(95) redundant, leading to eight total restrictions, and thus $22 - 8 = 14$ parameters.

Proposition 2 *If the short rate process follows a four-factor Markov process given by equations (104)-(107) where the parameters satisfy the admissibility conditions (108)-(114) then zero-coupon bond prices are given by:*

$$P(t, T) = \exp \left(A(T-t) - B_r(T-t) r_t - B_\mu(T-t) \mu_t^Q - B_\theta(T-t) \theta_t^Q \right), \quad (115)$$

where the deterministic functions $A(\tau)$, $B_r(\tau)$, $B_\mu(\tau)$, and $B_\theta(\tau)$ are given by:

$$B_r(\tau) = \frac{e^{c_{r\mu} \tau} (6 c_{r\mu} - 2 a_\theta)}{4 c_{r\mu}^2 - c_{r\mu} a_\theta} + \frac{e^{2 c_{r\mu} \tau} (3 c_{r\mu} - a_\theta)}{-10 c_{r\mu}^2 + 2 c_{r\mu} a_\theta} + \frac{7 c_{r\mu} - 3 a_\theta}{-6 c_{r\mu}^2 + 2 c_{r\mu} a_\theta} - \frac{2 c_{r\mu}^2 e^{(-3 c_{r\mu} + a_\theta) \tau}}{\Gamma} \quad (116)$$

$$B_\mu(\tau) = \frac{a_\theta}{2 c_{r\mu}^2 (-3 c_{r\mu} + a_\theta)} + \frac{e^{2 c_{r\mu} \tau} (2 c_{r\mu} - a_\theta)}{10 c_{r\mu}^3 - 2 c_{r\mu}^2 a_\theta} + \frac{e^{c_{r\mu} \tau} (c_{r\mu} - a_\theta)}{-4 c_{r\mu}^3 + c_{r\mu}^2 a_\theta} + \frac{3 c_{r\mu} e^{(-3 c_{r\mu} + a_\theta) \tau}}{\Gamma} \quad (117)$$

$$B_\theta(\tau) = \frac{e^{c_{r\mu} \tau}}{c_{r\mu}^2 (-4 c_{r\mu} + a_\theta)} + \frac{1}{6 c_{r\mu}^3 - 2 c_{r\mu}^2 a_\theta} + \frac{e^{2 c_{r\mu} \tau}}{10 c_{r\mu}^3 - 2 c_{r\mu}^2 a_\theta} - \frac{e^{(-3 c_{r\mu} + a_\theta) \tau}}{\Gamma} \quad (118)$$

$$A(\tau) = \int_0^\tau \left(\frac{\sigma_\mu^0}{2} B_\mu(s)^2 + \frac{\sigma_\theta^0}{2} B_\theta(s)^2 + B_r(s) B_\mu(s) c_{r\mu}^0 + B_r(s) B_\theta(s) c_{r\theta}^0 + B_\theta(s) B_\mu(s) c_{\mu\theta}^0 - B_\theta(s) a_0 \right) ds,$$

and where the parameters are given by:

$$\Gamma = (3 c_{r\mu} - a_\theta) (4 c_{r\mu} - a_\theta) (5 c_{r\mu} - a_\theta). \quad (119)$$

Proof: See Appendix A

4 Empirical Implementation

Of primary empirical interest is 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 USD swap rates to estimate a variety of maximal two-, three-, and four-factor affine models both with and without USV.

As discussed in the previous section, the volatility state variable does not enter into the bond price formulas for those models which exhibit USV. As such, the $A_1(3)$ USV model is effectively a two-factor model in the cross-sectional sense and therefore bears some resemblance to both the unrestricted $A_1(2)$ and $A_1(3)$ models. The latter model, with three factors in the yield curve, motivates examination of the $A_1(4)$ USV specification, which also has three factors in yields but which has an additional volatility factor that is free to explain time series patterns.

While USV seems desirable from evidence on derivatives-pricing (CDG 2002, Heiddari and Wu (2003)), it remains to be seen whether USV is too restrictive of an assumption for bond prices

themselves. 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.

4.1 Model specifications to be estimated

In Section 3 we introduced a representation of the $A_1(3)$ model to establish maximality. Note that this was accomplished even though we specified only the risk-neutral dynamics. To complete the model, however, we also need to specify the risk-premia $\{\lambda\}$, which link the Brownian motions under the historical measure and risk-neutral measure via:

$$dZ_i^P(t) = dZ_i^Q(t) - \lambda_i(t) dt, \quad \forall i = 1, 2, 3, \quad (120)$$

We specify the $\lambda_i(t)$ as:

$$\lambda_1(t) = \frac{\lambda_{10} + \lambda_{13}V_t}{\sqrt{V_t - \psi_1}} \quad (121)$$

$$\lambda_2(t) = \frac{\lambda_{20} + \lambda_{21}r_t + \lambda_{22}\mu_t^Q + \lambda_{23}V_t}{\sqrt{\sigma_2^2V_t - \psi_2}} \quad (122)$$

$$\lambda_3(t) = \frac{\lambda_{30} + \lambda_{31}r_t + \lambda_{32}\mu_t^Q + \lambda_{33}V_t}{\sqrt{\sigma_3^2V_t - \psi_3}}. \quad (123)$$

By including a term in (121) proportional to $1/\sqrt{V_t - \psi_1}$, we are in fact generalizing Duffee's (2002) essentially affine specification. While, the Novikov condition may not be satisfied, a simple application of theorem 7.19 p.294 in Liptser and Shiryaev (1974) shows that if zero is not accessible by $V_t - \psi_1$ under both measures then the two measures implicitly defined by the Market price of risks above are equivalent.¹⁸ We therefore impose the Feller condition, for both measures, as a constraint in the likelihood maximization. Combined with (31), the Feller condition implies that the Radon-Nikodym density obtained from Girsanov's theorem is integrable. Below, we will see whether or not the Feller constraints are binding.

The flexibility of this specification of risk-premia allows for every drift parameter in the r and μ^Q processes be adjusted when changing measures. For simplicity of exposition we use the following notation: we denote by λ_{xy} the adjustment in the drift of x to the loading on y , where $x \in \{r, \mu^Q, V\}$ and $y \in \{0, r, \mu^Q, V\}$ (where 0 denotes a constant). The completely affine risk premium of the first Brownian motion restricts the adjustment of the drift parameters of the V equation to be proportional. As a result, the dynamics of the state vector for the unrestricted $A_1(3)$ under the P measure are:

$$dV_t = \left(\gamma_V + \lambda_{V0} - (\kappa_V - \lambda_{VV})V_t \right) dt + \sigma_V \sqrt{V_t - \psi_1} dZ_1(t) \quad (124)$$

¹⁸See also the recent paper by Cheridito, Filipovic, and Kimmel (2003) on the subject.

$$dr_t = \left(\lambda_{r0} + \lambda_{rr}r_t + (1 + \lambda_{r\mu})\mu_t^Q + \lambda_{rV}V_t \right) dt + \sigma_1 \sqrt{V_t - \psi_1} dZ_1(t) + \sqrt{\sigma_2^2 V_t - \psi_2} dZ_2(t) + \sqrt{\sigma_3^2 V_t - \psi_3} dZ_3(t) \quad (125)$$

$$d\mu_t^Q = \left(m_0 + \lambda_{\mu 0} + (m_r + \lambda_{\mu r})r_t + (m_\mu + \lambda_{\mu\mu})\mu_t^Q + (m_V + \lambda_{\mu V})V_t \right) dt + \nu_1 \sqrt{V_t - \psi_1} dZ_1(t) + \nu_2 \sqrt{\sigma_2^2 V_t - \psi_2} dZ_2(t) + \nu_3 \sqrt{\sigma_3^2 V_t - \psi_3} dZ_3(t). \quad (126)$$

With this specification, the unrestricted $A_1(3)$ model has a total of 24 parameters (14 risk neutral and 10 risk-premium parameters). The USV model, on the other hand, has a only 17 parameters that can be estimated from bond prices (9 risk-neutral and 10 risk-premium parameters, but the two volatility risk premia parameters are not identifiable from bond prices alone).

For comparison, we also consider models with completely affine risk premia, so that

$$\lambda_1(t) = \lambda_1 \sqrt{V_t - \psi_1} \quad (127)$$

$$\lambda_2(t) = \lambda_2 \sqrt{\sigma_2^2 V_t - \psi_2} \quad (128)$$

$$\lambda_3(t) = \lambda_3 \sqrt{\sigma_3^2 V_t - \psi_3}. \quad (129)$$

For both USV and non-USV specifications, all three completely affine risk premia parameters are identifiable from bond prices.

For the $A_1(4)$ USV model, essentially affine risk premia can involve up to 17 parameters. Preliminary results suggested that this was too great a number to be estimated reliably, so our investigation of the essentially affine $A_1(4)$ USV specifications uses the restricted risk premia

$$\lambda_1(t) = \frac{\lambda_{10} + \lambda_{12}V_t}{\sqrt{V_t - \psi_1}} \quad (130)$$

$$\lambda_2(t) = \frac{\lambda_{20} + \lambda_{21}\mu_t^Q + \lambda_{22}V_t}{\sqrt{(1 - \sigma_1^2)V_t + \sigma_1^2\psi_1 + \psi_3 + \psi_4}} \quad (131)$$

$$\lambda_3(t) = \frac{\lambda_{30} + \lambda_{31}\mu_t^Q + \lambda_{32}V_t}{\sqrt{-\psi_3}} \quad (132)$$

$$\lambda_4(t) = \frac{\lambda_{40} + \lambda_{41}\mu_t^Q + \lambda_{42}V_t}{\sqrt{-\psi_4}}. \quad (133)$$

Note, however, that λ_{10} and λ_{12} are unidentified because of USV.

We include a dependence on μ_t^Q in each of the risk premia because of studies such as Fama and Bliss (1987), which document strong relations between slope and bond excess returns. Since μ^Q is proportional to the slope of the yield curve at zero, we would expect it to be related to more conventional measures of yield curve slope as well. We include dependence on V_t so that these risk premia nest the completely affine specification, which we also consider.

The unrestricted $A_1(2)$ model is the last specification considered. Its representation under the Q -measure is given by

$$dV_t = (\gamma_V - \kappa_V V_t) dt + \sigma_V \sqrt{V_t - \psi_1} dZ_1^Q(t) \quad (134)$$

$$dr_t = (\gamma_r - \kappa_{rr} r_t - \kappa_{rV} V_t) dt + \sigma_1 \sqrt{V_t - \psi_1} dZ_1^Q(t) + \sqrt{(1 - \sigma_1^2)V + \sigma_1^2 \psi_1} dZ_2^Q(t). \quad (135)$$

Generalized essentially affine risk premia for this model are

$$\lambda_1(t) = \frac{\lambda_{10} + \lambda_{12} V_t}{\sqrt{V_t - \psi_1}} \quad (136)$$

$$\lambda_2(t) = \frac{\lambda_{20} + \lambda_{21} r_t + \lambda_{22} V_t}{\sqrt{(1 - \sigma_1^2)V_t + \sigma_1^2 \psi_1}}, \quad (137)$$

yielding a total of 13 parameters (8 risk neutral plus 5 risk premia). Completely affine risk premia for this model are defined as usual.

4.2 Data

We use weekly LIBOR and swap rate data from Datastream from January 6, 1988, to December 30, 2003. 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. Swap rates are converted to zero-coupon rates assuming that they can be valued as par bond rates.¹⁹ Following Bliss (1997), we use the extended Nelson-Siegel method for bootstrapping.

A complication arises from the use of LIBOR rates because the swap rates used in our sample are quoted roughly nine hours later.²⁰ To overcome this problem, following Jones (2003), we estimate the ‘synchronized’ values of the LIBOR rates. The procedure is essentially a Bayesian smoothing algorithm that exploits the extremely high correlations between changes in LIBOR and swap rates of similar maturities. Jones (2003) shows that the errors of the procedure are typically on the order of one basis point, or roughly one third the magnitude of the errors one would make by using either the same morning’s LIBOR quote, the next morning’s quote, or the average of the two.

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

¹⁹If swap 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 market. 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 Sundareshan (2002)).

²⁰LIBOR rates are quoted by the British Bankers’ Association at 11:00am London time, while our swap rates are recorded at 5:00pm New York time.

underlying yield quote for each one. We therefore expect the bootstrapped yields to be particularly accurate for these maturities.

Ideally, we would fit the model to the data in their original form, without modification via temporal smoothing or bootstrapping. Our decision to ‘pre-process’ the raw data is for convenience only, as it linearizes the relation between observables (the yields) and unobservables (r_t , μ_t^Q , etc.). Using the raw swap and LIBOR rates would complicate an estimation procedure that is already computationally demanding due to the presence of latent variables. We proceed with these methodological caveats in place.

4.3 Econometric methodology

We estimate all models using a quasi-maximum likelihood approach similar to Fisher and Gilles (1996), Duffee (2002), and others. As is well known, for the unrestricted $A_1(3)$ model (and a given trial parameter vector), it is (theoretically) possible to identify the state variables given any three linear combinations of zero-coupon yields. Indeed, since any linear combination of yields is linear in the state variables, the state vector can be identified by matrix inversion. A standard procedure used in the literature is to assume that some arbitrarily chosen set of yields is observed without error. These yields are then used to identify the state variables. Further, it is assumed that the remaining yields are observed with error.²¹ The latter are interpreted as ‘measurement errors’ (e.g., Chen and Scott (1993) and Pearson and Sun (1994), Duffie and Singleton (1997), Collin-Dufresne and Solnik (2001), Duffee (2002)).

In contrast, in this paper we assume that the first two or three *principal components* of the term structure are observed without errors, and that the measurement errors apply to the remaining principal components. This approach has three advantages. First, it guarantees that our procedure will fit perfectly the most important principal components, which we know from Litterman and Scheinkman (1991) explain the vast majority of the variance of changes in yields. Second, by construction the factors are unconditionally orthogonal, which approximately ‘orthogonalizes’ the matrix of ‘measurement’ errors. Finally, this procedure circumvents the arbitrariness of the standard approach of fitting specific yields. In Table 2 we present the eigenvectors corresponding to the eigenvalues of the covariance matrix. The first two or three give the loadings of the linear combination of yields which we fit perfectly (i.e., those which are measured without ‘error’). Consistent with Litterman and Scheinkman (1991), the first three principal components can loosely be interpreted as level, slope, and curvature factors. Only the first six principal components are

²¹There are several notable exceptions to the practice of assuming that some yields are perfectly observed. Pennacchi (1991) and Brandt and He (2002) both assume that *all* yields are measured with error. In Pennacchi’s case, this assumption is tractable because his model is Gaussian. Brandt and He use Monte Carlo simulation to compute likelihoods for a two-factor CIR model.

reported, as the remaining two explain only 0.26% of the total variance in yield curve changes and appear to represent pure noise.²²

Note that the standard quasi-maximum likelihood approach relies on the ‘invertibility’ of the bond yields to identify the state vector. However, under USV there does not exist a one-to-one mapping between yields and factors since bond prices are independent of the volatility state variable. Moreover, the situation is even worse for the unrestricted $A_1(3)$ model if we assume that only two principal components are observed without error. Indeed, in that case none of the state variables can then be inverted from yields. To overcome this obstacle, we use the Efficient Importance Sampler (EIS) of Richard and Zhang (1996, 1997) to evaluate the quasi-likelihood function by Monte Carlo simulation.

This approach is somewhat unique in the literature on affine models, as authors such as Duan and Simonato (1999) and de Jong (2000) have applied the Kalman filter to affine models like ours in which both the mean and the variance are linear functions of state variables. A problem with this method is that for non-Gaussian models it is generally infeasible to derive the optimal nonlinear filter for the state variables, and a suboptimal linear filter must be used instead. This results in an incorrect specification of conditional variances, which, as Lund (1997) and de Jong (2000) argue, leads to inconsistent estimates. Interestingly, however, studies including Lund (1997), de Jong (2000), and 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 of this result is that the linear relationship between yields and latent state variables is strong in typical affine term structure models. Ignoring nonlinearities is therefore innocuous since even the linear filter results in a high degree of accuracy.

In the USV case, however, this result seems unlikely to hold. As Duan and Simonato (1999, p. 115) note, the Kalman filtered estimates of the state variables are linear projections on the observed yields. Since yield levels provide no information about the volatility state variable under USV, these projections should result in substantial errors in V_t , which drives all conditional variances. In this case, the inconsistency identified by Lund (1997) and de Jong (2000) could be especially severe. We therefore adopt the EIS algorithm of Richard and Zhang (1996, 1997), which computes quasi-likelihoods conditional on simulated paths of V_t . These likelihoods are then averaged to

²²There is one potential shortcoming resulting from the use of estimated principal components. When more data is added to the sample, revisions in parameter estimates will result both from the new information in those data and from the changes that they cause in the estimated principal component loadings.

obtain an unconditional value.²³

To describe the algorithm in more specific terms, assume that we are estimating one of the $A_1(3)$ specifications. Let ϕ denote the parameter vector and $\mathbf{P} = \{\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_T\}$ the time series of principal components of the yield curve, representing the data that the models are attempting to explain. The simulated QML approach is based on the observation that the likelihood function, $p(\mathbf{P}|\phi)$, can be written as the integral $\int p(\mathbf{P}, \mathbf{V}|\phi) d\mathbf{V}$, where $\mathbf{V} = \{V_1, V_2, \dots, V_T\}$ denotes the time series of the latent stochastic variance process.

The integral is evaluated using simulation. In general, the density $p(\mathbf{V}|\mathbf{P}, \phi)$ implied by the true model is unknown. But for an arbitrary density $p^*(\mathbf{V}|\mathbf{P}, \phi)$, we may still write

$$\begin{aligned} \int p(\mathbf{P}, \mathbf{V}|\phi) d\mathbf{V} &= \int p(\mathbf{P}, \mathbf{V}|\phi) \frac{p^*(\mathbf{V}|\mathbf{P}, \phi)}{p^*(\mathbf{V}|\mathbf{P}, \phi)} d\mathbf{V} \\ &\equiv \mathbf{E}^* \left[\frac{p(\mathbf{P}, \mathbf{V}|\phi)}{p^*(\mathbf{V}|\mathbf{P}, \phi)} \right]. \end{aligned} \quad (138)$$

We can therefore evaluate the integral by simulating stochastic variance paths from $p^*(\mathbf{V}|\mathbf{P}, \phi)$ and averaging the ratio inside the brackets. One can see that in the unlikely case that $p^*(\mathbf{V}|\mathbf{P}, \phi) = p(\mathbf{V}|\mathbf{P}, \phi)$, then

$$\frac{p(\mathbf{P}, \mathbf{V}|\phi)}{p^*(\mathbf{V}|\mathbf{P}, \phi)} = p(\mathbf{P}|\phi), \quad (139)$$

independent of V_t . In this case only one simulation would be needed, as every realization of \mathbf{V} would yield the same result.

The idea of Richard and Zhang's (1996, 1997) Efficient Importance Sampler is essentially to choose the $p^*(\mathbf{V}|\mathbf{P}, \phi)$ that minimizes the variation in

$$\ln p(\mathbf{P}, \mathbf{V}|\phi) - \ln p^*(\mathbf{V}|\mathbf{P}, \phi),$$

thereby reducing the number of simulations required to compute (138). Unlike more traditional importance sampling, such as Sandmann and Koopman (1998), EIS does not require the ad hoc construction of an approximate linearized model. The importance sampler is simply defined as the density (within a parametric class) that maximizes simulation efficiency. An appendix provides more details about the sampler.

Given each draw of \mathbf{V} , evaluation of $p(\mathbf{P}, \mathbf{V}|\phi)$ is fairly straightforward. Let $\mathcal{P}_t = \{\mathcal{P}_{0,t}, \mathcal{P}_{\epsilon,t}\}$, where $\mathcal{P}_{0,t}$ are principal components observed without error and $\mathcal{P}_{\epsilon,t}$ are the remaining principal components, assumed observed with some serially independent measurement errors. Then assuming that $\mathcal{P}_{0,t}$ has two elements, the *joint* process for \mathcal{P}_t and V_t is Markov. This follows from the

²³We attempted to use a variant of the better-known approach of Kim, Shephard, and Chib (1998) in a previous version of the paper, but we found generalizations of this method unreliable for non-USV specifications where the stochastic volatility process plays a somewhat complex role. We now find that the EIS approach generates a substantially different set of parameters and results in much higher likelihoods for the non-USV models, demonstrating that this approach is more appropriate in these cases. We thank the referee for pushing us in this direction.

fact that $\mathcal{P}_{0,t}$ and V_t jointly imply (both with and without USV) values for r_t , μ_t^Q , and V_t , which are themselves jointly Markov. We can therefore decompose the ‘augmented’ likelihood function as

$$\begin{aligned}
p(\mathbf{P}, \mathbf{V} | \phi) &= \prod_{t=1}^T p(\mathcal{P}_t, V_t | \mathcal{P}_{t-1}, V_{t-1}, \phi) \\
&= \prod_{t=1}^T p(\mathcal{P}_t, V_t | r_{t-1}, \mu_{t-1}^Q, V_{t-1}, \phi) \\
&= \prod_{t=1}^T p(\mathcal{P}_{0,t}, V_t | r_{t-1}, \mu_{t-1}^Q, V_{t-1}, \phi) p(\mathcal{P}_{\epsilon,t} | r_t, \mu_t^Q, V_t, \phi) \\
&= [\det(L_0^{*-1})]^T \prod_{t=1}^T p(r_t, \mu_t^Q, V_t | r_{t-1}, \mu_{t-1}^Q, V_{t-1}, \phi) p(\mathcal{P}_{\epsilon,t} | r_t, \mu_t^Q, V_t, \phi), \quad (140)
\end{aligned}$$

where L_0^* is defined below. The first equality follows from the Markov property, and the second from the invertibility of \mathcal{P}_t and V_t . In the third, we separate, using Bayes rule, those principal components that are assumed perfectly observed, $\mathcal{P}_{0,t}$, from those that are assumed observed with error, $\mathcal{P}_{\epsilon,t}$. Irrelevant conditioning arguments are also eliminated.

The final equality again makes use of the invertibility of \mathcal{P}_t and V_t . Specifically, let K_0 (1×2) and L_0 (3×2) be the matrices such that

$$\mathcal{P}_{0,t} = K_0 + [r_t \ \mu_t^Q \ V_t] L_0. \quad (141)$$

Note that K_0 is a weighted sum of $A(\tau)$ -coefficients, where the weights correspond to the principal component loadings. Similarly, L_0 is a weighted sum of $B(\tau)$ -coefficients. Then defining

$$L_0^* \equiv \begin{bmatrix} 0 \\ L_0 \\ 1 \end{bmatrix}, \quad (142)$$

we obtain

$$[\mathcal{P}_{0,t} \ V_t] = [K_0 \ 0] + [r_t \ \mu_t^Q \ V_t] L_0^*, \quad (143)$$

which implies the ‘inversion formula’

$$[r_t \ \mu_t^Q \ V_t] = ([\mathcal{P}_{0,t} \ V_t] - [K_0 \ 0]) L_0^{*-1}. \quad (144)$$

The term $[\det(L_0^{*-1})]^T$ in the likelihood decomposition thus reflects the change in variables from $\{\mathcal{P}_{0,t}, V_t\}$ to $\{r_t, \mu_t^Q, V_t\}$.²⁴

²⁴In the case of the $A_1(3)$ USV model, the last row of L_0 is a row of zeros, so the vector $[r_t \ \mu_t^Q]$ can be recovered perfectly by post-multiplying $\mathcal{P}_{0,t} - K_0$ by the inverse of the first two rows of L_0 .

To summarize, we use 1,000 simulations from the importance sampling density $\mathbf{p}^*(\mathbf{V}^i | \mathbf{P}, \phi)$ to approximate the likelihood function as

$$\frac{1}{1000} \sum_i \frac{[\det(L_0^{*-1})]^T \prod_{t=1}^T \mathbf{p}(r_t^i, \mu_t^{Q,i}, V_t^i | r_{t-1}^i, \mu_{t-1}^{Q,i}, V_{t-1}^i, \phi) \mathbf{p}(\mathcal{P}_{\epsilon,t} | r_t^i, \mu_t^{Q,i}, V_t^i, \phi)}{\mathbf{p}^*(\mathbf{V}^i | \mathbf{P}, \phi)}, \quad (145)$$

where the V_t^i are simulated from $\mathbf{p}^*(\mathbf{V}^i | \mathbf{P}, \phi)$ and $\{r_t^i, \mu_t^{Q,i}\}$ are inferred via equation (144). By design, $\mathbf{p}^*(\mathbf{V} | \mathbf{P}, \phi)$ is known in closed-form. As discussed above, $\mathbf{p}(r_t^i, \mu_t^{Q,i}, V_t^i | r_{t-1}^i, \mu_{t-1}^{Q,i}, V_{t-1}^i, \phi)$ is treated as Gaussian, where we use Fisher and Gilles' (1996) algorithm for computing conditional means and covariances. of mean reversion coefficients,

Finally, to compute $\mathbf{p}(\mathcal{P}_{\epsilon,t} | r_t, \mu_t^Q, V_t, \phi)$ we use the affine structure to find the K_ϵ and L_ϵ such that *without* measurement error we would have

$$\mathcal{P}_{\epsilon,t} = K_\epsilon + \begin{bmatrix} r_t & \mu_t^Q & V_t \end{bmatrix} L_\epsilon. \quad (146)$$

The density of $\mathcal{P}_{\epsilon,t}$ is computed assuming that measurement errors, causing deviations from these values, are Gaussian with mean zero with a diagonal covariance matrix Σ_ϵ .²⁵

An obvious improvement to the estimation procedure would be to use the true transition density for the state vector rather than the Gaussian approximation. Except for a few specific cases (such as the Gaussian or the independent square root models), the transition density for affine models is not known in closed-form. Given the computational demands of the importance sampling procedure, we refrain from considering more sophisticated density approximations such as those of Ait-Sahalia and Kimmel (2002).

4.4 Estimated state variables

To further understand the properties of the models, we will need methods to extract estimates of the latent state variables.²⁶ For the $A_1(2)$ model, the state vector is obtained by inverting the first two principal components. Analogously, for the $A_1(3)$ 3PC model, the state vector is obtained by inverting the first three principal components. For the remaining three models, however, it is not possible to identify all state variables from an inversion technique. Instead, smoothed estimates based on the entire sample of yields must be computed. Here we explain the procedure used to do so.

²⁵In many cases, assuming that the measurement error covariance matrix is diagonal would be unnatural. However, recall that we are working with principal components which are unconditionally orthogonal. Hence, cross-sectional error correlations are expected to be small enough to ignore.

²⁶We note that the use of the word 'latent' here differs from the way it was used in the first section of the paper, where latent referred to state variables that have no physical meaning independently of parameter values of the model. Here, the state variables do have physical meaning independent of parameter values, but they are 'latent' in that they are not directly observable in practice due to either 'noisy prices' or to the unavailability of data at a continuum of dates.

The model-implied short-rate (or short rate drift or variance) at time t for the unrestricted $A_1(3)$ model, for example, is computed as the expectation of r_t given the QML parameter estimates and the entire time series of principal components, or $E[r_t|\mathbf{P}, \phi]$. Following the estimation methodology, we can compute this expectation as

$$\begin{aligned}
\int r_t \mathbf{p}(r_t|\mathbf{P}, \phi) dr_t &= \int r_t \int \mathbf{p}(r_t, \mathbf{V}|\mathbf{P}, \phi) d\mathbf{V} dr_t \\
&= \int \int r_t \mathbf{p}(r_t|\mathbf{V}, \mathbf{P}, \phi) dr_t \mathbf{p}(\mathbf{V}|\mathbf{P}, \phi) d\mathbf{V} \\
&= \int E[r_t|\mathbf{V}, \mathbf{P}, \phi] \mathbf{p}(\mathbf{V}|\mathbf{P}, \phi) d\mathbf{V} \\
&= E^* \left[E[r_t|\mathbf{V}, \mathbf{P}, \phi] \frac{\mathbf{p}(\mathbf{V}|\mathbf{P}, \phi)}{\mathbf{p}^*(\mathbf{V}|\mathbf{P}, \phi)} \right], \tag{147}
\end{aligned}$$

where we have again used the importance sampling density introduced in equation (138). Given the joint invertibility of V_t and \mathcal{P}_t , r_t is nonstochastic given \mathbf{V} , \mathbf{P} , and ϕ , so $E[r_t|\mathbf{V}, \mathbf{P}, \phi]$ is simply equal to the value obtained via (144). The smoothed estimate of r_t can therefore be computed via simulation. $E[\mu_t^Q|\mathbf{P}, \phi]$ and $E[V_t|\mathbf{P}, \phi]$ can be computed analogously, the latter also being relevant for both USV models.

4.5 Estimation results

Table 3 presents likelihood-based analysis of several variations on each of five models. The first is the $A_1(2)$ model, which is an unrestricted two-factor model. Next are the $A_1(3)$ USV and unrestricted $A_1(3)$ models. Each of these three models assume that exactly two principal components are observed without error. This implies that the entire state vector $\{r_t, V_t\}$ of the $A_1(2)$ model can be inverted, without error, from the yield curve. For the $A_1(3)$ USV specification, only the state variables r_t and μ_t^Q can be inverted from yields, with V_t remaining latent. For the unrestricted $A_1(3)$ models, observing two yields is insufficient to infer any of the state variables.

The next model, labelled $A_1(3)$ 3PC, is identical to the unrestricted $A_1(3)$ specification except that it assumes that three principal components are perfectly observed. This allows the inversion of the entire state vector $\{r_t, \mu_t^Q, V_t\}$. Finally, the $A_1(4)$ USV model also assumes that three principal components are observed without error, which is sufficient to identify the state variables r_t , μ_t^Q , and θ_t , but not V_t .

We note that all the models other than $A_1(4)$ USV can be viewed as special cases of the unrestricted $A_1(3)$ model. The $A_1(3)$ USV model, for instance, imposes the four restrictions tabulated in equations (43) to (46). The $A_1(2)$ class, as defined by Dai and Singleton (2000), is a special case of the $A_1(3)$ class as well. Given our risk premia specification, there are total of seven restrictions in going from $A_1(3)$ to $A_1(2)$. Finally, the $A_1(3)$ 3PC model is a special case of $A_1(3)$ in that

the $A_1(3)$ 3PC model imposes the restriction that the measurement error standard deviation on the third principal component is zero, whereas the unrestricted model removes this restriction.

Table 3 first examines essentially affine specifications estimated using all six principal components over the 729 weeks of data from January 1988 to December 2001. Next we consider completely affine models estimated using the same data, and finally, where possible, we re-estimate the completely affine specifications using only the first two principal components of yields.

Panel A reports the statistics on essentially affine specifications. Among the five models, the $A_1(4)$ USV model displays the highest log likelihood both for the estimation sample and for a 105-week hold-out sample from January 2002 to December 2003. Akaike and Bayesian information criteria, which are lower for ‘preferred’ specifications, also favor $A_1(4)$ USV.

Given that three of the models are nested within the unrestricted $A_1(3)$ model, Table 3 includes likelihood ratio tests of the implicit restrictions. In Panel A, all restricted versions are strongly rejected, though we stress that these results completely ignore the test misspecification induced by the use of Gaussian transition densities. The rejection of the $A_1(2)$ and $A_1(3)$ USV models likely reflects the presence of at least three factors in the yield curve, as Litterman and Scheinkman (1991) find. The failure of $A_1(3)$ 3PC indicates a rejection of the hypothesis that the third principal component (which explains just over 8% of yield variation) is free of observation error.

Finally, Panel A reports that the Feller condition is binding for only one of the essentially affine specifications – the $A_1(2)$ model – and that it binds under the Q measure only. Below we will ignore this constraint in calculating standard errors, noting that results may be inaccurate because of it. Lastly, for both USV specifications the Feller condition can be evaluated only under the P measure, as volatility risk premia parameters are unobservable.

Panel B repeats the same analysis for models with completely affine risk premia. Little changes except that the Feller condition is always binding for non-USV specifications. Panel C then reports likelihood ratio tests comparing each completely affine model with its essentially affine counterpart. In all cases, there is a significant improvement from specifying more general risk premia, confirming the conclusions of Duffee (2002) that richer specifications are required to fit the data.

Panel D reconsiders completely affine specifications using a data set that consists only of the first two principal components of yields.²⁷ Likelihood ratio tests result in a rejection of the $A_1(2)$ against the unrestricted $A_1(3)$ model, but no rejection of $A_1(3)$ USV. We interpret this result as an indication that the failure of $A_1(3)$ USV is in fitting the cross-section of bond yields rather than in describing their time series properties. More evidence on this point is provided below.

²⁷Because of the use of a different data set, likelihood-based statistics are not comparable to those from other panels of the table.

Finally, Panel E reports the number of parameters used in each of the specifications. All models require the number of risk neutral parameters specified plus those parameters corresponding to the type of risk premia used. Models estimated using all six principal components also require the number of measurement error standard deviations given. These values imply the degrees-of-freedom used to obtain the LRT p-values.

Given the results in Table 3, we subsequently consider only the essentially affine specifications. In addition, we do not present results for the $A_1(3)$ 3PC model because they are not substantially different from the unrestricted $A_1(3)$ model.

Tables 4A and 4B therefore report parameter estimates, expressed on an annualized basis, for the remaining four essentially affine models. Standard errors are in parentheses. For completeness, we include estimates of restricted parameters under both USV specifications. For example, the parameter m_r is restricted to equal to $-2c_v^2$ for the $A_1(3)$ USV model, so the value reported for that model is implied by the estimate of c_v . Standard errors for these restricted parameters are calculated, where appropriate, using the delta method. In addition, we do not report the risk-neutral parameters γ_v and κ_v , but rather their P -measure counterparts $\gamma_v^P \equiv \gamma_v + \lambda_{v0}$ and $\kappa_v^P \equiv \kappa_v - \lambda_v$. We do so because only the latter are identified under USV.

In general, the four specifications offer dramatically different interpretations of the values of each parameter. For example, the $A_1(3)$ USV model results in an estimate of .465 for κ_v^P , implying a half-life of about 1.5 years for volatility shocks. The same half-life for the unrestricted $A_1(3)$ model, with $\kappa_v^P \approx .0071$, is almost 100 years, indicating vastly different time series behavior.

4.6 Yield curve fit

Rather than relying only on statistical evaluations, such as likelihood ratio tests, the appraisal of a term structure model must also account for that model's abilities in valuation and forecasting. In Table 5 we examine the accuracy of the model's in-sample fit of the yield curve, both in terms of bias and root mean squared error. Below, we also investigate out-of-sample performance.

Statistical tests for bias are relatively standard. Errors are defined as actual yields minus fitted yields, where fitted yields are computed via (8) using smoothed state variables calculated as in section 4.4. T-statistics are based on Newey-West (1987) standard errors. For ease of comparison, all standard errors in a given panel are calculated using the same lag length, which in the case of the top panel of Table 5 was 21. This lag length is 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 repeated in Tables 7, 9, and 10. Absolute values in excess of 1.96 and 2.56 are taken to imply significance at the 5% and 1% levels, respectively.

Estimated biases with these levels of significance are marked with one or two stars.

Statistical evaluation of RMSE is somewhat more complicated. Because there is no well-defined null hypothesis for the RMSE of a given model, the best we can do is to compare the RMSEs of two models to see if one is significantly higher than the other. In order to reduce the number of pairwise comparisons made, however, we report only those comparisons that we find interesting *ex ante*. These are:

- $A_1(2)$ versus $A_1(3)$ USV: Both of these models ‘explain’ the cross-section of yields with only two factors, so it is not obvious which model will out-perform the other.
- $A_1(3)$ USV versus unrestricted $A_1(3)$: Do the USV restrictions substantially worsen the ability of the model to fit the yield curve?
- Unrestricted $A_1(3)$ versus $A_1(4)$ USV: Does the addition of a ‘free’ volatility state variable allow the remaining three factors to better explain yields?

For each maturity, these three pairwise comparisons are made 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

$$\hat{e}_{1,t}^2 - \hat{e}_{2,t}^2. \quad (148)$$

In this case, a significantly positive mean would indicate the superiority of model 2 over model 1. Standard errors are again calculated using the method of Newey and West (1987, 1994) with 20 lags. 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 or two stars signifying the level of significance.

Table 5 reveals that all four models imply reasonably unbiased fits of individual yields, with few rejections of zero mean errors. Root mean squared errors are much clearer in their preference for the unrestricted $A_1(3)$ model and the $A_1(4)$ USV model, both of which have three factors that affect yields rather than the two implied by $A_1(2)$ and $A_1(3)$ USV. The errors from all four models are highly autocorrelated, an indication that all the models are misspecified.

In comparing $A_1(3)$ to the $A_1(4)$ USV model, the former seems to offer slightly better performance in sample. It is therefore surprising that this result is reversed in Table 6, which reports out-of-sample yield fits using data from January 2002 to December 2003. Significant biases in both models indicate that while both the $A_1(3)$ and $A_1(4)$ USV models fit the principal components reasonably well, they do not fit the individual yields as accurately. These deviations are

larger for $A_1(3)$ than they are for $A_1(4)$ USV model, as indicated by the former's significantly larger RMSEs.²⁸

One reason for the high RMSEs of the $A_1(2)$ and $A_1(3)$ USV models, both in and out of sample, is depicted in Figure 1, which shows time series of actual and model-implied curvature of yields. A vertical dotted line denotes the end of the estimation period. Curvature here is defined as defined as $(Y_{10y} - 2Y_{3y} + Y_{6m})$, which from Table 2 is seen to be similar in construction to the third principal component. It is therefore not that surprising that actual curvature, depicted by the solid black line, is indistinguishable with the curvature implied from the $A_1(3)$ and $A_1(4)$ USV models, both of which have a third factor that can explain variation in the third principal component. The $A_1(2)$ and $A_1(3)$ USV models, with only two cross-sectional factors, cannot generate fluctuations in curvature of a realistic magnitude.

4.7 Properties of model-implied time series

We now examine some properties of the model-implied state variables and other time series, where state variables are extracted using the methods in section 4.4.

Figure 2 shows the implied time series of $E[V_t | \mathbf{P}, \phi]$ for each specification along with volatilities from Nelson's (1991) EGARCH(1,1) model estimated from demeaned changes in the 6-month yield. A vertical dotted line again denotes the end of the estimation period. For both USV specifications, the model-implied and EGARCH volatilities track each other closely, but for the non-USV specifications there appears to be little or no relation at all.

Table 7 reports a variety of correlations between observed time-series and related model-implied variables over the full estimation period and two subsamples. Over all three periods, we see that every model is capable of matching both the average yield (defined as the average of the 0.5, 1, 2, 5, 7, and 10-year yields) and the slope of the yield curve (defined as $Y_{10y} - Y_{6m}$). As shown earlier, actual and model-implied curvature are also extremely close for the $A_1(3)$ and $A_1(4)$ USV models, but not so for $A_1(2)$ and $A_1(3)$ USV.

Model-implied volatilities for both USV specifications are closely related to volatilities from both EGARCH(1,1) and Bollerslev's (1986) GARCH(1,1) models, the latter two having a correlation of .951 over the entire sample. Volatility from the unrestricted $A_1(3)$ model is actually negatively correlated with both GARCH volatilities over all sample periods. Volatility from the $A_1(2)$ model is fairly highly correlated with both GARCH volatilities, though Figure 2 shows that the variation in $A_1(2)$ volatility is unreasonably low.

²⁸The superiority of the $A_1(4)$ USV model might be expected since it fits three principal components perfectly, rather than the two fitted by the $A_1(3)$ model. However, the $A_1(3)$ 3PC specification, which fits an additional principal component without error, delivers performance virtually indistinguishable from the otherwise identical $A_1(3)$ model.

The bottom panel Table 7 also reports correlations of model volatilities with an average of one-year implied volatilities from cap and floor contracts, which are available from DataStream after 1995. Since by convention implied volatilities are determined assuming LIBOR rates follow a geometric Brownian motion, this volatility may be interpreted as the volatility of proportional changes in (or logarithms of) one-year LIBOR rates. We therefore also report correlations with the product of implied volatility and the level of the one-year rate. Under reasonable assumptions, this will approximate the volatility of the *level* of the one-year rate, making it more comparable to the other volatility proxies included.

We find that volatilities from both USV specifications are moderately to highly correlated with the two implied volatility measures, and also that these implied volatilities are about as closely related to the two GARCH series. Volatilities from the non-USV specifications are weakly or negatively correlated with the option-implied series. As such, we speculate that the USV models would therefore be much more successful in pricing such derivatives.

The table also reports correlations between various volatility measures and the actual curvature in yields, as defined in Figure 1. In general, this relationship is weak for the USV specifications and for the GARCH and implied volatilities as well. For the $A_1(3)$ model, the relationship is stronger and positive, which is surprising given the negative correlation between volatility and curvature that Litterman, Scheinkman, and Weiss (1991) find more theoretically plausible.²⁹ Only for ‘volatility’ extracted from the $A_1(2)$ model is the correlation large and negative.

Finally, Table 7 reports correlations between model-implied estimates of the state variables and the values obtained from the ‘interpolation’ scheme of section 2.3. In general, interpolated state variables are highly related to the values obtained through estimation of the model, which we take as strong empirical evidence of the observability of the state vector under our model rotations.

These results highlight the dual role 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 towards explaining the cross section. The result is that volatilities imputed from the two models without USV restrictions are essentially nonsensical, being unrelated to all other volatility proxies. Instead, both models use the variance process to provide a better fit of the cross

²⁹Litterman, Scheinkman, and Weiss (1991) examine whether the theoretical relation between yields and volatility holds empirically between January 1984 and June 1988. Regressing realized volatility on short, medium, and long-term yields results in an R-squared of .7 and slope coefficients whose signs suggest a relation between volatility and curvature. We note that their sample is short and that no standard errors were provided for their estimates. Using simulations, we find that R-squares above .6, coupled with coefficient estimates of the same sign as those reported by Litterman, Scheinkman, and Weiss (1991), occur with about 10% probability under the $A_1(4)$ USV model, in which volatility and curvature are essentially unrelated in population.

section, as evidenced by a relation between V_t and curvature that holds only for these two models.

The $A_1(3)$ USV model, meanwhile, generates reasonable volatility dynamics but cannot match curvature, which simply reinforces Litterman and Scheinkman’s (1991) finding that three factors are required to drive the yield curve. Only 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 the yield curve level, slope, curvature, and volatility. 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 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.

4.8 Forecasting performance

To reinforce these results we examine the forecasting performance of the same four models, both for evidence of misspecification and for assessing their potential usefulness in securities pricing and hedging. Because our sample size is relatively short, we focus on short horizon (one-week) forecasts of changes in yields and of two different volatility proxies. All forecasts are constructed using the parameter estimates reported in Table 4, so the bulk of our forecasts are in-sample. After using two years of data to initialize the forecasts, we are left with a 625-week in-sample period. With our hold-out sample from 2002 and 2003, we perform a 105-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 similarly to Section 4.4. Here, however, 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 the current values of the state variables, we simulate ten thousand 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.

Results for in-sample forecasts of yield changes are reported in Table 8. Out-of-sample forecasts appear in Table 9. In both tables, forecast errors are defined as the actual yield change minus $E_t[Y_{t+1}] - \hat{Y}_t$, where $E_t[Y_{t+1}]$ is the model-implied expectation and \hat{Y}_t is the model’s current fitted value. The statistical significance of biases, which are averages of these errors, is assessed using Newey-West standard errors. Pairwise comparison tests for root mean squared errors are tested with the method of Diebold and Mariano (1995), also with Newey-West standard errors. The Newey-West lag length selection, following the procedure outlined for Tables 5 and 6, results in

lags of 11 and 3 for the two panels of Table 8. For Table 9, 1 and 2 lags are used.

Unfortunately, yield forecasts fail to clearly distinguish the models, most likely because our sample is too short to evaluate forecasts of a relatively unpredictable time series. In sample, we see weak evidence favoring $A_1(3)$ USV over the $A_1(2)$ model in terms of RMSE. Otherwise, there are no statistically significant results in sample. Out of sample forecasts are slightly more informative, with the $A_1(3)$ USV model displaying significant biases across all maturities and the unrestricted $A_1(3)$ model having a strong bias at the 6-month maturity only. In no cases do we find RMSEs significantly different from one another.

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

$$\hat{\sigma}_{t,\tau} \equiv \sqrt{\frac{1}{N} \sum_{i=1}^N \Delta Y(t, i, \tau)^2}, \quad (149)$$

where $Y(t, i, \tau)$ is the τ -maturity yield observed on the i^{th} 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.

In-sample results on forecasted volatility are reported in Table 10. Newey-West lag lengths for the four panels of the table are 5, 5, 16, and 16, in that order. Throughout the table, the best performance, in terms of RMSE, is generally registered by the $A_1(4)$ USV model, though even that model displays significant biases in its forecasts of daily realized volatility.

Out of sample, $A_1(4)$ USV continues to perform well in forecasting volatilities on short-maturity yields, though long yield volatilities are best predicted by the unrestricted $A_1(3)$ model. This is surprising given that model's generally unrealistic portrayal of V_t as a variable that is even related to volatility. We believe the reason for this is that for most of the in-sample period, all yield volatilities were close to one-percent per year. However, during the out-of-sample period, short run volatility dropped to 0.5% per year, while the long run volatilities jumped to about 1.25% per year. Hence, during the out-of-sample period the short-run volatility changes were negatively correlated with long-run volatility changes. Unfortunately, such behavior cannot be captured by a single state variable for volatility. Rather, this empirical observation suggests that we may need more state variables to capture the 'term structure of volatilities', consistent with the findings of Heidari and Wu (2003) and Han (2004).

4.9 The Dai and Singleton (2003) challenges

In their review article, Dai and Singleton (2003) identify two empirical observations that have each proven somewhat of a challenge for affine term structure models. The first, which they label LPY, refers to the observation that a linear projection of yield changes on the lagged slope of the yield curve typically results in negative slope coefficients that are decreasing in maturity. This observation from Fama and Bliss (1987) and Campbell and Shiller (1991), among others, is a motivation for more general risk premia specification such as those of Duffee (2002) and Duarte (2003).

Figure 3 plots estimates of the slope coefficient in the regression of yield changes, $Y(t + 1, \tau) - Y(t, \tau)$, on lagged slope, $(Y(t, \tau) - Y(t, .5)) / (\tau - .5)$.³⁰ For consistency with Dai and Singleton (2003), we look at changes over four-week intervals. In each panel, the thick grey line depicts the sample slope coefficients from actual data as a function of the maturity τ . The solid black line denotes the average values computed from artificial 14-year data sets simulated using each model under the parameter values given in Table 4. The 95% confidence intervals from these distributions are depicted by dashed lines.

The figure shows that in all cases confidence intervals are rather large, mirroring the conclusion from Tables 8 and 9 that the sample we are using is probably too short to distinguish models on the basis of their yield forecasting performance. We note, however, that the unrestricted $A_1(3)$ and $A_1(4)$ USV specifications are the only to reproduce, at least on average, the convex shape of the actual relation between maturity and slope coefficients.

Dai and Singleton's (2003) second observation, labeled CVY, refers to time variation in conditional volatility and the hump shape in unconditional volatility as a function of maturity. Figure 4 displays the relation between maturity and the unconditional volatility of four-week yield changes. Results from actual data over the 1988-2001 sample are again displayed as a thick grey line. Means and 95% confidence intervals of model-implied sampling distributions are depicted by solid and dashed black lines, respectively.

The top two panels of Figure 4 reveal separate failures of the $A_1(2)$ and $A_1(3)$ USV models in explaining the volatility hump. Both models come close to matching average volatilities across maturities but fail to match patterns related to maturity, with the $A_1(2)$ model overstating short-maturity volatility and understating long-maturity volatility. While the $A_1(3)$ USV model produces a much wider distribution of possible sample volatilities, it is clear that the model overstates long-maturity volatility by roughly 20%. Both the $A_1(3)$ and $A_1(4)$ USV models perform well.

³⁰Due to data constraints, our specification differs slightly from that of Dai and Singleton (2003), whose left hand side variable is defined as $Y(t + 1, \tau - 1) - Y(t, \tau)$.

5 Conclusion

We have proposed a representation for affine term structure models in terms of the derivatives of the term structure at zero and their quadratic co-variations. These state variables have simple physical interpretation such as level, slope, and curvature. They are by construction observable from the cross-section of the yield curve. Thus our representation is ‘maximal’ (i.e., econometrically identified). Further, model-insensitive estimates of the process of the state variable are readily available, which simplifies the empirical estimation of the model.

We apply this representation to two-, three-, and four-factor affine stochastic volatility models. 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). Maximum likelihood estimation results in more weight placed on the first role at the expense of the second. We then investigate two ‘unspanned stochastic volatility’ models, where volatility does not enter the cross-section of bond prices. The three-factor USV model, which is nested within the unrestricted $A_1(3)$ model, dramatically improves the estimates of volatility at the expense of an inadequate cross-sectional fit. A four-factor 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 new representation introduced in this paper.

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. However, we find that the $A_1(4)$ USV model is able to provide both a good cross-sectional fit and a good description of yield volatility.

References

A Proofs

A.1: Proof of Generality of equations (19), (20) and (21)

Consider a Markov state vector $\{X(t)\}$ of length N with general (i.e., non-affine) risk-neutral dynamics

$$dX_i = m_i^Q(\{X\}, t) dt + \sum_{k=1}^N \sigma_{ik}(\{X\}, t) dz_k^Q. \quad (\text{A.1})$$

Further, assume the spot rate is some arbitrary function of the state vector: $r = r(\{X\}, t)$. Using the shorthand notations $r_i \equiv \frac{\partial r}{\partial X_i}$, $m_i^Q = m_i^Q(\{X\}, t)$ and $\sigma_{ik} = \sigma_{ik}(\{X\}, t)$, we obtain from Ito's lemma the dynamics for r :

$$dr = r_t dt + \sum_{i=1}^N r_i \left[m_i^Q dt + \sum_{k=1}^N \sigma_{ik} dz_k^Q \right] + \frac{1}{2} \sum_{i,j,k=1}^N r_{ij} \sigma_{ik} \sigma_{jk} dt. \quad (\text{A.2})$$

Note that this allows us to define

$$\begin{aligned} \mu^Q(t) &\equiv \frac{1}{dt} \mathbb{E}_t^Q [dr] \\ &\equiv r_t + \sum_{i=1}^N r_i m_i^Q + \frac{1}{2} \sum_{i,j,k=1}^N r_{ij} \sigma_{ik} \sigma_{jk}. \end{aligned} \quad (\text{A.3})$$

$$\begin{aligned} V(t) &\equiv \frac{1}{dt} \text{Var}_t^Q [dr] \\ &\equiv \sum_{i,j,k=1}^N r_i r_j \sigma_{ik} \sigma_{jk}. \end{aligned} \quad (\text{A.4})$$

Finally, from Ito's lemma we have

$$\mathbb{E}_t^Q [d\mu^Q(t)] = \mu_t^Q dt + \sum_{i=1}^N \mu_i^Q m_i^Q dt + \frac{1}{2} \sum_{i,j,k=1}^N \mu_{ij}^Q \sigma_{ik} \sigma_{jk} dt. \quad (\text{A.5})$$

Using the relationship between yield to maturity and bond prices

$$P(t, (\{X\}), T) \equiv e^{-(T-t)Y(t, \{X\}, T)}, \quad (\text{A.6})$$

and similar notations as above, we find

$$P_t = [Y - (T-t)Y_t] P \quad (\text{A.7})$$

$$P_i = [-(T-t)Y_i] P \quad (\text{A.8})$$

$$P_{ij} = [(T-t)^2 Y_i Y_j - (T-t)Y_{ij}] P \quad (\text{A.9})$$

Bond prices satisfy the PDE

$$rP = P_t + \sum_{i=1}^N P_i m_i + \frac{1}{2} \sum_{i,j,k=1}^N P_{ij} \sigma_{ik} \sigma_{jk} \quad (\text{A.10})$$

Plugging in equations (A.7)-(A.9), we find

$$r = [Y - (T - t) Y_t] - (T - t) \sum_{i=1}^N Y_i m_i + \frac{1}{2} \sum_{ijk=1}^N [(T - t)^2 Y_i Y_j - (T - t) Y_{ij}] \sigma_{ik} \sigma_{jk}. \quad (\text{A.11})$$

Now we use a Taylor series expansion to write yields as

$$\begin{aligned} Y(t, \{X\}, T) &= Y(t, \{X\}, T = t) + (T - t) Y_T(t, \{X\}, T = t) + \frac{1}{2} (T - t)^2 Y_{TT}(t, \{X\}, T = t) + \dots \\ &\equiv Y^0(t, \{X\}) + (T - t) Y^1(t, \{X\}) + \frac{1}{2} (T - t)^2 Y^2(t, \{X\}) + \dots \end{aligned}$$

Plugging this Taylor expansion into equation (A.11), and collecting terms of different orders of $(T - t)$, we find

$$(T - t)^0: Y^0(t, \{X\}) = r(t, \{X\}) \quad (\text{A.12})$$

$$(T - t)^1: Y^1(t, \{X\}) = \frac{1}{2} \mu^Q(t) \quad (\text{A.13})$$

$$(T - t)^2: Y^2(t, \{X\}) = \frac{1}{3} \left[\frac{1}{dt} \mathbf{E}_t^Q [d\mu^Q] - V(t) \right], \quad (\text{A.14})$$

which is what we wished to prove. \square

A.2: Proof of Proposition 1

To prove the proposition note that 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 equation (64). Indeed, in that case we have $e^{-\int_0^t r_s ds} P(t, T) = \mathbf{E}_t^Q \left[e^{-\int_0^T r_s ds} P(T, T) \right]$, which implies

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

since equations (64)-(67) 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 equation (64). Using the fact that the functions $A(\cdot)$, $B_r(\cdot)$ and $B_\mu(\cdot)$ satisfy the system of ODE:

$$B'_r = -2(c_V)^2 B_\mu + 1 \quad (\text{A.15})$$

$$B'_\mu = B_r + 3c_V B_\mu \quad (\text{A.16})$$

$$A' = \frac{1}{2} B_\mu^2 \sigma_0^\mu - B_\mu m_0 + B_r B_\mu c_0, \quad (\text{A.17})$$

and that, in particular, we have:

$$B_r = -c_V B_\mu + \sqrt{2B_\mu}, \quad (\text{A.18})$$

we find that

$$\mathbf{E}^Q [dP(t, T) - r_t P(t, T)] = 0 \quad (\text{A.19})$$

Thus,

$$\begin{aligned}
e^{-\int_0^t r_s ds} P(t, T) &= - \int_0^t \sqrt{2B_\mu(s)} \left(\sigma_1 \sqrt{V_s - \psi_1} dZ_1^Q(s) + \sqrt{(1 - \sigma_1^2)V_s + \sigma_1^2 \psi_1 + \psi_2} dZ_3^Q(s) \right) \\
&\quad - \int_0^t (B_r(s) + \nu_2 B_\mu(s)) \sqrt{-\psi_2} dZ_2^Q(s). \tag{A.20}
\end{aligned}$$

This shows that $e^{-\int_0^t r_s ds} P(t, T)$ is indeed a Q -martingale.

□

A.3: Proof of Proposition 2

To prove the proposition 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 equation (64). Indeed, in that case we have $e^{-\int_0^t r_s ds} P(t, T) = \mathbb{E}_t^Q \left[e^{-\int_0^T r_s ds} P(T, T) \right]$, which implies

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

since equations (115)-(119) 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 equation (115). Using the fact that the functions $A(\cdot)$, $B_r(\cdot)$ and $B_\mu(\cdot)$ satisfy the system of ODE:

$$B_r' = a_r B_\mu + 1 \tag{A.21}$$

$$B_\mu' = B_r + a_\mu B_\theta \tag{A.22}$$

$$B_\theta' = B_\mu + a_\theta B_\theta \tag{A.23}$$

$$A' = \frac{\sigma_\mu^0}{2} B_\mu^2 + \frac{\sigma_\theta^0}{2} B_\theta^2 + B_r B_\mu c_{r\mu}^0 + B_r B_\theta c_{r\theta}^0 + B_\theta B_\mu c_{\mu\theta}^0 - B_\theta, \tag{A.24}$$

and that, in particular, because of the restrictions on a_r, a_μ given in equations (89) and (90), we have:

$$B_r = -c_{r\mu}(B_\mu + c_{r\mu} B_\theta) + \sqrt{2B_\mu + 6c_{r\mu} B_\theta}, \tag{A.25}$$

we find that

$$\mathbb{E}^Q [dP(t, T) - r_t P(t, T)] = 0. \tag{A.26}$$

Therefore,

$$\begin{aligned}
e^{-\int_0^t r_s ds} P(t, T) &= \tag{A.27} \\
&\quad - \int_0^t \sqrt{2B_\mu(s) + 6c_{r\mu} B_\theta(s)} \left(\sigma_1 \sqrt{V_s - \psi_1} dZ_1^Q(s) + \sqrt{(1 - \sigma_1^2)V_s + \sigma_1^2 \psi_1 + \psi_3 + \psi_4} dZ_2^Q(s) \right) \\
&\quad - \int_0^t (B_r(s) + \nu_3 B_\mu(s) + \eta_3 B_\theta(s)) \sqrt{-\psi_3} dZ_3^Q(s) - \int_0^t (B_r(s) + \nu_4 B_\mu(s) + \eta_4 B_\theta(s)) \sqrt{-\psi_4} dZ_4^Q(s).
\end{aligned}$$

This shows that $e^{-\int_0^t r_s ds} P(t, T)$ is indeed a Q-martingale.

□

Note that the function $A(\tau)$ can be obtained in closed-form since it is composed of integrals of exponential functions of time. But for conciseness, we leave it in integral form.

B Efficient Importance Sampling

The true density

$$p(\mathbf{V}|\mathbf{P}, \phi)$$

may be decomposed as the product of conditional densities of the form

$$p(V_t|V_{t-1}, \mathbf{P}, \phi) \propto p(\mathcal{P}_{t+1}, \dots, \mathcal{P}_T|V_t, V_{t-1}, \mathcal{P}_1, \dots, \mathcal{P}_t, \phi) p(V_t|V_{t-1}, \mathcal{P}_1, \dots, \mathcal{P}_t, \phi). \quad (\text{B.28})$$

Using the Markovian nature of $\{\mathcal{P}_t, V_t\}$ to eliminate irrelevant conditioning arguments, this reduces to

$$p(\mathcal{P}_{t+1}, \dots, \mathcal{P}_T|V_t, \mathcal{P}_t, \phi) p(V_t|V_{t-1}, \mathcal{P}_{t-1}, \mathcal{P}_t, \phi). \quad (\text{B.29})$$

Note that within the context of the Gaussian quasi-likelihood, the second density is normal. Denote its mean as $\mu_{0,t}$ and its standard deviation as $\sigma_{0,t}$.

Following Liesenfeld and Richard (2002), we construct the importance sampler by replacing the first term, $p(\mathcal{P}_{t+1}, \dots, \mathcal{P}_T|V_t, \mathcal{P}_t, \phi)$, with a Gaussian kernel. Lognormal and inverted gamma densities were also applied, with no improvements in performance. Our importance sampling density is therefore defined as

$$p_t^*(V_t) \propto \exp(a_{1,t}V_t + a_{2,t}V_t^2) \exp\left(-\frac{1}{2}\left(\frac{V_t - \mu_{0,t}}{\sigma_{0,t}}\right)^2\right) \equiv k(V_t, a_t). \quad (\text{B.30})$$

Liesenfeld and Richard (2002) note that the normalized density $p_t^*(V_t)$ is Gaussian with mean

$$\mu_t = \sigma_t^2 \left(\frac{\mu_{0,t}}{\sigma_{0,t}^2} + a_{1,t} \right)$$

and variance

$$\sigma_t^2 = \frac{\sigma_{0,t}^2}{1 - 2\sigma_{0,t}^2 a_{2,t}}.$$

Thus implicitly defining $\chi(V_{t-1}, a_t)$ by

$$\frac{k(V_t, a_t)}{\chi(V_{t-1}, a_t)} = \frac{k(V_t, a_t)}{\int k(V_t, a_t) dV_t} = p_t^*(V_t) = \frac{1}{\sigma_t \sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{V_t - \mu_t}{\sigma_t}\right)^2\right) \quad (\text{B.31})$$

we have

$$\ln \chi(V_{t-1}, a_t) = \frac{1}{2} \ln(\sigma_t^2) - \frac{\mu_{0,t}^2}{2\sigma_{0,t}^2} + \frac{\mu_t^2}{2\sigma_t^2} + \frac{1}{2} \ln(2\pi).$$

The likelihood function is thus the expectation, under the importance sampling density, of

$$\frac{\mathbf{p}(\mathbf{P}, \mathbf{V}|\phi)}{\mathbf{p}^*(\mathbf{V}|\mathbf{P}, \phi)} = \frac{\mathbf{p}(\mathcal{P}_1, V_1)}{m_1(V_1)} \times \prod_{t=2}^T \frac{\mathbf{p}(\mathcal{P}_t, V_t | \mathcal{P}_{t-1}, V_{t-1})}{p_t^*(V_t)} \quad (\text{B.32})$$

$$= \frac{\mathbf{p}(V_1 | \mathcal{P}_1) \mathbf{p}(\mathcal{P}_1)}{p_1^*(V_1)} \times \prod_{t=2}^T \frac{\mathbf{p}(V_t | \mathcal{P}_{t-1}, \mathcal{P}_t, V_{t-1}) \mathbf{p}(\mathcal{P}_t | \mathcal{P}_{t-1}, V_{t-1})}{p_t^*(V_t)} \quad (\text{B.33})$$

$$= \frac{\mathbf{p}(V_1 | \mathcal{P}_1) \mathbf{p}(\mathcal{P}_1) \mathbf{p}(\mathcal{P}_2 | \mathcal{P}_1, V_1)}{p_1^*(V_1)} \times \prod_{t=2}^{T-1} \frac{\mathbf{p}(V_t | \mathcal{P}_{t-1}, \mathcal{P}_t, V_{t-1}) \mathbf{p}(\mathcal{P}_{t+1} | \mathcal{P}_t, V_t)}{p_t^*(V_t)} \times \frac{\mathbf{p}(V_T | \mathcal{P}_{T-1}, \mathcal{P}_T, V_{T-1})}{p_T^*(V_T)} \quad (\text{B.34})$$

$$= \frac{\mathbf{p}(V_1 | \mathcal{P}_1) \mathbf{p}(\mathcal{P}_1) \mathbf{p}(\mathcal{P}_2 | \mathcal{P}_1, V_1)}{k_1(V_1, a_1) / \chi(a_1)} \times \prod_{t=2}^{T-1} \frac{\mathbf{p}(V_t | \mathcal{P}_{t-1}, \mathcal{P}_t, V_{t-1}) \mathbf{p}(\mathcal{P}_{t+1} | \mathcal{P}_t, V_t)}{k_t(V_t, a_t) / \chi(V_{t-1}, a_t)} \times \frac{\mathbf{p}(V_T | \mathcal{P}_{T-1}, \mathcal{P}_T, V_{T-1})}{k_T(V_T, a_T) / \chi(V_{T-1}, a_T)} \quad (\text{B.35})$$

$$= \mathbf{p}(\mathcal{P}_1) \chi(a_1) \frac{\mathbf{p}(V_1 | \mathcal{P}_1) \mathbf{p}(\mathcal{P}_2 | \mathcal{P}_1, V_1)}{k_1(V_1, a_1) / \chi(V_1, a_2)} \times \prod_{t=2}^{T-1} \frac{\mathbf{p}(V_t | \mathcal{P}_{t-1}, \mathcal{P}_t, V_{t-1}) \mathbf{p}(\mathcal{P}_{t+1} | \mathcal{P}_t, V_t)}{k_t(V_t, a_t) / \chi(V_t, a_{t+1})} \times \frac{\mathbf{p}(V_T | \mathcal{P}_{T-1}, \mathcal{P}_T, V_{T-1})}{k_T(V_T, a_T)} \quad (\text{B.36})$$

$$= \mathbf{p}(\mathcal{P}_1) \chi(a_1) \frac{\frac{1}{\sigma_{0,1} \sqrt{2\pi}} \mathbf{p}(\mathcal{P}_2 | \mathcal{P}_1, V_1) \chi(V_1, a_2)}{\exp(a_{1,1} V_1 + a_{2,1} V_1^2)} \times \prod_{t=2}^{T-1} \frac{\frac{1}{\sigma_{0,t} \sqrt{2\pi}} \mathbf{p}(\mathcal{P}_{t+1} | \mathcal{P}_t, V_t) \chi(V_t, a_{t+1})}{\exp(a_{1,t} V_t + a_{2,t} V_t^2)} \times \frac{\frac{1}{\sigma_{0,T} \sqrt{2\pi}}}{\exp(a_{1,T} V_T + a_{2,T} V_T^2)} \quad (\text{B.37})$$

$$= \frac{\mathbf{p}(\mathcal{P}_1) \chi(a_1)}{\sigma_{0,1}} \frac{\frac{1}{\sigma_{0,2} \sqrt{2\pi}} \mathbf{p}(\mathcal{P}_2 | \mathcal{P}_1, V_1) \chi(V_1, a_2)}{\exp(a_{1,1} V_1 + a_{2,1} V_1^2)} \times \prod_{t=2}^{T-1} \frac{\frac{1}{\sigma_{0,t+1} \sqrt{2\pi}} \mathbf{p}(\mathcal{P}_{t+1} | \mathcal{P}_t, V_t) \chi(V_t, a_{t+1})}{\exp(a_{1,t} V_t + a_{2,t} V_t^2)} \times \frac{\frac{1}{\sqrt{2\pi}}}{\exp(a_{1,T} V_T + a_{2,T} V_T^2)} \quad (\text{B.38})$$

$$= \frac{\mathbf{p}(\mathcal{P}_1) \chi(a_1)}{\sigma_{0,1} (2\pi)^{T/2}} \frac{\mathbf{p}(\mathcal{P}_2 | \mathcal{P}_1, V_1) \chi(V_1, a_2)}{\sigma_{0,2} \exp(a_{1,1} V_1 + a_{2,1} V_1^2)} \times \prod_{t=2}^{T-1} \frac{\mathbf{p}(\mathcal{P}_{t+1} | \mathcal{P}_t, V_t) \chi(V_t, a_{t+1})}{\sigma_{0,t+1} \exp(a_{1,t} V_t + a_{2,t} V_t^2)} \times \frac{1}{\exp(a_{1,T} V_T + a_{2,T} V_T^2)} \quad (\text{B.39})$$

In the above calculations, equation (B.34) is simply a rearrangement terms, and (B.35) merely invokes the definitions of p_t^* , k_t , and χ . (B.36) is another rearrangement, while (B.37) results from cancelling out the $\mathbf{p}(V_t | V_{t-1}, \mathcal{P}_{t-1}, \mathcal{P}_t, \phi)$ terms. In (B.38), the $\sigma_{0,t}$ terms are “transferred back” in t because they are functions of V_{t-1} , not V_t . The final equality just simplifies the expressions slightly.

The purpose of this representation is to gain insight as to how $a_{1,t}$ and $a_{2,t}$ should be optimally

selected. Efficient Importance Sampling suggests that the numerator and denominator of each of the ratios in () be as close as possible. We therefore choose a_t to minimize deviations between their logarithms. For $t = T$, we simply choose $a_{1,T} = a_{2,T} = 0$. This sets the last ratio in () equal to a constant.

For $1 \leq t < T$, we choose $a_{1,t}$ and $a_{2,t}$ based on a set of simulated paths of the variance process, to minimize variation in the logarithm of

$$\frac{\mathbf{p}(\mathcal{P}_{t+1}|\mathcal{P}_t, V_t) \chi(V_t, a_{t+1})}{\sigma_{0,t+1} \exp(a_{1,t} V_t + a_{2,t} V_t^2)}.$$

Given the simulated V_t , we do so by regressing

$$\ln \mathbf{p}(\mathcal{P}_{t+1}|\mathcal{P}_t, V_t) + \ln \chi(V_t, a_{t+1}) - \ln \sigma_{0,t+1} \quad (\text{B.40})$$

on V_t and V_t^2 . $a_{1,t}$ and $a_{2,t}$ are then set equal to the two slope coefficients. Finally, these values are used to construct $\chi(V_{t-1}, a_t)$, which are then used in the time $t-1$ regression to compute $a_{1,t-1}$ and $a_{2,t-1}$. The process is repeated until time 1.

In practice, the EIS algorithm is initiated by simulating from $\mathbf{p}_t^*(V_t)$ based on some guess of the a_t . In the likelihood maximization process, very good guesses are typically obtained from the values used in the previous evaluation of the likelihood function. Given this first set of simulated V_t , new a_t are calculated using the procedure above, and the results generally converge after three or four iterations.

The performance of the sampler is analyzed in Figure A1. The top three panels plot recursive estimates of the maximized likelihood for three models that require the use of importance sampling. The bottom panels replicate Shephard and Koopman's (2002) tests of the ξ parameter, which measures the tail thickness of the distribution of

$$\frac{\mathbf{p}(\mathbf{P}, \mathbf{V}|\phi)}{\mathbf{p}^*(\mathbf{V}|\mathbf{P}, \phi)},$$

as well as its 95% critical value. Values of ξ above .5 indicate a distribution with no finite variance.

From Figure A1, we see that the $A_1(3)$ USV model is clearly the worst performer of the three. Likelihoods do not appear to stabilize completely at even 10,000 simulations, though the instability in the likelihood seems unlikely to explain the large differences in likelihoods between this model and the others. Furthermore, the values of ξ in the lower panel indicate that the above variance likely does not exist. While this does not strictly invalidate the use of importance sampling, it makes inferences about this model somewhat unreliable, and it is possible that the parameter vector resulting from the maximization of the simulated likelihood function is not close to the true MLE estimate. It is possible, therefore, that the performance of the $A_1(3)$ USV model might improve given a better estimator.

For the $A_1(3)$ model, in the middle panel, the likelihood stabilizes almost immediately and there is no cause for concern. The sampler performance for the $A_1(4)$ USV model is somewhat worse, though likelihoods still converge reasonably well and ξ is well below its critical value.

Table 1**Observability of state variables**

The table contains output from the regressions

$$\begin{aligned} \text{true } r_t &= \alpha^r + \beta^r \times \text{estimated } r_t + \epsilon_t^r \\ \text{true } \mu_t^Q &= \alpha^\mu + \beta^\mu \times \text{estimated } \mu_t^Q + \epsilon_t^\mu, \end{aligned}$$

where r_t is the instantaneous short rate and μ_t^Q is its drift under the risk-neutral measure. Ten-year samples of weekly short rate data are simulated from the two-factor CIR model $dx_{i,t} = \kappa_i(\theta_i - x_{i,t})dt + \sigma_i\sqrt{x_{i,t}}dz_{i,t}$, $r_t = x_{1,t} + x_{2,t}$, with parameter values from Table I of Duffie and Singleton (1997). Zero coupon yields with maturities $\tau = \{.5, 1, 2, 5, 7, 10\}$ years are computed under the risk-neutralized process $dx_{i,t} = [\kappa_i(\theta_i - x_{i,t}) - \lambda_i x_{i,t}]dt + \sigma_i\sqrt{x_{i,t}}dz_{i,t}^Q$, and then modified by adding i.i.d. measurement errors with standard deviations of either 2 or 5 basis points. Quadratic and cubic polynomials in τ are used to fit these yields by OLS. The value of the polynomial at zero and twice the value of its slope at zero are taken as estimates of r_t and μ_t^Q , respectively. Numbers in the table are means and standard deviations (in parentheses) from 5000 simulated data samples.

	2 b.p. measurement error		5 b.p. measurement error	
	quadratic	cubic	quadratic	cubic
Instantaneous Short Rate				
$\alpha^r \times 100$	-0.303 (0.292)	-0.074 (0.069)	-0.299 (0.286)	-0.064 (0.059)
β^r	1.033 (0.017)	1.008 (0.004)	1.032 (0.017)	1.005 (0.005)
R^2	0.999 (0.001)	0.999 (0.000)	0.998 (0.002)	0.997 (0.002)
Short Rate Drift				
$\alpha^\mu \times 100$	-0.042 (0.008)	0.024 (0.020)	-0.013 (0.023)	0.155 (0.088)
β^μ	1.631 (0.006)	1.129 (0.014)	1.599 (0.022)	1.026 (0.058)
R^2	0.996 (0.002)	0.980 (0.010)	0.976 (0.012)	0.890 (0.049)

Table 2**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 2001. 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	0.09	3.15	1.81	7.09	9.36	26.42
1-year	0.12	2.60	0.34	-6.64	-16.14	-89.59
2-year	0.14	1.04	-1.44	-5.93	7.09	168.50
3-year	0.14	-0.03	-1.49	0.62	8.50	-52.72
4-year	0.14	-0.70	-0.94	4.63	0.30	-123.67
5-year	0.13	-1.16	-0.27	5.77	-7.21	-38.92
7-year	0.13	-1.73	0.94	2.66	-10.44	187.29
10-year	0.12	-2.17	2.05	-7.20	9.55	-76.30
% explained	63.92	18.38	8.36	4.51	3.36	1.20

Total % explained by first six principal components: 99.74

Table 3**Likelihood analysis of various models**

For Panels A and B of the table, likelihoods were maximized over the 1988 to 2001 sample from a dataset consisting of six principal components of zero-coupon yields bootstrapped from swap and LIBOR rates. Panel A reports statistics on essentially affine specifications, while Panel B describes completely affine models. Out-of-sample likelihoods use the parameters estimated over that period to compute a likelihood for the 105 weekly observations of 2002 and 2003. The Akaike and Bayesian Information Criteria (AIC) and (BIC) are calculated as $-L + N$ and $-L + .5N \ln T$, respectively, where L is the sample log likelihood, N is the number of model parameters, and $T = 729$ is the number of weeks in the sample. Likelihood ratio statistics are used to test the restricted $A_1(2)$, $A_1(3)$ USV, and $A_1(3)$ 3PC models against the more general $A_1(3)$ specification. P-values are in parentheses. Finally, these panels report whether the Feller constraint was binding under the P and Q measures. Note that volatility drift parameters under the risk-neutral measure are unidentified for USV specifications.

Panel C of the table contains likelihood ratio statistics corresponding to the test of each completely affine specification against its more general essentially affine counterpart.

Panel D again reports results for the completely affine specifications, but models are now estimated using data only on the first two principal components. Statistics are therefore not comparable to those from other panels of the table.

Panel E summarizes the number of parameters estimated in each specification. All models require the number of risk neutral parameters specified plus those parameters corresponding to the type of risk premia used. Models estimated using all six principal components also require the number of measurement error standard deviations given.

	$A_1(2)$	$A_1(3)$ USV	$A_1(3)$	$A_1(3)$ 3PC	$A_1(4)$ USV
Panel A: Essentially affine specifications estimated using all 6 principal components					
Sample log likelihood	14118.26	14742.08	16193.24	16103.89	16249.97
Out-of-sample log likelihood	1768.35	2080.07	2046.72	2111.28	2221.96
AIC	-14101.26	-14721.08	-16165.24	-16076.89	-16223.97
BIC	-14062.23	-14672.86	-16100.95	-16014.91	-16164.27
LRT relative to $A_1(3)$	4149.96	2902.32		178.69	
LRT p-value	(0.0000)	(0.0000)		(0.0000)	
Feller binding under $P/Q?$	no/yes	no/NA	no/no	no/yes	no/NA
Panel B: Completely affine specifications estimated using all 6 principal components					
Sample log likelihood	13942.18	14709.27	15882.56	15867.19	16242.38
Out-of-sample log likelihood	1835.94	2066.84	1996.32	2081.28	2213.97
AIC	-13928.18	-14693.27	-15861.56	-15847.19	-16221.38
BIC	-13896.04	-14656.54	-15813.34	-15801.28	-16173.17
LRT relative to $A_1(3)$	3880.75	2346.57		30.73	
LRT p-value	(0.0000)	(0.0000)		(0.0000)	
Feller binding under $P/Q?$	yes/yes	no/no	yes/yes	yes/yes	no/no
Panel C: Essentially versus completely affine LR tests					
LRT of EA vs CA	352.15	45.53	621.36	473.40	15.17
LRT p-value	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0097)
Panel D: Completely affine specifications estimated using only 2 principal components					
Sample log likelihood	6487.34	6612.18	6616.60		
Out-of-sample log likelihood	914.87	927.67	920.37		
AIC	-6477.34	-6600.18	-6599.60		
BIC	-6454.38	-6572.63	-6560.57		
LRT relative to $A_1(3)$	258.51	8.84			
LRT p-value	(0.0000)	(0.1155)			
Feller binding under $P/Q?$	no/no	no/no	yes/yes		
Panel E: Numbers of model parameters					
risk-neutral	8	9	14	14	14
essentially affine risk premia	5	8	10	10	9
completely affine risk premia	2	3	3	3	4
measurement error std. devs.	4	4	4	3	3

Table 4A
Parameter estimates

Parameter estimates and standard errors are calculated by quasi-maximum likelihood from weekly bootstrapped yields from January 1988 to December 2001.

	$A_1(2)$	$A_1(3)$ USV	$A_1(3)$	$A_1(4)$ USV
σ_1	0.0677 (0.0333)	0.2454 (0.0411)	-0.1305 (0.0501)	-0.1631 (0.0446)
σ_2		0 NA	0.9898 (0.0183)	0.9866 (0.0074)
σ_3		0.9694 (0.0104)	0.0567 (0.3054)	0 NA
ν_1		-0.0274 (0.0046)	6.6379 (4.1839)	0.0842 (0.0230)
ν_2		-84.4499 (2184.1561)	-3.7288 (0.8022)	-0.5160 (0.0063)
ν_3		-0.1116 (0.0019)	3.7282 (3.2253)	-430.3842 (2377.8647)
ν_4				-76.7401 (8.4888)
$\sigma_V \times 10^3$	0.9942 (0.1404)	8.1198 (0.2545)	1.8390 (0.2764)	5.0652 (0.1936)
$\psi_1 \times 10^5$	0.0002 (2.1015)	0.0112 (0.0357)	2.3310 (1.1669)	0.0471 (0.0331)
$\psi_2 \times 10^5$		-0.0004 (0.0022)	2.2838 (1.8648)	0.0082 (0.0024)
$\psi_3 \times 10^5$		0.0002 (0.0033)	-2.3235 (1.8699)	0.0001 (0.0001)
$\psi_4 \times 10^5$				-0.0094 (0.0023)
κ_V^P	0.1559 (0.0980)	0.4650 (0.1655)	0.0071 (0.1667)	0.2527 (0.0301)
$\gamma_V^P \times 10^4$	0.2694 (0.1572)	0.4810 (0.0706)	0.0192 (0.2275)	0.2540 (0.0239)
$m_0 \times 10^2$		-0.1513 (0.0085)	-15.3115 (1.1706)	
$m_r \times 10^2$		-2.4899 (0.0853)	-104.9432 (3.6014)	
m_μ		-0.3347 (0.0057)	-2.3793 (0.0585)	
m_V		1 NA	-521.8524 (101.9732)	
Additional parameters for the $A_1(2)$ model				
	γ_r	κ_{rr}	κ_{rV}	
	0.0102 (0.0068)	0.3297 (0.0069)	-247.4986 (46.2590)	
Additional parameters for the $A_1(4)$ USV model				
$a_0 \times 10^4$	a_θ	η_3	η_4	$c_{r\mu}$
-1.3742 (2.5943)	-1.5576 (0.0118)	-714.8040 (122.3995)	113.8364 (13.4077)	-0.5160 (0.0063)

Table 4B**Parameter estimates, cont.**

Parameter estimates and standard errors are calculated by quasi-maximum likelihood from weekly bootstrapped yields from January 1988 to December 2001.

	$A_1(2)$	$A_1(3)$ USV	$A_1(3)$	$A_1(4)$ USV
λ_{r0}	-0.0768 (0.0400)	-0.0289 (0.0061)	-0.0382 (0.0364)	-0.0066 (0.0013)
λ_{rr}	0.3022 (0.2176)	0.4265 (0.0941)	0.3681 (0.2323)	
$\lambda_{r\mu}$		1.6801 (0.2838)	-0.1711 (0.1816)	-0.0472 (0.1340)
λ_{rV}	170.7519 (263.8621)	-244.5913 (35.4346)	73.9945 (202.5647)	2.5103 (37.9357)
$\lambda_{\mu0}$		0.0224 (0.0078)	0.2550 (0.1222)	-0.0152 (0.0065)
$\lambda_{\mu r}$		-0.3143 (0.1017)	-0.3266 (0.8028)	
$\lambda_{\mu\mu}$		-0.7552 (0.3365)	0.3485 (0.6648)	-0.5597 (0.5148)
$\lambda_{\mu V}$		93.3392 (23.4874)	248.4142 (673.2265)	-1.1103 (65.2360)
$\lambda_{\theta0}$				0.0252 (0.0095)
$\lambda_{\theta\mu}$				0.9195 (0.7437)
$\lambda_{\theta V}$				-2.7537 (91.4982)
$\lambda_{V0} \times 10^4$	0.2644 (0.1571)		0.6549 (0.2274)	
λ_{VV}	-0.1559 (0.0980)		-0.3672 (0.1668)	

Table 5
In-sample yield fits

This table contains statistics on the in-sample fits of zero coupon yields (Y). For each model, fitted yields (\hat{Y}_t) are calculated for .5, 1, 2, 3, 4, 5, 7, and 10-year maturities. The table examines the bias, root mean squared error, and autocorrelation of $\hat{e}_t = Y_t - \hat{Y}_t$, 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). 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 729 weeks.

	$A_1(2)$		$A_1(3)$ USV		$A_1(3)$		$A_1(4)$ USV
	mean \hat{e} (basis points)						
6-month	0.81		2.07		-0.37		0.76
1-year	-0.49		0.01		0.27		-0.64
2-year	-1.08		-1.53		0.45		-0.79
3-year	-0.45		-1.48		0.07		-0.01
4-year	0.50		-0.81		-0.28		0.61
5-year	1.29*		-0.05		-0.47		0.80*
7-year	1.55*		1.08		-0.32		0.22
10-year	-2.00		1.73		0.59		-0.84
	RMSE (basis points)						
6-month	13.86	>*	13.09	>**	3.59	<*	4.40
1-year	5.18		5.40	>**	4.00	<*	4.75
2-year	9.47	>*	9.09	>**	2.71		2.93
3-year	8.32		7.95	>**	1.28		1.27
4-year	6.05		5.88	>**	2.22		2.49
5-year	4.11		4.15	>*	2.74		2.85
7-year	5.53	>*	5.34	>**	1.82		1.65
10-year	13.73		12.92	>**	3.65		3.58
	autocorrelation of \hat{e}						
6-month	0.96		0.96		0.92		0.95
1-year	0.94		0.95		0.90		0.93
2-year	0.95		0.95		0.87		0.90
3-year	0.95		0.96		0.62		0.85
4-year	0.96		0.97		0.88		0.94
5-year	0.95		0.96		0.90		0.91
7-year	0.93		0.93		0.79		0.80
10-year	0.96		0.97		0.87		0.91

Table 6**Out-of-sample yield fits**

This table contains statistics on the in-sample fits of zero coupon yields (Y). For each model, fitted yields (\hat{Y}_t) are calculated for .5, 1, 2, 3, 4, 5, 7, and 10-year maturities. The table examines the bias, root mean squared error, and autocorrelation of $\hat{e}_t = Y_t - \hat{Y}_t$, 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). 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 105 weeks.

	$A_1(2)$	$A_1(3)$ USV		$A_1(3)$		$A_1(4)$ USV
mean \hat{e} (basis points)						
6-month	-8.44	0.52		5.40**		2.04*
1-year	0.22	3.30**		-4.36**		-0.85
2-year	3.82	-0.62		-6.08**		-3.85**
3-year	5.00	-2.87		-0.20		-1.03**
4-year	5.11	-3.21		4.32**		2.29**
5-year	4.04**	-2.57*		5.96**		3.98**
7-year	-1.14	0.14		2.99**		2.39**
10-year	-13.10*	6.73		-7.24**		-4.64**
RMSE (basis points)						
6-month	19.77	15.94	>*	8.60	>**	5.23
1-year	5.37	5.96	<*	8.07	>**	5.00
2-year	12.34	10.88		7.95	>**	5.20
3-year	14.72	12.34	>**	1.30		1.43
4-year	12.13	9.49		6.12	>**	3.65
5-year	7.23	5.08		7.83	>**	5.31
7-year	7.35	7.12	>*	3.64	>**	2.94
10-year	27.36	20.13		9.68	>**	6.38
autocorrelation of \hat{e}						
6-month	0.92	0.91		0.82		0.70
1-year	0.73	0.69		0.82		0.68
2-year	0.91	0.91		0.85		0.73
3-year	0.95	0.94		0.68		0.64
4-year	0.94	0.93		0.86		0.70
5-year	0.89	0.83		0.87		0.75
7-year	0.93	0.93		0.87		0.84
10-year	0.95	0.94		0.87		0.76

Table 7**Correlations of observed and model-implied time series**

This table reports correlations between actual and model-implied series. Average yield is simply the average of the .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. GARCH(1,1) and EGARCH(1,1) volatilities are calculated from demeaned changes in the six-month rate. Interpolated r , μ^Q , and θ are calculated using a third-order polynomial regression of yields on maturity. The Implied Volatility series, obtained from one-year cap and floor contracts, is an average of Black-Scholes implied volatilities on the logarithm of the one-year LIBOR rate. Multiplying the implied volatilities by the level of the one-year rate yields an approximate volatility of the level of one-year LIBOR. Note that the Implied Volatility series is available only starting in 1995.

	$A_1(2)$	$A_1(3)$ USV	$A_1(3)$	$A_1(4)$ USV	EGARCH (1,1)	GARCH (1,1)	Implied Volatility	Imp. Vol. $\times Y_1$
Full Sample								
Actual vs. model average yield	1.000	1.000	1.000	1.000				
Actual vs. model slope	0.998	0.998	0.998	0.998				
Actual vs. model curvature	0.006	0.357	0.999	1.000				
EGARCH vs. model volatility	0.669	0.862	-0.693	0.875				
GARCH vs. model volatility	0.697	0.905	-0.704	0.917	0.951			
Interpolated vs. model r	0.996	0.996	0.986	0.997				
Interpolated vs. model μ^Q		0.880	0.611	0.884				
Interpolated vs. model θ				0.845				
Actual curvature vs. model volatility	-0.400	-0.066	0.253	-0.101	-0.067	-0.111		
Actual curvature vs. model variance	-0.395	-0.062	0.262	-0.101	-0.040	-0.103		
1988 to 1994								
Actual vs. model average yield	1.000	1.000	1.000	1.000				
Actual vs. model slope	0.998	0.998	0.998	0.998				
Actual vs. model curvature	0.179	0.373	0.998	1.000				
EGARCH vs. model volatility	0.467	0.743	-0.542	0.747				
GARCH vs. model volatility	0.545	0.809	-0.603	0.807	0.925			
Interpolated vs. model r	0.997	0.997	0.990	0.998				
Interpolated vs. model μ^Q		0.864	0.485	0.845				
Interpolated vs. model θ				0.815				
Actual curvature vs. model volatility	-0.471	-0.159	0.269	-0.196	-0.101	-0.211		
Actual curvature vs. model variance	-0.471	-0.103	0.275	-0.149	-0.031	-0.159		
1995 to 2001								
Actual vs. model average yield	1.000	1.000	1.000	1.000				
Actual vs. model slope	0.998	0.996	0.999	0.998				
Actual vs. model curvature	-0.269	0.415	0.999	1.000				
EGARCH vs. model volatility	0.209	0.871	-0.086	0.883				
GARCH vs. model volatility	0.204	0.878	-0.068	0.894	0.946			
Imp. vol. vs. model volatility	0.072	0.742	0.057	0.742	0.843	0.857		
Imp. vol. $\times Y_1$ vs. model volatility	0.279	0.385	-0.378	0.392	0.540	0.484	0.592	
Interpolated vs. model r	0.987	0.990	0.964	0.992				
Interpolated vs. model μ^Q		0.892	0.820	0.944				
Interpolated vs. model θ				0.943				
Actual curvature vs. model volatility	-0.786	0.052	0.619	-0.001	0.021	-0.002	0.157	0.109
Actual curvature vs. model variance	-0.790	0.016	0.615	-0.028	-0.005	-0.020	0.067	0.105

Table 8**In-sample yield forecasts**

This table contains statistics on the in-sample one-week forecasts of zero coupon yield changes. For each model, expected yield changes are calculated for .5, 1, 2, 3, 4, 5, 7, and 10-year maturities as differences between the model expectations of future yields and the current model fitted values. The table examines the bias, root mean squared error, and autocorrelation of $\hat{e}_{t+1} = (Y_{t+1} - Y_t) - (E_t[Y_{t+1}] - \hat{Y}_t)$, where $E_t[Y_{t+1}]$ is the model-implied expectation and \hat{Y}_t is 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). 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 625 weeks.

	$A_1(2)$	$A_1(3)$ USV	$A_1(3)$	$A_1(4)$ USV
mean \hat{e} (basis points)				
6-month	-0.24	-0.92	-0.30	-0.65
1-year	-0.25	-0.84	0.35	-0.03
2-year	-0.24	-0.68	0.89	0.66
3-year	-0.26	-0.58	1.03	0.92
4-year	-0.31	-0.52	1.04	0.99
5-year	-0.35	-0.49	1.01	0.99
7-year	-0.44	-0.44	0.95	0.92
10-year	-0.52	-0.41	0.89	0.81
RMSE (basis points)				
6-month	11.33	>* 11.08	11.12	11.19
1-year	13.34	13.18	13.20	13.25
2-year	14.76	14.66	14.68	14.73
3-year	14.93	14.85	14.86	14.91
4-year	14.72	14.66	14.67	14.72
5-year	14.48	14.42	14.43	14.48
7-year	14.12	14.09	14.08	14.13
10-year	14.05	14.03	14.02	14.06
autocorrelation of \hat{e}				
6-month	0.04	-0.01	-0.02	0.02
1-year	0.01	-0.02	-0.03	0.00
2-year	0.01	-0.01	0.00	0.01
3-year	-0.01	-0.02	-0.02	0.00
4-year	-0.03	-0.04	-0.03	-0.02
5-year	-0.04	-0.05	-0.05	-0.04
7-year	-0.07	-0.07	-0.06	-0.06
10-year	-0.08	-0.08	-0.08	-0.08

Table 9**Out-of-sample yield forecasts**

This table contains statistics on the in-sample one-week forecasts of zero coupon yield changes. For each model, expected yield changes are calculated for .5, 1, 2, 3, 4, 5, 7, and 10-year maturities as differences between the model expectations of future yields and the current model fitted values. The table examines the bias, root mean squared error, and autocorrelation of $\hat{e}_{t+1} = (Y_{t+1} - Y_t) - (E_t[Y_{t+1}] - \hat{Y}_t)$, where $E_t[Y_{t+1}]$ is the model-implied expectation and \hat{Y}_t is 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). 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 105 weeks.

	$A_1(2)$	$A_1(3)$ USV	$A_1(3)$	$A_1(4)$ USV
	mean \hat{e} (basis points)			
6-month	1.16	-1.90**	2.13**	-0.75
1-year	0.90	-2.14*	1.21	-0.63
2-year	0.03	-2.97*	-0.15	-0.91
3-year	-0.51	-3.47*	-0.79	-1.04
4-year	-0.81	-3.71*	-1.04	-1.01
5-year	-0.96	-3.81*	-1.12	-0.91
7-year	-1.08	-3.78*	-1.07	-0.66
10-year	-1.08	-3.54*	-0.88	-0.33
	RMSE (basis points)			
6-month	6.70	6.82	7.31	6.79
1-year	12.07	12.23	12.25	12.13
2-year	15.72	15.95	15.71	15.70
3-year	16.68	16.96	16.61	16.61
4-year	17.35	17.65	17.26	17.25
5-year	17.71	18.02	17.61	17.61
7-year	17.67	17.99	17.58	17.59
10-year	17.14	17.43	17.05	17.08
	autocorrelation of \hat{e}			
6-month	0.08	0.07	0.18	0.13
1-year	-0.18	-0.19	-0.16	-0.17
2-year	-0.12	-0.13	-0.12	-0.12
3-year	-0.05	-0.06	-0.05	-0.05
4-year	-0.02	-0.04	-0.03	-0.03
5-year	-0.02	-0.03	-0.03	-0.03
7-year	-0.04	-0.05	-0.04	-0.04
10-year	-0.08	-0.08	-0.08	-0.08

Table 10
In-sample volatility forecasts

This table contains statistics on in-sample one-week forecasts of different volatility proxies. For each model, expected squared yield changes ($E[\Delta Y^2]$) and expected “realized volatility” ($E[\hat{\sigma}]$) are calculated for .5, 1, 2, 3, 4, 5, 7, and 10-year maturities. Realized volatility is defined as $\hat{\sigma}_{t,\tau}^2 = \frac{1}{5} \sum_{i=1}^5 \Delta Y(t, i, \tau)^2$, and is calculated using daily data. The table examines the forecast bias (actual minus forecast) and root mean squared error of ΔY^2 and $\hat{\sigma}$, where all yields are expressed in basis points. * and ** denote statistical significance at the 5% and 1% levels, respectively, where standard errors are calculated using the method of Newey and West (1987). 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 625 weeks.

	$A_1(2)$		$A_1(3)$ USV		$A_1(3)$		$A_1(4)$ USV
bias in weekly ΔY^2							
6-month	-126.44**		-19.78		-41.24*		-25.08
1-year	-44.48*		28.82		-16.20		-13.17
2-year	41.05*		56.40**		-9.58		-22.98
3-year	74.25**		42.65**		2.68		-23.87
4-year	85.67**		17.32		11.10		-17.74
5-year	88.79**		-7.22		15.07		-10.20
7-year	87.59**		-41.08**		16.44		1.54
10-year	86.50**		-51.45**		17.49		14.58
RMSE of weekly ΔY^2							
6-month	450.91	> **	433.47		438.94		432.39
1-year	464.67		463.33		466.81	> *	461.40
2-year	426.59		427.50		429.47	> *	424.19
3-year	402.55		397.75		399.95		396.23
4-year	382.06		373.48		375.93		373.11
5-year	367.38		357.76		359.60		357.11
7-year	349.64		341.99		341.32		338.88
10-year	335.73		328.26		328.41	> *	325.18
bias in $\hat{\sigma}$							
6-month	-3.56**		-1.56**		-2.29**		-1.73**
1-year	-1.93**		-0.44*		-1.51**		-1.33**
2-year	-0.04		0.44*		-0.82**		-0.96**
3-year	0.74**		0.37		-0.41		-0.74**
4-year	1.04**		0.01		-0.21		-0.60**
5-year	1.15**		-0.36		-0.12		-0.48*
7-year	1.20**		-0.85**		-0.06		-0.28
10-year	1.28**		-0.93**		0.06		0.02
RMSE of $\hat{\sigma}$							
6-month	4.29	> **	2.90	< *	3.42	> *	2.93
1-year	3.25	> **	2.76	< *	3.10		2.93
2-year	2.96		2.99		3.16		3.08
3-year	3.09		3.01		3.12		3.08
4-year	3.09		2.92		2.99		2.97
5-year	3.04		2.86		2.87		2.85
7-year	2.94		2.84		2.74		2.69
10-year	2.92		2.78		2.72	> **	2.63

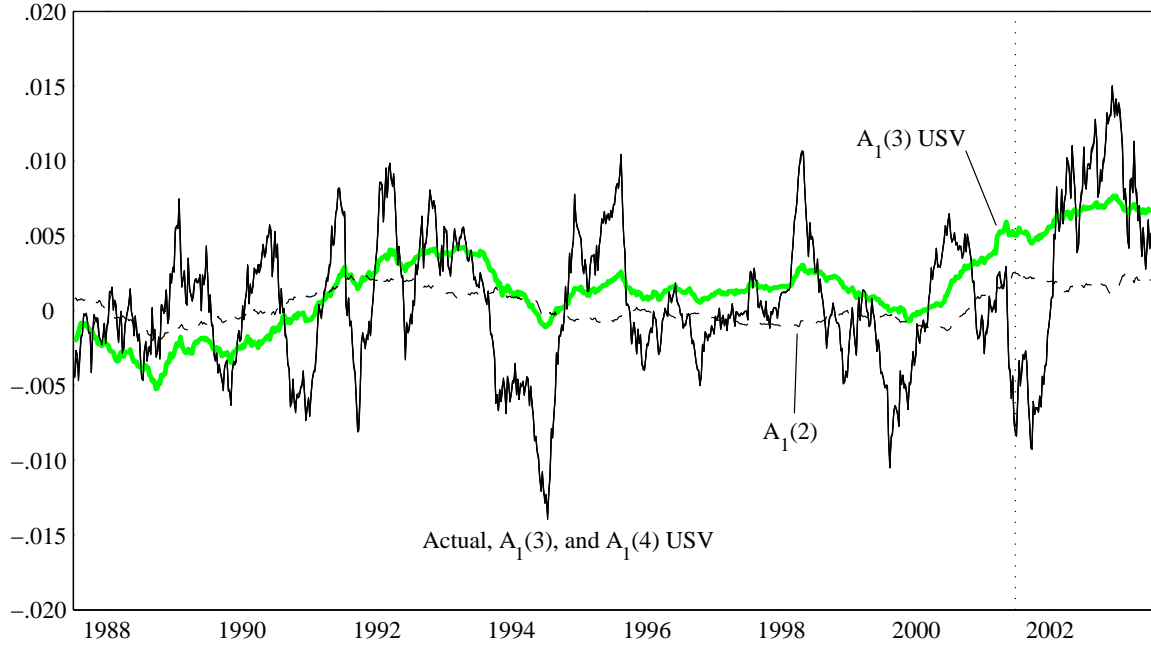
Table 11**Out-of-sample volatility forecasts**

This table contains statistics on in-sample one-week forecasts of different volatility proxies. For each model, expected squared yield changes ($E[\Delta Y^2]$) and expected “realized volatility” ($E[\hat{\sigma}]$) are calculated for .5, 1, 2, 3, 4, 5, 7, and 10-year maturities. Realized volatility is defined as $\hat{\sigma}_{t,\tau}^2 = \frac{1}{5} \sum_{i=1}^5 \Delta Y(t, i, \tau)^2$, and is calculated using daily data. The table examines the forecast bias (actual minus forecast) and root mean squared error of ΔY^2 and $\hat{\sigma}$, where all yields are expressed in basis points. * and ** denote statistical significance at the 5% and 1% levels, respectively, where standard errors are calculated using the method of Newey and West (1987). 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 105 weeks.

	$A_1(2)$		$A_1(3)$ USV		$A_1(3)$		$A_1(4)$ USV
bias in weekly ΔY^2							
6-month	-180.75**		-71.11**		-150.98**		-101.45**
1-year	-51.17*		26.61		-67.53**		-37.63
2-year	91.68**		112.62**		-5.99		14.20
3-year	148.09**		121.63**		26.94		38.16
4-year	186.13**		121.62**		60.57		71.45
5-year	207.67**		113.85**		81.83*		97.49*
7-year	214.23**		84.20*		88.68*		116.57**
10-year	196.53**		52.72		69.79*		112.09**
RMSE of weekly ΔY^2							
6-month	192.31	>**	109.59	<**	165.29	>**	131.85
1-year	235.70		233.56		244.18		236.81
2-year	367.68	<**	375.36		359.04		359.77
3-year	399.07		393.19	>*	372.58		375.36
4-year	428.49	>**	408.36	>*	390.30		394.66
5-year	445.38	>**	413.75	>*	401.11	<*	407.38
7-year	435.10	>**	390.47		386.46	<**	396.70
10-year	402.74	>**	356.19		355.34	<**	368.63
bias in $\hat{\sigma}$							
6-month	-3.68**		-1.68**		-3.11**		-2.35**
1-year	-1.33**		0.18		-1.52**		-1.12**
2-year	1.12**		1.63**		-0.36		-0.10
3-year	2.23**		1.87**		0.31		0.46
4-year	2.89**		1.85**		0.83*		0.98**
5-year	3.23**		1.69**		1.11**		1.34**
7-year	3.39**		1.26**		1.23**		1.65**
10-year	3.18**		0.84*		1.00**		1.65**
RMSE of $\hat{\sigma}$							
6-month	3.94	>**	2.36	<**	3.44	>**	2.87
1-year	2.55		2.24	<**	2.71	>**	2.46
2-year	2.83	<**	3.13	>*	2.69		2.62
3-year	3.50		3.38	>**	2.78		2.78
4-year	4.04	>**	3.48	>**	2.98		3.02
5-year	4.36	>**	3.48	>**	3.15	<*	3.24
7-year	4.51	>**	3.31		3.22	<**	3.41
10-year	4.40	>**	3.19		3.20	<**	3.45

Figure 1

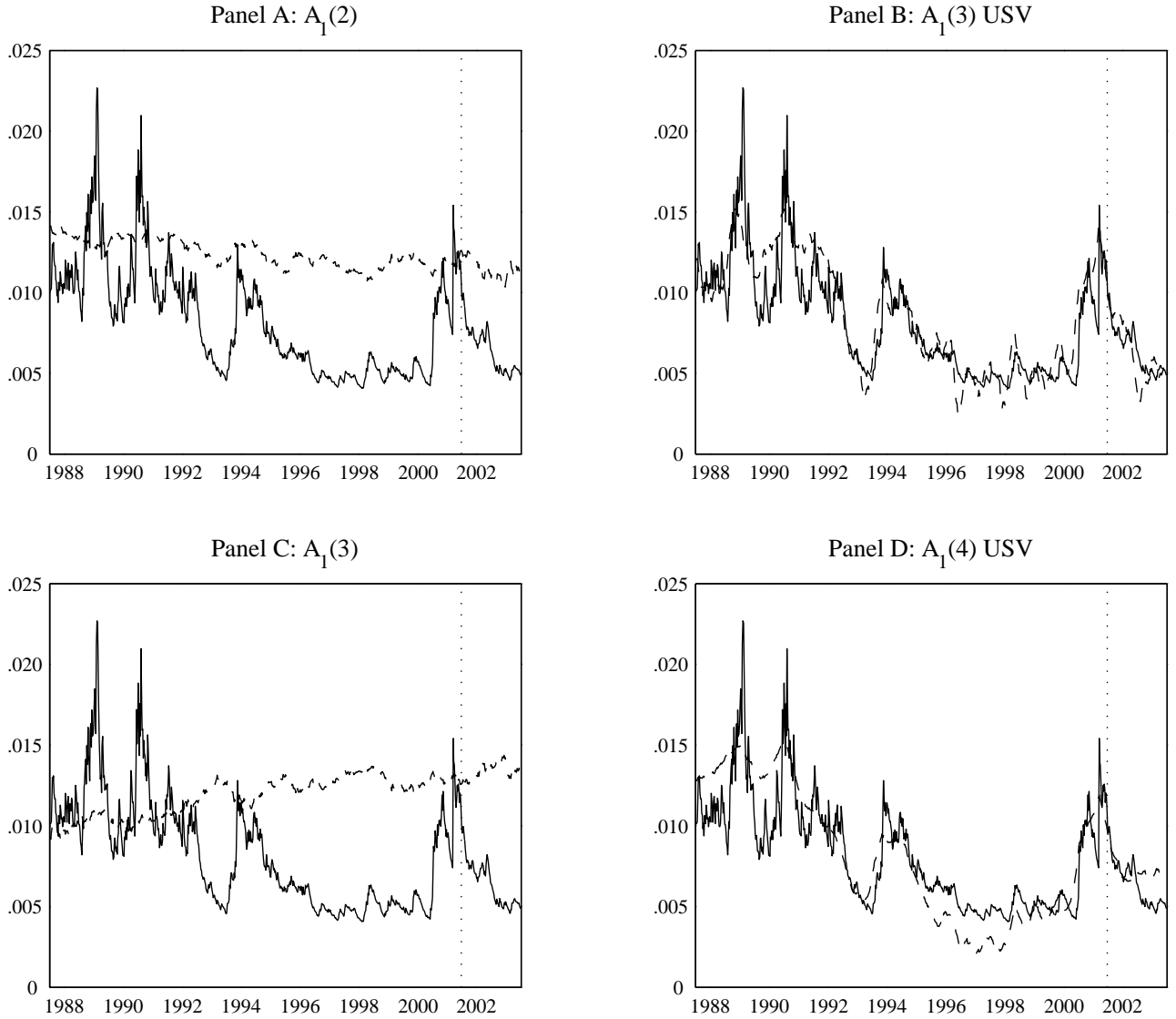
Actual and model-implied curvature



Actual curvature, depicted by the solid black line, is defined as $Y_{10y} - 2Y_{3y} + Y_{6m}$. Model implied curvature is calculated using smoothed estimates of the model state variables. For the $A_1(3)$ and $A_1(4)$ USV models, fitted curvatures are almost indistinguishable from the actual. The vertical dotted line denotes the end of the estimation period.

Figure 2

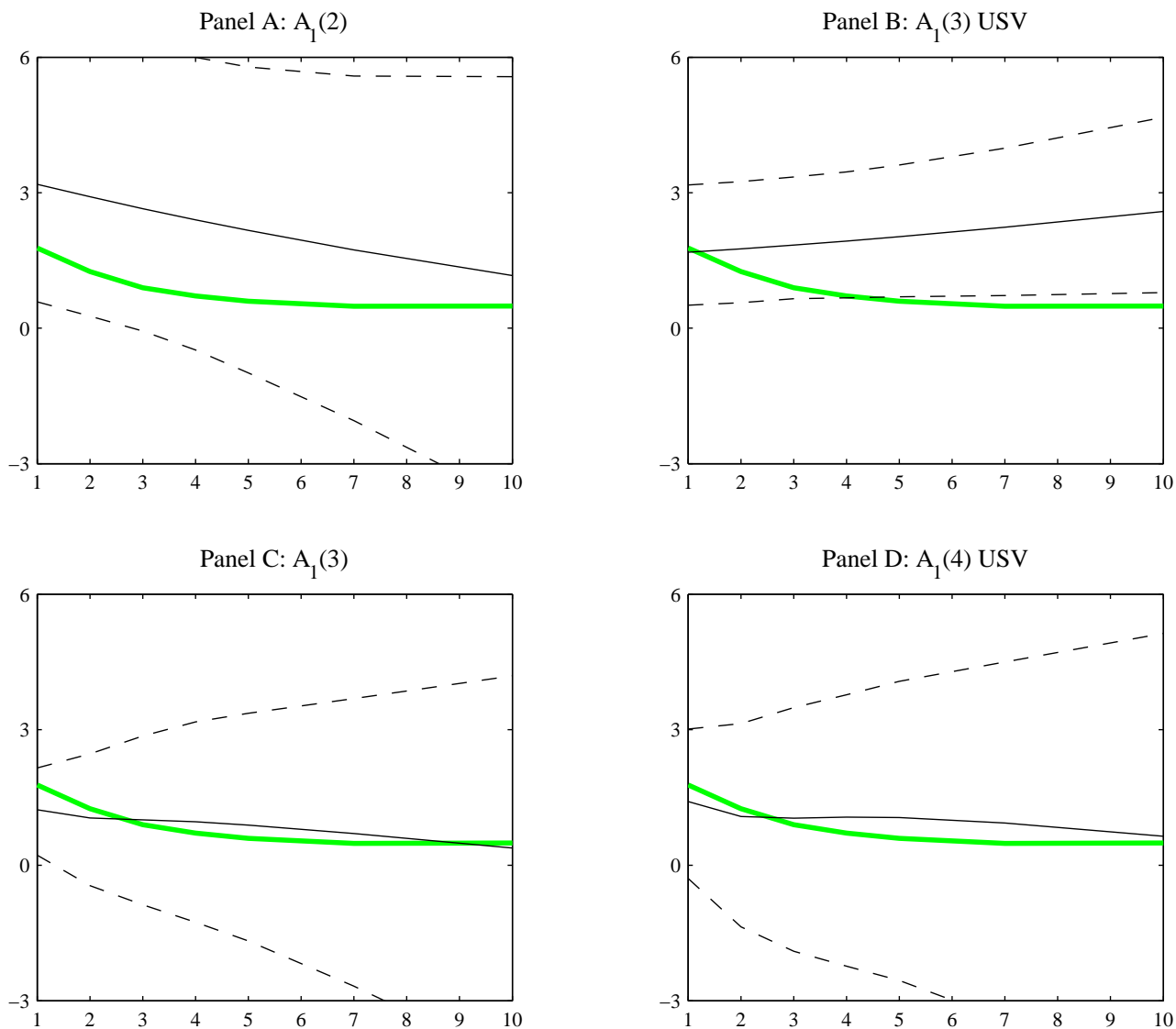
EGARCH and model-implied short rate volatility



In each panel, the solid line depicts the fitted path of the volatility of the 6-month yield that is implied by an EGARCH(1,1) model. The dashed lines correspond to smoothed estimates of instantaneous volatility implied by each of the affine specifications in Table 4. The vertical dotted lines denote the end of the estimation period.

Figure 3

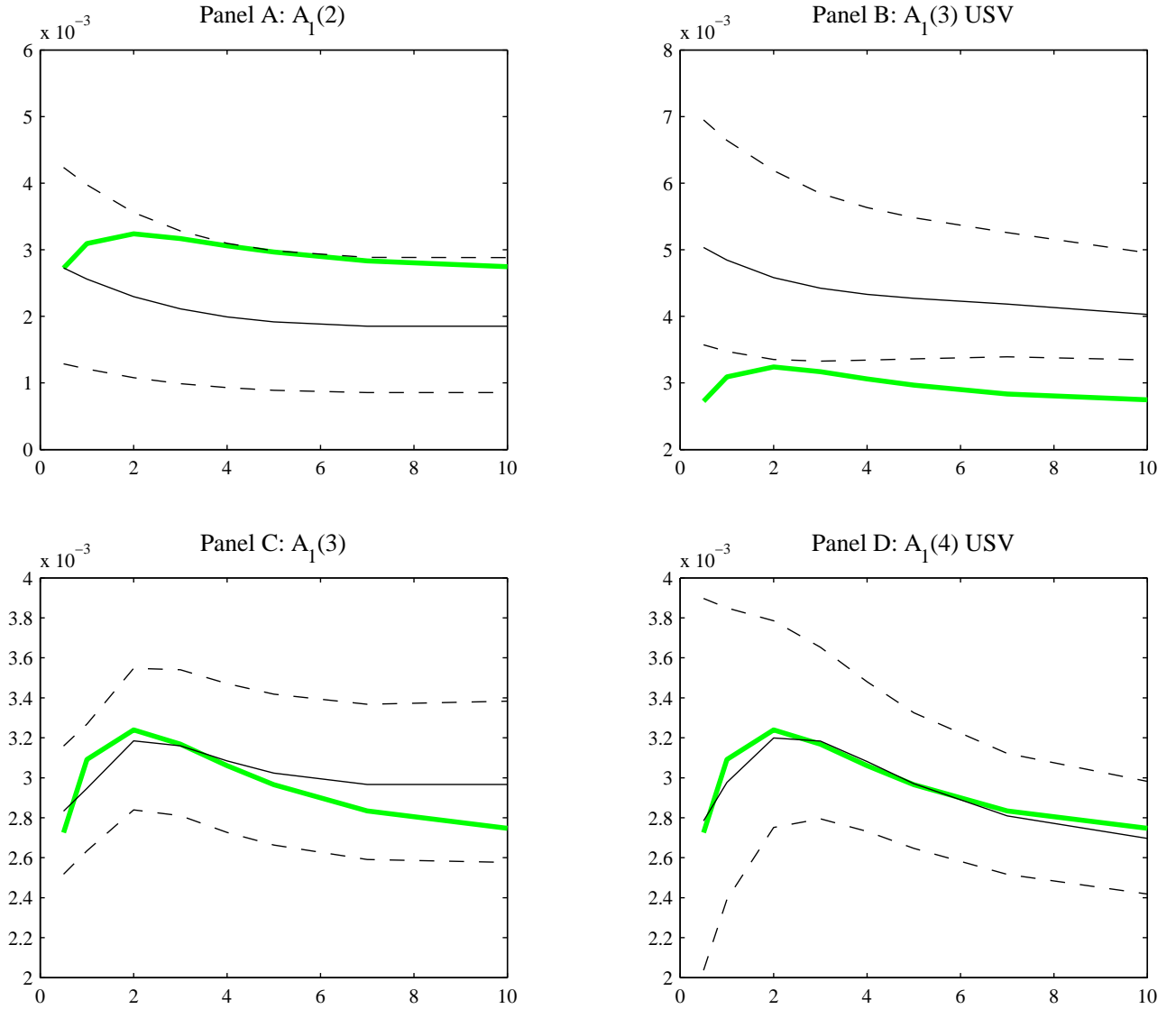
Estimated and model-implied estimates of the slope coefficient of the regression of yield changes on lagged yield curve slope



In each panel, the thick grey line depicts the sample slope coefficient of the regression of $Y(t + 1, \tau) - Y(t, \tau)$ on $(Y(t, \tau) - Y(t, .5)) / (\tau - .5)$ as a function of maturity (τ). Distributions of model-implied regression coefficients were calculated by simulation under the parameter values given in Table 4. The means and 95% confidence intervals of these distributions are depicted by solid and dashed black lines, respectively.

Figure 4

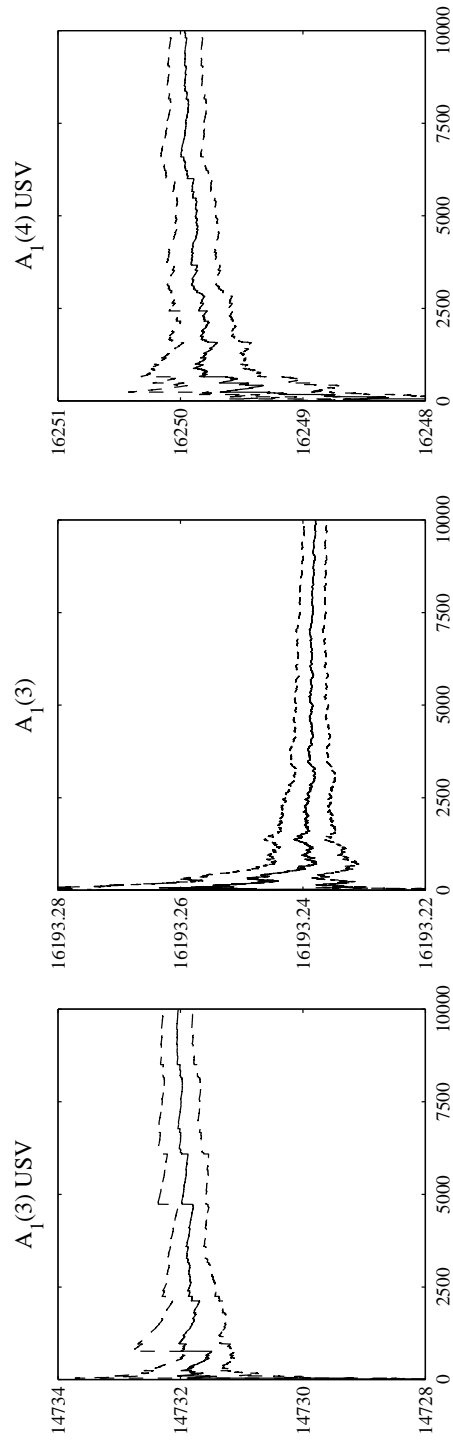
The maturity/volatility relation



In each panel, the thick grey line depicts the sample standard deviation of monthly changes in yields as a function of maturity. Distributions of model-implied sample standard deviations were calculated by simulation under the parameter values given in Table 4. The means and 95% confidence intervals of these distributions are depicted by solid and dashed black lines, respectively.

Figure A1

Recursive log likelihoods and 95% confidence intervals



Koopman and Shephard's ξ parameter (solid) and its 95% critical value (dashed)

