NBER WORKING PAPER SERIES

CAN INTEREST RATE VOLATILITY BE EXTRACTED FROM THE CROSS SECTION OF BOND YIELDS? AN INVESTIGATION OF UNSPANNED STOCHASTIC VOLATILITY

Pierre Collin-Dufresne
Robert S. Goldstein
Christopher S. Jones

Working Paper 10756
http://www.nber.org/papers/w10756

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 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. The views expressed herein are those of the author(s) and not necessarily those of the National Bureau of Economic Research.

Can Interest Rate Volatility be Extracted from the Cross Section of Bond Yields?
An Investigation of Unspanned Stochastic Volatility
Pierre Collin-Dufresne, Robert S. Goldstein, and Christopher S. Jones
NBER Working Paper No. 10756
September 2004
JEL No. G1, C4

ABSTRACT

Most affine models of the term structure with stochastic volatility (SV) 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 empirically that the $A1(3)$ SV 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).' Of the models tested, only the $A1(4)$ USV model is found to generate both realistic volatility estimates and a good cross-sectional fit. Our findings suggests that interest rate volatility cannot be extracted from the cross-section of bond prices. Separately, we propose an alternative to the canonical representation of affine models introduced by Dai and Singleton (2001). This representation has several advantages, including: (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 econometrically identifiable, (iv) model-insensitive estimates of the state vector process implied from the term structure are readily available, and (v) it isolates those parameters which are not identifiable from bond prices alone if the model is specified to exhibit USV.

Pierre Collin-Dufresne
Haas School of Business
University of California, Berkeley
545 Student Services Building #1900
Berkeley, CA 94720-1900
and NBER
dufresne@haas.berkeley.edu

Robert S. Goldstein
Carlson School of Management
University of Minnesota
Room 30122
321-19th Avenue, South
rgoldstein@csom.umn.edu

Christopher S. Jones
Marshall School of Business
University of Southern California
701 Hoffman Hall
Los Angeles, CA 90089
christopher.jones@marshall.usc.edu

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 Umantsev (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 report another trade-off between capturing cross-sectional and time-series properties of the term structure. Here, however, the trade-off involves 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. The former property implies that it should be possible to extract spot rate volatility solely from the cross-section of bond prices, independent of any time-series information. 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 (i.e., the 'term structure-implied volatility') 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.

We interpret these findings as evidence that the $A_1(3)$ model cannot simultaneously describe the

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

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

[3] See also Chacko (1997).

[4] Note that since the volatility structure is invariant under transformation from the historical measure to the risk-neutral measure, proposing a more general risk-premia specification will not overcome this problem as it did in Duffee (2002).

yield curve's level, slope, curvature, and volatility. That is, volatility is unable to play the dual role that the $A_1(3)$ model predicts it does. The estimation of such a model therefore presents a tradeoff between choosing volatility dynamics that are more consistent with one role or the other. For the data set we investigate, and with no parameter 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 series are (unconditionally, anyway) weakly correlated, suggesting that there may be no three-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 even roughly coincides with the quadratic variation of the spot rate, we also empirically investigate three and four-factor models that exhibit unspanned stochastic volatility, as defined by 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 cannot 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 good cross sectional and time series fits of yields. An implication of our findings is that any strategy that attempts to hedge the volatility risk inherent in fixed income derivatives (if feasible at all) must be substantially more complex than the convexity-based 'butterfly' positions discussed by Litterman, Scheinkman, and Weiss (1993). Indeed, our results suggest that implied spot rate volatility measures extracted from the cross-section of the yield curve are likely to be bad estimates of actual volatility.[6] 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 the 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

[5] We note that Brandt and Chapman (2004) report an estimate of a three-factor quadratic Gaussian model that performs very well with respect to the moments they choose to capture. However, they do not attempt to match, for example, the correlation between variance and curvature.

[6] This contrasts with results from the equity literature which show that implied volatility estimates backed out from a cross-section of option prices are in general good predictors of spot volatility. We speculate that the difference is due to the difference between bond and option payoffs. The latter are more non-linear.

stressed by DS, an important econometric issue in dealing with latent variables is that some of the 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.[7] 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 models in terms of the physical measure dynamics of a latent state vector.

In contrast to the 'invariant transformation' approach of DS, we identify below a set of rotations which allows us to transform an underidentified affine model with latent state variables into an 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 the state vector so that it is composed of two types of variables: (i) the first few components in the Taylor series expansion of the yield curve which have physical interpretations such as level, slope and curvature, and (ii) their quadratic covariation. This representation has several advantages over those previously proposed in the literature. First, since under this rotation the state variables have a physical meaning independent of the model, their time series estimates are readily comparable across models. This contrasts with latent variable representation. (For example, when going from a two-factor latent variable model (X_1, X_2) to a three-factor model (X_1, X_2, X_3), the latent variables X_1 and X_2 are not readily comparable across models.)

Second, this representation allows us to demonstrate that the issue of identification rests mainly with the specification of the risk-neutral dynamics and is mostly independent of the risk premia specification. Indeed, while DS identify a maximal model when the risk premia structure is restricted to the 'completely affine' structure, the models of Duffee (2002) and Cheredito et al. (2004) show that there are additional parameters that can be identified even though the model remains affine under both the P and Q measures. Moreover, as shown in Duarte (2003), models can be empirically tractable even if their P-dynamics are non-affine, so long as their Q-dynamics are affine. Below, we show how our rotation naturally generates a Q-maximal representation: it identifies the maximum number of risk-neutral parameters that can be estimated for a given class of affine models. Intuitively, the problem of parameter identification in traditional affine models is due to the state vector being latent. Instead, if the state vector is chosen to be a set of observable quantities, then financial contracts can be written on future values of these observable variables, in turn permitting the risk-neutral parameters to be identified. Furthermore, given a time-series of the state vector, the historical-measure parameters can be identified via vector autoregression. Hence, the issue of identification becomes moot once the state vector is written in terms of observables. We note that Duffie and Kan (1996) were the first to suggest rotating the state vector from a set of latent variables to a set of observable yields. Unfortunately, their rotation to a state vector of yields of

[7] DS identify three such types of 'invariant transformations': (i) rotation of the state vector $\mathcal{T}_{\mathcal{A}}$, (ii) Diffusion rescaling $\mathcal{T}_{\mathcal{D}}$, (iii) Brownian motion rotation $\mathcal{T}_{\mathcal{O}}$.

finite maturity is difficult to implement. We circumvent this obstacle by effectively choosing yields (and derivatives thereof) with *infinitesimal* maturities, which in turn generates a tractable affine observable state vector. Below, we use the DK framework to (i) clearly demonstrate why there is an identification problem when the state vector is latent, (ii) show that the issue basically only affects the risk-neutral specification, and (iii) show why the finite-maturity rotation suggested by DK is not tractable, while our rotation to infinitesimal-maturity yields (and their covariations) do provide a tractable framework.

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. The intuition for this result is as follows: For a given state vector X, bond prices satisfy the partial differential equation

$$rP = P_t + \sum_{i=1}^{N} P_{X_i}\, \mu^Q_{X_i} + \frac{1}{2} \sum_{i,j=1}^{N} P_{X_i X_j}\, \sigma_{X_i X_j}.$$

Hence, if bond prices do not span the fixed income market, that is, if they are independent of some state variable X_j so that $P_{X_j} = 0$ and $P_{X_j X_i} = 0\ \forall i$, then the risk-neutral parameters contained in $\mu^Q_{X_j}$ and $\sigma_{X_i X_j}\ \forall i$ do not show up in the PDE and hence cannot affect bond prices. This in turn implies that these parameters cannot be determined from the cross section of bond prices.[8] However, they can be identified from the prices of other fixed income securities. This example emphasizes that the notion of maximality introduced by DS implicitly assumes that, in addition to bonds, other fixed income derivatives are observed.[9] Because our representation rotates the state vector to include the 'unspanned stochastic variance' state variable, it isolates those parameters which cannot be identified from bond prices alone. Furthermore, this representation simplifies the form of the parameter constraints imposed by USV, in turn facilitating empirical investigation.

Below we provide a full characterization of the Q-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

[8] If the risk premia specification is somewhat restricted, some of these parameters might be identifiable from time-series information.

[9] Effectively, 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.

$A_1(3)$ and $A_1(4)$ models exhibiting USV. In Section 4 we describe the estimation technique used to account for USV, while Section 5 includes all empirical results. We conclude in Section 5.

2 Identifying Maximal Affine Models with Observable State Vectors

Mostly following the notation of Duffie and Kan (DK 1996), and Dai and Singleton (DS 2000), the risk-neutral dynamics of a Markov state vector X within an affine framework can be specified by as:

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

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

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

The spot rate is an affine function of X:

$$r(t) = \delta_0 + \delta_x^{\top} X(t)\,, \tag{3}$$

where δ_x is an N dimensional vector. Assuming the system is admissible (i.e., that the stochastic differential equation admits a unique strong solution[10]), then bond prices take the form:

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

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

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

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

and the initial conditions:

$$A(0) = 0, \qquad B(0) = 0. \tag{7}$$

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:

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

DK used this observation to suggest the possibility of rotating the system from a latent state vector X to observable yields Y. While such a rotation is not tractable in general for reasons we discuss below, it does provide important insights into the issue of parameter identification. In particular, note that by inverting equation (8) we may express the latent state variable X_i as a linear

[10] Sufficient conditions are given in Duffie and Kan (1996).

combination of yields.[11] The relative weightings of these yields depend on the functions $A(\tau)$ and $B(\tau)$, which in turn are functions of the risk-neutral parameters. Hence, each time a new trial parameter vector is considered, the physical definition of the latent variables change in that the relative weights of the observable yields change. That is, the latent state vector X does not have a physical definition independent of the values of the parameter vector! In contrast, yields can be extracted from market prices, and hence have a physical meaning independent of the parameter vector. Below, we will show that the problem of model identification stems from writing the system in terms of latent state variables, and that this issue vanishes when we rotate to a state vector which is observable.

While (8) provides a complete description of the relation between yields and state variables, a characterization that is more useful for our purposes may be obtained from Taylor series expansions of both the yield curve and the time-dependent coefficients $A(\tau)$ and $B(\tau)$. To simplify notations we define $\partial^n_{\tau=c} f(t,\tau) := \frac{\partial^n}{\partial \tau^n} f(t,\tau)\big|_{\tau=c}$. Performing these expansions at $\tau = 0$, we find:

$$Y(t,\tau) = Y(t,0) + \tau\,\partial_{\tau=0} Y(t,\tau) + \left(\frac{\tau^2}{2!}\right)\partial^2_{\tau=0} Y(t,\tau) + \ldots \tag{9}$$

$$A(\tau) = A(0) + \tau\,\partial_{\tau=0} A(\tau) + \left(\frac{\tau^2}{2!}\right)\partial^2_{\tau=0} A(\tau) + \ldots \tag{10}$$

$$B(\tau) = B(0) + \tau\,\partial_{\tau=0} B(\tau) + \left(\frac{\tau^2}{2!}\right)\partial^2_{\tau=0} B(\tau) + \ldots. \tag{11}$$

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 expansions:

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

Equation (12) implies that the $\{Y_n\}$ variables, representing the derivatives of the yield curve at $\tau = 0$, 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 $\mathbf{Y} = \{Y_0, Y_1, \ldots\}$ of the (observable) yield curve, which are therefore observable as well. As is well-known, any linear rotation of an affine model will generate a model that maintains the 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

[11] Here we ignore the USV issue of non-invertiblity.

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 $\{Y_0,\, Y_1,\, Y_2\}$:

$$\begin{aligned}
\partial_{\tau=0} A(\tau) &= -\delta_0 && (13)\\
\partial_{\tau=0} B(\tau) &= \delta_x && (14)\\
\partial^2_{\tau=0} A(\tau) &= -\theta^{Q\top}\mathcal{K}^{Q\top}\delta_x && (15)\\
\partial^2_{\tau=0} B(\tau) &= -\mathcal{K}^{Q\top}\delta_x && (16)\\
\partial^3_{\tau=0} A(\tau) &= \theta^{Q\top}\mathcal{K}^{Q\top}\mathcal{K}^{Q\top}\delta_x + \sum_{i=1}^{N}[\Sigma^\top\delta_x]_i^2\alpha_i && (17)\\
\partial^3_{\tau=0} B(\tau) &= \mathcal{K}^{Q\top}\mathcal{K}^{Q\top}\delta_x - \sum_{i=1}^{N}[\Sigma^\top\delta_x]_i^2\beta_i. && (18)
\end{aligned}$$

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

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

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

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

Hence, the level ($Y(t,0)$), slope ($\partial_{\tau=0}Y(t,\tau)$), and curvature ($\partial^2_{\tau=0}Y(t,\tau)$) 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.

The above discussion suggests an alternative to the canonical representation of DS (2000). To that effect let us define the following.

Definition 1 *A* Canonical Q-Representation *is an invariant transformation*[12] *of the latent state vector X given in equation (1) to a N-dimensional state vector $\mathbf{H} \equiv [\hat{\mathbf{Y}}, \hat{\mathbf{V}}]$ that combines elements of $\mathbf{Y}$ and $\mathbf{V}$ such that $\mathbf{H}$ (i) contains Y_0 and (ii) has jointly Markov dynamics under the risk-neutral measure.*

Let us note a few characteristics of our canonical Q-representation:

- By definition of the vectors $\mathbf{Y}$ and $\mathbf{V}$, the state variables in this representation are observable in that they have physical interpretations independent of the choice of the parameter vector.
- Since $Y_0(t) \equiv r(t)$, this definition insures that this system of observable state variables captures the dynamics of the entire term structure as well as fixed-income derivatives.[13]
- Since $\mathbf{H}$ is an invariant transformation of X it has jointly Markov affine dynamics:

$$d\mathbf{H}(t) = \mu(t)dt + \Sigma(t)\, dZ^Q(t) \tag{22}$$

with

$$\mu_i = m_0^i + m_Y^i \hat{\mathbf{Y}} + m_V^i \hat{\mathbf{V}} \quad \forall i \in (1, N) \tag{23}$$

$$\{\Sigma^\top \Sigma\}_{ij} = s_0^{ij} + s_V^{ij} \hat{\mathbf{V}} \quad \forall i, j \in (1, N) \tag{24}$$

where m_0^i, s_0^{ij} are constants and m_Y^i, m_V^i, s_V^{ij} are vectors of parameters.[14]

To illustrate our proposed representation, consider the $A_0(3)$ sub-family of models. Note that the covariance matrix of state vector dynamics is constant for this family. As such, all the state variables of our proposed representation must come from $\mathbf{Y}$ (i.e., from the Taylor series expansion of the yield curve) and not from $\mathbf{V}$. An appropriate state vector for this class of models would thus consist of $(Y_0(t),\, Y_1(t),\, Y_2(t))$, or equivalently, $\mathbf{H}(t) = (r(t),\, \mu^Q(t),\, \theta^Q(t))$, where $\theta^Q(t) \equiv \mathrm{E}^Q[d\mu^Q(t)]/dt$ is the expected change in the drift of the short rate. The equivalence between the two representations follows from the definitions of $Y_0(t),\, Y_1(t),\, Y_2(t)$ given in (19)-(21) above and the fact that $V(t)$ is constant in Gaussian models. We consider another Gaussian case in more detail in section 2.1 below.

As an alternative example, consider the $A_1(3)$ sub-family of models, where one state variable drives $V(t) = \frac{1}{dt}(dY_0(t))^2 \equiv \frac{1}{dt}(dr(t))^2$. For that case, it may be convenient to rotate the state vector to $\mathbf{H} = (r,\, \mu^Q,\, V)$ as we show in section 3.1 below. Note that the variance state variable is observable as well in that it has a physical interpretation independent of the model's parameter values, and, in theory, can be inferred perfectly form the quadratic variation of the short rate.

[12] The notion of invariant transformation is defined in DS (2000). See also footnote 7.

[13] Knowledge of the risk-neutral short rate process is sufficient to describe prices of all fixed-income derivatives (e.g., Glasserman and Jin (2001)).

[14] We emphasize that the rotation itself may naturally 'restrict' the parameters (to be consistent with the definition of the state variables). For example, suppose $H_1 = Y_0 \equiv r$ and $H_2 = Y_1 \equiv \frac{1}{2}\mu^Q$. Then the rotation implies that $m_0^1 = 0$ and $m_Y^1 = \{0, 2, 0 \ldots\}$, and $m_V^1 = \mathbf{0}$.

The main property of our proposed canonical Q-representation is Q-maximality, defined as follows:

Definition 2 *An $A_m(n)$ model is* Q-maximal *if (i) all the parameters of its Q-transition density can be identified solely from the cross-sectional information in bond and fixed-income derivative prices and (ii) the Q-measure transition density of the short rate process nests that of all other $A_m(n)$ models.*

We claim the following:

Proposition 1 *The canonical Q-representation of the state vector $\mathbf{H}$ is Q-maximal.*

Proof First we show that all parameters are identified from cross-section of derivative data. In particular, consider the fixed income derivative which has a payoff equal to $e^{\int_0^T r_s ds} e^{i\xi\cdot\mathbf{H}(T)}$ for some vector ξ of N constants. By definition the price of this derivative is the Q-measure characteristic function of the state vector

$$\Phi^Q(\xi, T) = \mathrm{E}^Q\left[e^{i\xi\cdot\mathbf{H}(T)}\right].$$

We can assume that such derivative exists since the components of $\mathbf{H}$ in our representation are observable. Hence, contracts whose payoffs are tied to future values of the state vector can be written, and agents will agree on the final payoff of such contracts independent of what model they use and what their parameter estimates are.[15] Since we may assume that a continuum (in T and ξ) of these derivatives are traded, this implies that conditional on a large panel data set of these derivatives, all Q-measure conditional moments of the state vector are observable. In particular, this implies that we can infer from derivatives data the left hand side of equations (23) and (24) for every date in our data set. Since the state vector $\mathbf{H}(t)$ is also observable at these dates we obtain, by stacking equations (23) and (24) for specific dates where we observe different $\mathbf{H}(t)$ values, a system of non-redundant linear equations, which with a large enough panel data can be perfectly inverted for all the parameters.

Second, the fact that the term structure dynamics generated by our model cannot be nested within a more general model follows directly from the fact that the transformation performed on the initial state vector X given in equation (1) to obtain $\mathbf{H}$ is an invariant transformation as defined in Dai and Singleton (2000). □

We emphasize that the notion of Q-maximality is independent of the specification of the risk premia. That is, it is independent of the P-measure dynamics of the state variables. This contrasts with the notion of maximality introduced in Dai and Singleton (2000), which is based on the P-measure likelihood. One drawback of their approach is that the notion of maximality changes as we consider more general risk-premium specifications. In particular, note that their canonical

[15] To accommodate the fact that the characteristic function has an imaginary component and there are no 'imaginary dollar,' simply assume that the derivative contract specifies that the value of the real part is paid separately from that of the imaginary part (or in some different numeraire).

representation is not maximal if one relaxes the 'completely affine' (e.g., DS (2000)) risk-premia specification in favor of the 'essentially affine' (e.g., Duffee (2002)) or the 'generalized essentially affine' (e.g., Cheridito et al (2004)). In contrast, all of these models are Q-maximal in the sense that they all possess the same number of risk-neutral parameters. Moreover, as emphasized by the work of Duarte (2003), one can obtain tractable models even if the P-measure dynamics are not affine, so long as the Q-measure dynamics remain affine. Our framework identifies the maximal number of risk-neutral parameters for a given class of models.

From there, a Q-maximal model can be combined with any form of risk-premium specification so long as the resulting P-dynamics of the state variables are identifiable. Since the diffusion parameters are unchanged between measures, only P-measure drift parameters must be checked. A common case would be when the P-measure drift of $\mathbf{H}$ is linear in state variables and/or parameter-free functions of state variables. Examples are affine models or the 'semi-affine' class of Duarte (2003), where drifts are linear in the state variables and their square roots. In those cases, the P-measure drift of H can be written as

$$\begin{aligned} \mu^P(t) &= C\,[1 \quad g_1(X(t)) \quad g_2(X(t)) \quad \cdots \quad g_M(X(t))] && (25) \\ &\equiv C\,G(t), && (26) \end{aligned}$$

and identification merely requires that (i) $G(t)$ is comprised of linearly independent time series and (ii) the mapping from underlying model parameters to the matrix C is one-to-one. Essentially, we are merely imposing the same conditions that one would require of a standard discrete-time vector autoregression, and we are only precluding such obviously degenerate cases as $y_i = \alpha + (\beta + \gamma)\, x_i + \epsilon_i$ or $y_i = \alpha + \beta x_i + \gamma\,(2x_i) + \epsilon_i$.

Thus, the notion of Q-maximality emphasizes that the problem of identification in affine models rests with the risk-neutral specification of the model. This is because we observe bond and fixed income derivative prices but not the latent state variables themselves. Only when a model's Q-measure parameters are identified can we extract the latent state variables from the data. Given the time series of these state variables, whether or not the remaining unknown P-measure parameters are identified is typically obvious.

2.1 Relation to Duffie and Kan's 'yield-factor model'

Conceptually, the rotation of the state vector to observables we propose 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, in practice, their approach turns out to be intractable for two reasons. First, for the subset of models exhibiting USV, the rotation proposed by DS fails as not all state variables can be written as linear combination of yields. Second, even for non-USV models where the rotation is, in principle, possible, the identification restrictions take the form of restrictions on the solution of the Riccatti equations, which are not generally known in closed-form. To provide

some intuition, consider a two-factor Gaussian (i.e., non-USV) model of the short rate r and a latent variable x:

$$dr_t = (\alpha_r + \beta_{rr}\, r_t + \beta_{rx}\, x_t)dt + \sigma_r\, dZ_r^Q(t) \tag{27}$$

$$dx_t = (\alpha_x + \beta_{xr}\, r_t + \beta_{xx}\, x_t)dt + \sigma_x\, dZ_x^Q(t) \tag{28}$$

Allowing for a correlation coefficient ($dZ_r^Q(t)\, dZ_x^Q(t) = \rho_{r,x} dt$), this model has a total of 9 risk-neutral parameters. We emphasize that *if* one could observe the risk-neutral trajectories of the two state variables, then one could estimate all 9 parameters from observing fixed income securities. However, in practice only data on yields and other fixed-income derivative securities are available. As a result, we show below that at most 6 risk-neutral parameters can be identified from observing fixed income derivatives data. We emphasize that this result depends solely on the risk-neutral dynamics of the state vector and is independent of the physical measure dynamics (which depends on a particular choice of risk-premia).

Since yields of arbitrary maturities are linear in r and x we have

$$Y(t,\tau) = -\frac{A(\tau)}{\tau} + \frac{B_r(\tau)}{\tau} r_t + \frac{B_x(\tau)}{\tau} x_t.$$

We can thus rotate from the latent state vector (r, x) to the observable state vector $(r, Y(t,\hat{\tau}))$ for some specific choice of $\hat{\tau} > 0$. As shown by DK, the dynamics of this state vector must be affine, i.e.:

$$dr_t = (\hat{\alpha}_r + \hat{\beta}_{rr}\, r_t + \hat{\beta}_{ry}\, Y(t,\hat{\tau}))dt + \sigma_r\, dZ_{r,t}^Q \tag{29}$$

$$dY(t,\hat{\tau}) = (\hat{\alpha}_y + \hat{\beta}_{yr}\, r_t + \hat{\beta}_{yy}\, Y(t,\hat{\tau})dt + \sigma_y\, dZ_{y,t}^Q \tag{30}$$

and yields are still affine in both state variables, i.e.:

$$\forall \tau \quad Y(t,\tau) = -\frac{\hat{A}(\tau)}{\tau} + \frac{\hat{B}_r(\tau)}{\tau} r_t + \frac{\hat{B}_y(\tau)}{\tau} Y(t,\hat{\tau}).$$

In particular, this must hold for the special case $\tau = \hat{\tau}$, which introduces three additional constraints, namely:

$$\hat{A}(\hat{\tau}) = 0, \quad \hat{B}_r(\hat{\tau}) = 0 \ \text{ and } \ \hat{B}_y(\hat{\tau}) = \hat{\tau}. \tag{31}$$

Intuitively, these additional (non-linear) restrictions will imply that at least three parameters in the specification of (29) and (30) above cannot be freely chosen. Hence, while the latent state vector representation (equations (27) and (28)) seems to suggest that there are nine free risk-neutral parameters, by rotating to an observable vector, we see that there are *probably* only six. Unfortunately, this 'yield-based' approach proposed by Duffie and Kan (1996) is intractable, because the coefficients $A(\tau)$ and $B(\tau)$ are generally not known in closed-form[16], making it difficult to impose the constraints implied in equation (31).

[16] We note that for the particular $A_0(2)$ model at hand, we do have analytic solutions for $A(\tau)$ and $B(\tau)$. But this does not affect the general point we are trying to make.

In contrast, our proposed representation circumvents the practical issues associated with DK's choice of finite maturity yields by choosing a different set of observable state variables, namely yields with infinitesimal maturity, or equivalently the derivatives of the term structure at zero $\{Y_0, Y_1 \ldots\}$. Indeed, our approach only involves the solution of these Ricatti equations and their higher order derivatives at zero, all of which are known functions of the parameters. Using our proposed representation, we would rotate from (r, x) to (r, μ^Q) which (as discussed above) is equivalent to (Y_0, Y_1). Using the definition of μ^Q and equation (27), we find $\mu^Q = \alpha_r + \beta_{rr}\, r_t + \beta_{rx}\, x_t$. Hence, the dynamics of the system become:

$$
\begin{aligned}
dr(t) &= \mu^Q(t)dt + \sigma_r\, dZ_r^Q(t) && (32)\\
d\mu^Q(t) &= (\beta_0 + \beta_1\, r_t + \beta_2\, \mu^Q(t))dt + \sigma_m\, dZ_m^Q(t) && (33)
\end{aligned}
$$

where $\sigma_m dZ_m^Q(t) = \beta_{rx}\sigma_x dZ_x^Q(t) + \beta_{rr}\sigma_r dZ_r^Q(t)$ and we have the following relation between parameters:

$$
\begin{aligned}
\beta_0 &= \beta_{rx}\alpha_x - \beta_{xx}\alpha_r && (34)\\
\beta_1 &= \beta_{rx}\beta_{xr} - \beta_{xx}\beta_{rr} && (35)\\
\beta_2 &= \beta_{rr} + \beta_{xx} && (36)\\
\sigma_m^2 &= \beta_{rx}^2\sigma_x^2 + \beta_{rr}^2\sigma_r^2 + 2\rho_{r,x}\beta_{rx}\sigma_x\beta_{rr}\sigma_r && (37)\\
\rho_{r,m}\sigma_m\sigma_r &= \rho_{r,x}\beta_{rx}\sigma_r\sigma_x + \beta_{rr}\sigma_r^2 && (38)
\end{aligned}
$$

Note that with 'no effort' our representation demonstrates that at most 6 (linear combination of) risk-neutral parameters are identifiable ($\sigma_r, \beta_0, \beta_1, \beta_2, \sigma_m, \rho_{r,m}$). Indeed, any choice of parameters in model (27)-(28) that leaves the left hand side of equations (34)-(38) unchanged, generates a short rate process which is path by path identical to that of model (32)-(33). Consequently, both models are observationally equivalent conditional on observing only term structure and fixed-income derivatives. In other words, only the left hand side of equations (34)-(38) are separately identifiable from fixed-income derivatives data.

Note further, that our discussion is valid irrespective of the risk-premium specification chosen. Starting from such a Q-maximal model, any risk-premium specification that is 'reasonable' in the sense that it leads to a P-measure state variable process identifiable from vector autoregression, will be maximal in the sense of DS (2000).

Up to this point, our example demonstrates that six risk-neutral parameters *at most* are identifiable from the cross section of fixed-income prices. As such, to demonstrate Q-maximality, we still need to show that all remaining parameters are in fact identifiable. Proposition 1 shows that this is in fact the case. Here we illustrate this idea in a different way by introducing a set of contingent claims whose prices will directly allow us to infer the parameters.

Suppose that a futures contract is traded with time to maturity Δ written on the *observable*

underlying μ^Q. The corresponding futures price $F(t,\Delta)$ is given by:

$$F(t,\Delta) = \mathrm{E}_t^Q[\mu^Q(t+\Delta)]$$

By definition when $\Delta \to 0$ we have the relation:

$$F(t) = \mu^Q(t) + (\beta_0 + \beta_1 r(t) + \beta_2 \mu^Q(t))\Delta$$

Since both (r, μ^Q) are observable, all we need is three set of (distinct) observations $\hat{F} \equiv \{F(t_1), F(t_2), F(t_3)\}$ to perfectly infer the parameters $(\beta_0, \beta_1, \beta_2)$ by running a cross-sectional 'regression' of $\hat{F}$ onto the observed values $\{r(t_1), r(t_2), r(t_3)\}$ and $\{\mu^Q(t_1), \mu^Q(t_2), \mu^Q(t_3)\}$. Similarly, the variance-covariance matrix parameters can be inferred from futures contracts written on $r(t+\Delta)^2$, on $\mu^Q(t+\Delta)^2$ and on $r(t+\Delta)\mu^Q(t+\Delta)$. Note that at no point does this approach require any assumption about risk-premia, as it uses only the cross-section of our panel data and not any of its time series properties.

Availability of derivative data and observability of the state variables are both crucial to our argument. While the fixed income securities we have used in the example above do not exist in practice, this argument guarantees nevertheless that theoretically our canonical Q-representation is identifiable. In practice, bonds alone are usually sufficient for identifying all risk-neutral parameters. When the model is specified to exhibit USV, bonds and simple fixed income derivatives such as caps are sufficient.

The next section documents using simulations, that the model-independent 'observability' of our state variables may also be of practical interest.

2.2 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 2bp 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:

$$\text{true } r_t = \alpha^r + \beta^r \times \text{estimated } r_t + \epsilon_t^r$$

$$\text{true } \mu_t^Q = \alpha^\mu + \beta^\mu \times \text{estimated } \mu_t^Q + \epsilon_t^\mu,$$

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.[17] 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 Q-maximality for stochastic volatility models

Below we focus on three and four factor models of the term structure which have only one factor driving stochastic volatility. This seems natural for two reasons. First, in their study of three factor models DS (2000) have shown that the $A_1(3)$ model is the least mis-specified at fitting various moments of the term structure. Second, Duffee (2002) shows that among three-factor models, Gaussian models perform best at capturing predictability regressions and for out of sample forecasts. However, there is also clear evidence that the conditional variance of the short rate is time-varying (e.g., Andersen, Benzoni and Lund (2003)). Thus it seems natural to only allow for one factor to

[17] We 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.

drive conditional variances. Finally, as we will see below the results of our investigation of three factor models call for the addition of a fourth factor.

3.1 Q-maximality and $A_1(3)$ Model

Consider the $A_1(3)$ model in the terminology of Dai and Singleton (2000). It is defined by 3 state variables, one of which follows a square-root process. One of the latent variable representations under the risk-neutral measure has 19 parameters:

$$dv = (\gamma_v - \kappa_v v)\,dt + \sigma_v\sqrt{v}\,dZ_3^Q \tag{39}$$

$$d\theta = [\gamma_\theta - \kappa_\theta\theta - \kappa_{\theta r}r - \kappa_{\theta v}v]\;dt + \sigma_{\theta r}\sqrt{\alpha_r + \alpha_v v}\,dZ_1^Q + \sqrt{\beta_\theta + \beta_v v}\,dZ_2^Q + \sigma_{\theta v}\sqrt{v}\,dZ_3^Q \tag{40}$$

$$dr = [\gamma_r - \kappa_r r - \kappa_{r\theta}\theta - \kappa_{rv}v]\;dt + \sqrt{\alpha_r + \alpha_v v}\,dZ_1^Q + \sigma_{r\theta}\sqrt{\beta_\theta + \beta_v v}\,dZ_2^Q + \sigma_{rv}\sqrt{v}\,dZ_3^Q\,. \tag{41}$$

Further, since we are interested in models where the short rate displays stochastic volatility we assume that

$$\alpha_v + \sigma_{r\theta}^2\beta_v + \sigma_{rv}^2 > 0.$$

DS demonstrate that this model is not identified, 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) defined by:[18]

$$\mu^Q = \gamma_r - \kappa_r r - \kappa_{r\theta}\theta - \kappa_{rv}v \tag{42}$$

$$V = \alpha_r + \sigma_{r\theta}^2\beta_\theta + (\alpha_v + \sigma_{r\theta}^2\beta_v + \sigma_{rv}^2)v \tag{43}$$

This rotation takes a model with 19 parameters, not all of which are identifiable, to a maximal model with 14 identifiable parameters inherent in its dynamics. Indeed, it is a matter of straightforward (but tedious) verification combining their definition given in equations (42) and (43) with the original dynamics of (39)-(41) we obtain:

$$dV_t = (\gamma_V - \kappa_V V_t)dt + \sigma_V\sqrt{V_t - \psi_1}\,dZ_1^Q(t) \tag{44}$$

$$dr_t = \mu_t^Q dt + \sigma_1\sqrt{V_t - \psi_1}\,dZ_1^Q(t) + \sqrt{\sigma_2^2 V_t - \psi_2}\,dZ_2^Q(t) + \sqrt{\sigma_3^2 V_t - \psi_3}\,dZ_3^Q(t) \tag{45}$$

$$\begin{aligned} d\mu_t^Q = {} & (m_0 + m_r r_t + m_\mu \mu_t^Q + m_V V_t)dt \\ & + \nu_1\sqrt{V_t - \psi_1}\,dZ_1^Q(t) + \nu_2\sqrt{\sigma_2^2 V_t - \psi_2}\,dZ_2^Q(t) + \nu_3\sqrt{\sigma_3^2 V_t - \psi_3}\,dZ_3^Q(t), \end{aligned} \tag{46}$$

where by definition of V_t we have:

$$\sigma_1^2 + \sigma_2^2 + \sigma_3^2 = 1 \tag{47}$$

$$\sigma_1^2\psi_1 + \psi_2 + \psi_3 = 0. \tag{48}$$

[18] Since we have restricted ourselves to model where the short rate displays stochastic volatility (i.e., $\alpha_v + \sigma_{r\theta}^2\beta_v + \sigma_{rv}^2 > 0$), such a rotation is always feasible. More generally, if one wanted to avoid this restriction, then a Q-maximal representation of the model would involve four state variables (r, μ^Q, θ^Q, V) (which would reduce to three when volatility is constant). For simplicity and given our focus on SV models we choose to impose the parameter restriction.

The model is admissible if[19]

$$\gamma_{_V} \geq \kappa_{_V}\psi_1 \tag{49}$$

$$\psi_1 \geq \max\left(\frac{\psi_2}{\sigma_2^2}, \frac{\psi_3}{\sigma_3^2}\right). \tag{50}$$

Note that all the parameters in equations 44-46 above can be expressed as linear combinations of parameters appearing in equations (42) and (43) above that are not separately identifiable. For example, for the volatility dynamics we have:[20]

$$\psi_1 = \alpha_{_r} + \beta_{_\theta}\sigma^2_{_{r\theta}} \tag{51}$$

$$\gamma_{_V} = (\alpha_{_r} + \beta_{_\theta}\sigma^2_{_{r\theta}})\kappa_{_v} + (\alpha_{_v} + \beta_{_v}\sigma^2_{_{r\theta}} + \sigma^2_{_{rv}})\gamma_{_v} \tag{52}$$

$$\kappa_{_V} = \kappa_{_v} \tag{53}$$

$$\sigma_{_V} = \sigma_{_v}\sqrt{\alpha_{_v} + \beta_{_v}\sigma^2_{_{r\theta}} + \sigma^2_{_{rv}}} \tag{54}$$

In addition to the advantages mentioned in the previous section, our proposed Q-representation is especially valuable 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.

The $A_{_1}(3)$ model is written above in equations (44)-(46). Alternatively, and for future reference, we can express the restrictions imposed by the maximality condition on the drift vector and instantaneous covariance matrix in the following form:

$$\begin{bmatrix} \frac{1}{dt}\mathrm{E}^{^Q}[dr] \\ \frac{1}{dt}\mathrm{E}^{^Q}[d\mu^{^Q}] \\ \frac{1}{dt}\mathrm{E}^{^Q}[dV] \end{bmatrix} = \begin{bmatrix} \mu^{^Q} \\ m_{_0} + m_{_r}r + m_{_\mu}\mu^{^Q} + m_{_V}V \\ \gamma_{_V} - \kappa_{_V}V \end{bmatrix} \tag{55}$$

$$\Sigma^2 = \begin{bmatrix} V & c_{_0} + c_{_V}V & \sigma_{_1}\sigma_{_V}\,(V - \psi_{_1}) \\ c_{_0} + c_{_V}V & \sigma_{_0}^{^\mu} + \sigma_{_V}^{^\mu}V & \nu_{_1}\sigma_{_V}\,(V - \psi_{_1}) \\ \sigma_{_1}\sigma_{_V}\,(V - \psi_{_1}) & \nu_{_1}\sigma_{_V}\,(V - \psi_{_1}) & \sigma_{_V}^2\,(\,V - \psi_{_1}) \end{bmatrix}, \tag{56}$$

where, by definition

$$c_{_V} = \sigma_{_1}\nu_{_1} + \sigma^2_{_2}\nu_{_2} + \sigma^2_{_3}\nu_{_3} \tag{57}$$

$$\sigma_{_V}^{^\mu} = \nu^2_{_1} + \nu^2_{_2}\sigma^2_{_2} + \nu^2_{_3}\sigma^2_{_3} \tag{58}$$

$$\sigma_{_0}^{^\mu} = -(\nu^2_{_1}\psi_{_1} + \nu^2_{_2}\psi_{_2} + \nu^2_{_3}\psi_{_3}) \tag{59}$$

$$c_{_0} = -(\sigma_{_1}\nu_{_1}\psi_{_1} + \nu_{_2}\psi_{_2} + \nu_{_3}\psi_{_3}). \tag{60}$$

[19]Note 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.

[20]We omit similar identities for the other parameters for the sake of brevity. They are available upon request.

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.

3.2 Q-maximality and the $A_1(3)$ USV Model

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

$$m_r = -2c_V^2 \tag{61}$$

$$m_\mu = 3c_V \tag{62}$$

$$m_V = 1 \tag{63}$$

$$\sigma_V^\mu = c_V^2. \tag{64}$$

Interestingly, note that our representation leads naturally to the condition in equation (63). 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 (61)-(64) and the admissibility conditions given in equations (49) and (50).

Combining the USV conditions (62) and (64) we see that to obtain stationary model under the Q measure (which requires that $m_\mu < 0$), the parameters must satisfy:

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

Hence, from the definitions in equations (57) and (58), it follows that the parameters must satisfy the following system of equations:

$$\begin{cases} \sigma_1\nu_1 + \sigma_2^2\nu_2 + \sigma_3^2\nu_3 = -\sqrt{\nu_1^2 + \nu_2^2\sigma_2^2 + \nu_3^2\sigma_3^2} \\ \sigma_1^2 + \sigma_2^2 + \sigma_3^2 = 1. \end{cases} \tag{65}$$

If we can find parameters that satisfy these equations, then the three USV conditions (61)-(63) can be satisfied by appropriately choosing the parameter values for m_r, m_μ, m_V. Further, the admissibility conditions (49) and (50) can be satisfied by appropriately choosing values for $\{\psi_1, \psi_2, \psi_3\}$.

To show there exists a solution to the system in equations (65), note that if we define the two vectors $\{u,\, v\}$ in $\Re^3$ 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:

$$\left\{ \begin{array}{rcl} ||u|| & = & 1 \\ u \cdot \frac{v}{||v||} & = & -1. \end{array} \right. \tag{66}$$

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

$$\sigma_1 = -\frac{\nu_1}{\sqrt{\nu_1^2 + \nu_2^2\sigma_2^2 + \nu_3^2\sigma_3^2}} \tag{67}$$

$$\sigma_2 = -\frac{\nu_2\sigma_2}{\sqrt{\nu_1^2 + \nu_2^2\sigma_2^2 + \nu_3^2\sigma_3^2}} \tag{68}$$

$$\sigma_3 = -\frac{\nu_3\sigma_3}{\sqrt{\nu_1^2 + \nu_2^2\sigma_2^2 + \nu_3^2\sigma_3^2}}\,. \tag{69}$$

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.[21] Hence, we focus our attention on Case 2.

In this case equation (68) holds for any value of ν_2. Equations (67) and (69) further imply that

$$\nu_1 = \sigma_1 c_V \tag{70}$$

$$\nu_3 = c_V\,. \tag{71}$$

Thus the system of equations becomes:

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

$$dr_t = \mu_t^Q dt + \sigma_1 \sqrt{V_t - \psi_1}\, dZ_1^Q(t) + \sqrt{(1-\sigma_1^2)V_t + \sigma_1^2\psi_1 + \psi_2}\, dZ_3^Q(t) + \sqrt{-\psi_2}\, dZ_2^Q(t) \tag{73}$$

$$\begin{aligned} d\mu_t^Q &= (m_0 - 2c_V^2 r_t + 3c_V\mu_t^Q + V_t)\, dt \\ &\quad + c_V\left(\sigma_1\sqrt{V_t - \psi_1}\, dZ_1^Q(t) + \sqrt{(1-\sigma_1^2)V_t + \sigma_1^2\psi_1 + \psi_2}\, dZ_3^Q(t)\right) + \nu_2\sqrt{-\psi_2}\, dZ_2^Q(t), \end{aligned} \tag{74}$$

with the following conditions:

$$\kappa_V > 0 \quad \text{for stationarity} \tag{75}$$

$$c_V < 0 \quad \text{for stationarity} \tag{76}$$

$$\gamma_V - \kappa_V\psi_1 > 0 \quad \text{for admissibility} \tag{77}$$

$$-\psi_2 > 0 \quad \text{for admissibility} \tag{78}$$

$$1 > \sigma_1^2 \quad \text{for admissibility} \tag{79}$$

$$\psi_1 + \psi_2 > 0 \quad \text{for admissibility.} \tag{80}$$

[21] The proofs are available upon request.

Thus the model has 9 parameters under the Q measure:

$$\gamma_{V},\, \kappa_{V},\, \sigma_{V},\, \psi_{1},\, \sigma_{1},\, c_{V},\, \psi_{2},\, m_{0},\, \nu_{2}.$$

Finally, we verify that the short rate process given by equations (72)-(74) above exhibits USV in that the zero-coupon bond price is not a function of the volatility state variable V_{t}:

Proposition 2 *If the short rate process follows the three-factor Markov process given by equations (72), (73) and (74), where the parameters satisfy the admissibility conditions (75)-(80), then zero-coupon bond prices are given by:*

$$P(t,T) = \exp\left(A(T-t) - B_{r}(T-t)r_{t} - B_{\mu}(T-t)\mu_{t}^{Q}\right), \tag{81}$$

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

$$B_{r}(\tau) = \frac{-3 + 4e^{c_{V}\tau} - e^{2c_{V}\tau}}{2c_{V}} \tag{82}$$

$$B_{\mu}(\tau) = \frac{(1 - e^{c_{V}\tau})^{2}}{2c_{V}^{2}} \tag{83}$$

$$A(\tau) = \int_{0}^{\tau} \left(\frac{1}{2}\sigma_{0}^{\mu} B_{\mu}^{2}(s) - m_{0} B_{\mu}(s) + c_{0} B_{r}(s) B_{\mu}(s)\right) ds \tag{84}$$

and where the parameters c_{0} and σ_{0}^{μ} can be written as

$$c_{0} = \psi_{2}\left(c_{V} - \nu_{2}\right) \tag{85}$$

$$\sigma_{0}^{\mu} = \psi_{2}\left[c_{V}^{2} - \nu_{2}^{2}\right]. \tag{86}$$

Proof See Appendix A □

From equations (82)-(84) 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 $(\gamma_{V},\, \kappa_{V})$ of V.[22] Rather, prices of other fixed income derivatives (e.g., caps) must be used to infer these risk-neutral parameters.

Note that bond prices would retain their exponential-affine form in the above model for *any* specification of the process for V_{t}. Indeed, the proof of Proposition 2 does not depend in on the specific process followed by the variance of the short rate.[23] 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

[22] This statement assumes that the risk premia are general enough so that the risk-neutral parameters $(\gamma_{V},\, \kappa_{V})$ are distinct from their physical-measure counterparts.

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

bond prices. Interestingly, the expression obtained for the term structure displays strong similarities to that of two-factor Gaussian model (such as that of Jegadeesh and Pennacchi (1996) for example) despite the fact that the short rate has stochastic volatility. Finally, we note that the integral in (84) can be obtained in closed-form as well, but to simplify notation we leave it in integral form.

3.3 Q-maximality and the $A_1(4)$ 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 $\theta_t^Q \equiv 3Y_{2,t} = \left(\frac{1}{dt}\mathrm{E}_t^Q[d\mu_t^Q] - V_t\right)$. The maximal $A_1(4)$ model is given by:

$$
\begin{aligned}
dV_t &= (\gamma_V - \kappa_V V_t)\,dt + \sigma_V\sqrt{V_t - \psi_1}\,dZ_1^Q(t) && (87)\\
dr_t &= \mu_t^Q dt + \sigma_1\sqrt{V_t - \psi_1}\,dZ_1^Q(t) + \sum_{i=2}^{4}\sqrt{\sigma_i^2 V_t - \psi_i}\,dZ_i^Q(t) && (88)\\
d\mu_t^Q &= (\theta_t^Q + V_t)dt + \nu_1\sqrt{V_t - \psi_1}\,dZ_1^Q(t) + \sum_{i=2}^{4}\nu_i\sqrt{\sigma_i^2 V_t - \psi_i}\,dZ_i^Q(t) && (89)\\
d\theta_t^Q &= (a_0 + a_r r_t + a_\mu \mu_t^Q + a_\theta \theta^Q + a_V V_t)dt \\
&\quad + \eta_1\sqrt{V_t - \psi_1}\,dZ_1^Q(t) + \sum_{i=2}^{4}\eta_i\sqrt{\sigma_i^2 V_t - \psi_i}\,dZ_i^Q(t), && (90)
\end{aligned}
$$

where by definition of V_t we have:

$$
\begin{aligned}
\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2 &= 1 && (91)\\
\sigma_1^2\psi_1 + \psi_2 + \psi_3 + \psi_4 &= 0. && (92)
\end{aligned}
$$

The model is admissible if

$$
\begin{aligned}
\gamma_V &\geq \kappa_V\psi_1 && (93)\\
\psi_1 &\geq \max\frac{\psi_i}{\sigma_i^2} \quad \forall\, i = 2,3,4 \quad \text{s.t.} \quad \sigma_i \neq 0. && (94)
\end{aligned}
$$

Note that the maximal unrestricted $A_1(4)$ model has a total of 22 free risk-neutral parameters $\left(\gamma_V, \kappa_V, \sigma_V,\ \{\psi_i, \nu_i, \eta_i, \sigma_i\}|_{i=1}^4\ , a_0, a_r, a_\theta, a_\mu, a_V\ \right)$, and two restrictions from equations (91) and (92).

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

$$
\begin{aligned}
\sigma &\equiv (\sigma_1,\ \sigma_2,\ \sigma_3,\ \sigma_4) && (95)\\
\nu &\equiv (\nu_1,\ \nu_2\sigma_2,\ \nu_3\sigma_3,\ \nu_4\sigma_4) && (96)\\
\eta &\equiv (\eta_1,\ \eta_2\sigma_2,\ \eta_3\sigma_3,\ \eta_4\sigma_4). && (97)
\end{aligned}
$$

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):

$$\Sigma^2 = \begin{bmatrix} V & c_{r\mu}V + c_{r\mu}^0 & c_{r\theta}V + c_{r\theta}^0 \\ c_{r\mu}V + c_{r\mu}^0 & \sigma_\mu^2 V + \sigma_\mu^0 & c_{\mu\theta}V + c_{\mu\theta}^0 \\ c_{r\theta}V + c_{r\theta}^0 & c_{\mu\theta}V + c_{\mu\theta}^0 & \sigma_\theta^2 V + \sigma_\theta^0 \end{bmatrix}, \tag{98}$$

where, by definition

$$c_{r\mu} \equiv \sigma \cdot \nu \tag{99}$$
$$c_{r\theta} \equiv \sigma \cdot \eta \tag{100}$$
$$c_{\mu\theta} \equiv \nu \cdot \eta \tag{101}$$
$$\sigma_\theta^2 \equiv \|\eta\|^2 \tag{102}$$
$$\sigma_\mu^2 \equiv \|\nu\|^2. \tag{103}$$

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

$$a_r = -2c_{r\mu}^2(3c_{r\mu} - a_\theta) \tag{104}$$
$$a_\mu = 7c_{r\mu}^2 - 3c_{r\mu}a_\theta \tag{105}$$
$$a_V = 3c_{r\mu} \tag{106}$$
$$\sigma_\mu = -c_{r\mu} \tag{107}$$
$$\sigma_\theta = c_{r\theta} \tag{108}$$
$$c_{r\theta} = c_{r\mu}^2 \tag{109}$$
$$c_{\mu\theta} = c_{r\mu}^3 \tag{110}$$

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 (91), (99), (103), and (107) imply that the vectors σ and ν are collinear but pointing in opposite directions. Similarly, equations (91), (100), (102) and (108) imply that the vectors σ and η are collinear and pointing in the same direction. Combining these results with the implications from equations (99), (100) and (109) we conclude that

$$\sigma = \frac{\eta}{\|\eta\|} = -\frac{\nu}{\|\nu\|} \tag{111}$$
$$\|\eta\| = \|\nu\|^2. \tag{112}$$

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 (111) 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 $\sqrt{\overline{\sigma^2}V - \overline{\psi}}\, dB^Q(t) \equiv \sum_{i=2}^4 \sqrt{\sigma_i^2 V - \psi_i}\, dZ_i^Q(t)$, where $\overline{\sigma^2} = \sum_{i=2}^4 \sigma_i^2$ and $\overline{\psi} = \sum_{i=2}^4 \psi_i$. 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 (111) and (112) imply:

$$\begin{aligned} \nu_1 &= c_{r\mu}\sigma_1 && (113)\\ \nu_2 &= c_{r\mu} && (114)\\ \eta_1 &= c_{r\mu}^2\sigma_1 && (115)\\ \eta_2 &= c_{r\mu}^2. && (116) \end{aligned}$$

Further, from equations (91) and (92) we have:

$$\begin{aligned} \sigma_1^2 + \sigma_2^2 &= 1 && (117)\\ \sigma_1^2\psi_1 + \psi_2 + \psi_3 + \psi_4 &= 0. && (118) \end{aligned}$$

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

$$dV_t = (\gamma_V - \kappa_V V_t)dt + \sigma_V\sqrt{V_t - \psi_1}\, dZ_1^Q(t) \quad (119)$$

$$\begin{aligned} dr_t &= \mu_t^Q\, dt + \sigma_1\sqrt{V_t - \psi_1}\, dZ_1^Q(t) + \sqrt{(1-\sigma_1^2)V_t + \sigma_1^2\psi_1 + \psi_3 + \psi_4}\, dZ_2^Q(t)\\ &\quad + \sqrt{-\psi_3}\, dZ_3^Q(t) + \sqrt{-\psi_4}\, dZ_4^Q(t) \quad (120) \end{aligned}$$

$$\begin{aligned} d\mu_t^Q &= (\theta_t^Q + V_t)\, dt + c_{r\mu}\sigma_1\sqrt{V_t - \psi_1}\, dZ_1^Q(t) + c_{r\mu}\sqrt{(1-\sigma_1^2)V_t + \sigma_1^2\psi_1 + \psi_3 + \psi_4}\, dZ_2^Q(t)\\ &\quad + \nu_3\sqrt{-\psi_3}\, dZ_3^Q(t) + \nu_4\sqrt{-\psi_4}\, dZ_4^Q(t) \quad (121) \end{aligned}$$

$$\begin{aligned} d\theta_t^Q &= \left(a_0 - 2c_{r\mu}^2(3c_{r\mu} - a_\theta)\, r_t + (7c_{r\mu}^2 - 3c_{r\mu}a_\theta)\, \mu_t^Q + a_\theta\, \theta^Q + 3c_{r\mu}\, V_t\right) dt\\ &\quad + c_{r\mu}^2\sigma_1\sqrt{V_t - \psi_1}\, dZ_1^Q(t) + c_{r\mu}^2\sqrt{(1-\sigma_1^2)V_t + \sigma_1^2\psi_1 + \psi_3 + \psi_4}\, dZ_2^Q(t)\\ &\quad + \eta_3\sqrt{-\psi_3}\, dZ_3^Q(t) + \eta_4\sqrt{-\psi_4}\, dZ_4^Q(t). \quad (122) \end{aligned}$$

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.[24]

[24]Note that the two restrictions $\sigma_3 = \sigma_4 = 0$ makes one of the seven restrictions from equations (104)-(110) redundant, leading to eight total restrictions, and thus 22 - 8 = 14 parameters.

The admissibility restrictions are:

$$\kappa_{_V} > 0 \qquad \text{for stationarity} \tag{123}$$

$$c_{_{r\mu}} < 0 \qquad \text{for stationarity: see equation (107)} \tag{124}$$

$$a_{_\theta} - 3c_{_{r\mu}} < 0 \qquad \text{for stationarity: see, for example, equation (131)} \tag{125}$$

$$\gamma_{_V} - \kappa_{_V}\psi_1 > 0 \qquad \text{for admissibility} \tag{126}$$

$$\psi_{_3}, \psi_{_4} < 0 \qquad \text{for admissibility} \tag{127}$$

$$1 > \sigma_{_1}^2 \qquad \text{for admissibility: see equation (117)} \tag{128}$$

$$\psi_{_1} + \psi_{_3} + \psi_{_4} > 0 \qquad \text{for admissibility.} \tag{129}$$

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.

Proposition 3 *If the short rate process follows a four-factor Markov process given by equations (119)-(122) where the parameters satisfy the admissibility conditions (123)-(129) then zero-coupon bond prices are given by:*

$$P(t,T) = \exp\left(A(T-t) - B_{_r}(T-t)\, r_{_t} - B_{_\mu}(T-t)\, \mu_{_t}^{Q} - B_{_\theta}(T-t)\, \theta_{_t}^{Q}\right), \tag{130}$$

where the deterministic functions $A(\tau)$, $B_{_r}(\tau)$, $B_{_\mu}(\tau)$, and $B_{_\theta}(\tau)$ are given by:

$$B_{_r}(\tau) = \frac{e^{c_{_{r\mu}}\,\tau}\,\left(6\,c_{_{r\mu}} - 2\,a_{_\theta}\right)}{4\,{c_{_{r\mu}}}^2 - c_{_{r\mu}}\,a_{_\theta}} + \frac{e^{2\,c_{_{r\mu}}\,\tau}\,\left(3\,c_{_{r\mu}} - a_{_\theta}\right)}{-10\,{c_{_{r\mu}}}^2 + 2\,c_{_{r\mu}}\,a_{_\theta}} + \frac{7\,c_{_{r\mu}} - 3\,a_{_\theta}}{-6\,{c_{_{r\mu}}}^2 + 2\,c_{_{r\mu}}\,a_{_\theta}} - \frac{2\,{c_{_{r\mu}}}^2\,e^{\left(-3\,c_{_{r\mu}} + a_{_\theta}\right)\tau}}{\Gamma} \tag{131}$$

$$B_{_\mu}(\tau) = \frac{a_{_\theta}}{2\,{c_{_{r\mu}}}^2\,\left(-3\,c_{_{r\mu}} + a_{_\theta}\right)} + \frac{e^{2\,c_{_{r\mu}}\,\tau}\,\left(2\,c_{_{r\mu}} - a_{_\theta}\right)}{10\,{c_{_{r\mu}}}^3 - 2\,{c_{_{r\mu}}}^2\,a_{_\theta}} + \frac{e^{c_{_{r\mu}}\,\tau}\,\left(c_{_{r\mu}} - a_{_\theta}\right)}{-4\,{c_{_{r\mu}}}^3 + {c_{_{r\mu}}}^2\,a_{_\theta}} + \frac{3\,c_{_{r\mu}}\,e^{\left(-3\,c_{_{r\mu}} + a_{_\theta}\right)\tau}}{\Gamma} \tag{132}$$

$$B_{_\theta}(\tau) = \frac{e^{c_{_{r\mu}}\,\tau}}{{c_{_{r\mu}}}^2\,\left(-4\,c_{_{r\mu}} + a_{_\theta}\right)} + \frac{1}{6\,{c_{_{r\mu}}}^3 - 2\,{c_{_{r\mu}}}^2\,a_{_\theta}} + \frac{e^{2\,c_{_{r\mu}}\,\tau}}{10\,{c_{_{r\mu}}}^3 - 2\,{c_{_{r\mu}}}^2\,a_{_\theta}} - \frac{e^{\left(-3\,c_{_{r\mu}} + a_{_\theta}\right)\tau}}{\Gamma} \tag{133}$$

$$A(\tau) = \int_0^\tau \left(\frac{\sigma_{_\mu}^0}{2}B_{_\mu}(s)^2 + \frac{\sigma_{_\theta}^0}{2}B_{_\theta}(s)^2 + B_{_r}(s)B_{_\mu}(s)c_{_{r\mu}}^0 + B_{_r}(s)B_{_\theta}(s)c_{_{r\theta}}^0 + B_{_\theta}(s)B_{_\mu}(s)c_{_{\mu\theta}}^0 - B_{_\theta}(s)a_{_0}\right) ds, \tag{134}$$

and where

$$\Gamma = \left(3\,c_{_{r\mu}} - a_{_\theta}\right)\,\left(4\,c_{_{r\mu}} - a_{_\theta}\right)\,\left(5\,c_{_{r\mu}} - a_{_\theta}\right). \tag{135}$$

Proof See Appendix A □

As before, we note that the integral in (134) can be obtained in closed-form as well, but to simplify notation we leave it in integral form. We now turn to the estimation of three and four-factor Q-maximal SV and USV models.

4 Empirical Approach

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 Q-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:

$$dZ_i^P(t) = dZ_i^Q(t) - \lambda_i(t)\,dt,\ \forall\, i = 1,2,3, \tag{136}$$

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

$$\lambda_1(t) = \frac{\lambda_{10} + \lambda_{13} V_t}{\sqrt{V_t - \psi_1}} \tag{137}$$

$$\lambda_2(t) = \frac{\lambda_{20} + \lambda_{21} r_t + \lambda_{22}\mu_t^Q + \lambda_{23} V_t}{\sqrt{\sigma_2^2 V_t - \psi_2}} \tag{138}$$

$$\lambda_3(t) = \frac{\lambda_{30} + \lambda_{31} r_t + \lambda_{32}\mu_t^Q + \lambda_{33} V_t}{\sqrt{\sigma_3^2 V_t - \psi_3}}. \tag{139}$$

By including a term in (137) 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 in Liptser and Shiryaev (1974, p. 294) 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.[25] We therefore impose the Feller condition for both measures as

[25] Cheridito, Filipovic, and Kimmel (2004) recently offer an alternative proof of this result in the context of affine models. Liptser and Shiryaev's result however applies more generally to any process of the 'diffusion type' (see their definition 7 p. 118).

a constraint in the likelihood maximization.[26] Combined with condition (50), the Feller condition implies that the Radon-Nikodym density defined by the risk-premia specification in (136) above integrates to one.

The flexibility of this specification of risk-premia allows for every drift parameter in the r, μ^Q, and V 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 dynamics of the state vector for the unrestricted $A_1(3)$ under the P measure are:

$$dV_t = \left(\gamma_V + \lambda_{V0} - (\kappa_V - \lambda_{VV})\, V_t\right) dt + \sigma_V \sqrt{V_t - \psi_1}\, dZ_1(t) \tag{140}$$

$$\begin{aligned} dr_t = &\left(\lambda_{r0} + \lambda_{rr} r_t + (1 + \lambda_{r\mu})\mu_t^Q + \lambda_{rV}\, V_t\right) dt \\ &+\sigma_1 \sqrt{V_t - \psi_1}\, dZ_1(t) + \sqrt{\sigma_2^2 V_t - \psi_2}\, dZ_2(t) + \sqrt{\sigma_3^2 V_t - \psi_3}\, dZ_3(t) \end{aligned} \tag{141}$$

$$\begin{aligned} d\mu_t^Q = &\left(m_0 + \lambda_{\mu 0} + (m_r + \lambda_{\mu r}) r_t + (m_\mu + \lambda_{\mu\mu})\mu_t^Q + (m_V + \lambda_{\mu V})\, V_t\right) dt \\ &+\nu_1 \sqrt{V_t - \psi_1}\, dZ_1(t) + \nu_2 \sqrt{\sigma_2^2 V_t - \psi_2}\, dZ_2(t) + \nu_3 \sqrt{\sigma_3^2 V_t - \psi_3}\, dZ_3(t). \end{aligned} \tag{142}$$

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

$$\lambda_1(t) = \lambda_1 \sqrt{V_t - \psi_1} \tag{143}$$

$$\lambda_2(t) = \lambda_2 \sqrt{\sigma_2^2 V_t - \psi_2} \tag{144}$$

$$\lambda_3(t) = \lambda_3 \sqrt{\sigma_3^2 V_t - \psi_3}. \tag{145}$$

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

$$\lambda_1(t) = \frac{\lambda_{10} + \lambda_{12} V_t}{\sqrt{V_t - \psi_1}} \tag{146}$$

$$\lambda_2(t) = \frac{\lambda_{20} + \lambda_{21}\mu_t^Q + \lambda_{22} V_t}{\sqrt{(1-\sigma_1^2)V_t + \sigma_1^2\psi_1 + \psi_3 + \psi_4}} \tag{147}$$

$$\lambda_3(t) = \frac{\lambda_{30} + \lambda_{31}\mu_t^Q + \lambda_{32} V_t}{\sqrt{-\psi_3}} \tag{148}$$

[26] The Feller condition for the Q-measure parameters of the process $V - \psi_1$ is simply $2(\gamma_V - \kappa_V \psi_1) > \sigma_V^2$. A similar condition applies for the P-measure parameters.

$$\lambda_4(t) \;=\; \frac{\lambda_{40} + \lambda_{41}\mu_t^Q + \lambda_{42}V_t}{\sqrt{-\psi_4}}. \tag{149}$$

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

$$dV_t \;=\; (\gamma_V - \kappa_V V_t)\,dt + \sigma_V\sqrt{V_t - \psi_1}\,dZ_1^Q(t) \tag{150}$$

$$dr_t \;=\; (\gamma_r - \kappa_{rr}r_t - \kappa_{rV}V_t)\,dt + \sigma_1\sqrt{V_t - \psi_1}\,dZ_1^Q(t) + \sqrt{(1-\sigma_1^2)V + \sigma_1^2\,\psi_1}\,dZ_2^Q(t). \tag{151}$$

Generalized essentially affine risk premia for this model are

$$\lambda_1(t) \;=\; \frac{\lambda_{10} + \lambda_{12}V_t}{\sqrt{V_t - \psi_1}} \tag{152}$$

$$\lambda_2(t) \;=\; \frac{\lambda_{20} + \lambda_{21}r_t + \lambda_{22}V_t}{\sqrt{(1-\sigma_1^2)V_t + \sigma_1^2\psi_1}}, \tag{153}$$

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.[27] 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.[28] 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

[27] If 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 Sundaresan (2002)).

[28] LIBOR 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.

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

represent pure noise.[29]

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, these studies, as well as that of Duffee and Stanton (2002), have found that methods based on the Kalman filter perform well in simulated samples, with minimal biases and relatively high accuracy. A natural explanation 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 suboptimal 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 obtain an unconditional value.[30]

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 $\mathbf{P} = \{\mathcal{P}_1, \mathcal{P}_2, ..., \mathcal{P}_T\}$ 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,

[29]There 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.

[30]We 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.

$\mathrm{p}(\mathbf{P}|\phi)$, can be written as the integral $\int \mathrm{p}(\mathbf{P},\mathbf{V}|\phi)\,d\mathbf{V}$, where $\mathbf{V} = \{V_1, V_2, ..., V_T\}$ denotes the time series of the latent stochastic variance process. The integral is evaluated using simulation. In general, the density $\mathrm{p}(\mathbf{V}|\mathbf{P},\phi)$ implied by the true model is unknown. But for an arbitrary density $\mathrm{p}^*(\mathbf{V}|\mathbf{P},\phi)$, we may still write

$$\begin{aligned}\int \mathrm{p}(\mathbf{P},\mathbf{V}|\phi)\,d\mathbf{V} &= \int \mathrm{p}(\mathbf{P},\mathbf{V}|\phi)\frac{\mathrm{p}^*(\mathbf{V}|\mathbf{P},\phi)}{\mathrm{p}^*(\mathbf{V}|\mathbf{P},\phi)}d\mathbf{V} \\ &\equiv \mathrm{E}^*\left[\frac{\mathrm{p}(\mathbf{P},\mathbf{V}|\phi)}{\mathrm{p}^*(\mathbf{V}|\mathbf{P},\phi)}\right]. \end{aligned} \tag{154}$$

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

$$\frac{\mathrm{p}(\mathbf{P},\mathbf{V}|\phi)}{\mathrm{p}^*(\mathbf{V}|\mathbf{P},\phi)} = \mathrm{p}(\mathbf{P}|\phi), \tag{155}$$

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 $\mathrm{p}^*(\mathbf{V}|\mathbf{P},\phi)$ that minimizes the variation in

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

thereby reducing the number of simulations required to estimate (154) accurately. 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 $\mathrm{p}(\mathbf{P},\mathbf{V}|\phi)$ 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 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

$$\begin{aligned}\mathrm{p}(\mathbf{P},\mathbf{V}|\phi) &= \prod_{t=1}^{T} \mathrm{p}(\mathcal{P}_t, V_t|\mathcal{P}_{t-1}, V_{t-1}, \phi) \\ &= \prod_{t=1}^{T} \mathrm{p}(\mathcal{P}_t, V_t|r_{t-1}, \mu_{t-1}^Q, V_{t-1}, \phi) \\ &= \prod_{t=1}^{T} \mathrm{p}(\mathcal{P}_{0,t}, V_t|r_{t-1}, \mu_{t-1}^Q, V_{t-1}, \phi)\ \mathrm{p}(\mathcal{P}_{\epsilon,t}|r_t, \mu_t^Q, V_t, \phi) \\ &= \left[\det(L_0^{*-1})\right]^T \prod_{t=1}^{T} \mathrm{p}(r_t, \mu_t^Q, V_t|r_{t-1}, \mu_{t-1}^Q, V_{t-1}, \phi)\ \mathrm{p}(\mathcal{P}_{\epsilon,t}|r_t, \mu_t^Q, V_t, \phi), \end{aligned} \tag{156}$$

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

$$\mathcal{P}_{0,t} = K_0 + \begin{bmatrix} r_t & \mu_t^Q & V_t \end{bmatrix} L_0. \tag{157}$$

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

$$L_0^* \equiv \begin{bmatrix} & & 0 \\ L_0 & & 0 \\ & & 1 \end{bmatrix}, \tag{158}$$

we obtain

$$\begin{bmatrix} \mathcal{P}_{0,t} & V_t \end{bmatrix} = [K_0 \;\; 0] + \begin{bmatrix} r_t & \mu_t^Q & V_t \end{bmatrix} L_0^*, \tag{159}$$

which implies the 'inversion formula'

$$\begin{bmatrix} r_t & \mu_t^Q & V_t \end{bmatrix} = \left(\begin{bmatrix} \mathcal{P}_{0,t} & V_t \end{bmatrix} - [K_0 \;\; 0] \right) L_0^{*-1}. \tag{160}$$

The term $\left[\det(L_0^{*-1})\right]^T$ in the likelihood decomposition thus reflects the change in variables from $\left\{\mathcal{P}_{0,t}, V_t\right\}$ to $\left\{r_t, \mu_t^Q, V_t\right\}$.[31]

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

$$\frac{1}{1000} \sum_i \frac{\left[\det(L_0^{*-1})\right]^T \prod_{t=1}^T \mathrm{p}(r_t^i, \mu_t^{Q,i}, V_t^i | r_{t-1}^i, \mu_{t-1}^{Q,i}, V_{t-1}^i, \phi) \; \mathrm{p}(\mathcal{P}_{\epsilon,t} | r_t^i, \mu_t^{Q,i}, V_t^i, \phi)}{\mathrm{p}^*(\mathbf{V}^i|\mathbf{P}, \phi)}, \tag{161}$$

where the V_t^i are simulated from $\mathrm{p}^*(\mathbf{V}^i|\mathbf{P}, \phi)$ and $\{r_t^i, \mu_t^{Q,i}\}$ are inferred via equation (160). By design, $\mathrm{p}^*(\mathbf{V}|\mathbf{P}, \phi)$ is known in closed-form. As discussed above, $\mathrm{p}(r_t^i, \mu_t^{Q,i}, V_t^i | r_{t-1}^i, \mu_{t-1}^{Q,i}, V_{t-1}^i, \phi)$ is treated as Gaussian, where we use Fisher and Gilles' (1996) algorithm for computing exact conditional means and covariances.

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

$$\mathcal{P}_{\epsilon,t} = K_\epsilon + \begin{bmatrix} r_t & \mu_t^Q & V_t \end{bmatrix} L_\epsilon. \tag{162}$$

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.[32]

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

[32] In 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.

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). We note, however, that with a sufficiently short sampling frequency (weekly in our case) this is unlikely to have a big impact.

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.[33] 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.

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 $\mathrm{E}[r_t|\mathbf{P},\phi]$. Following the estimation methodology, we can compute this expectation as

$$\begin{aligned}
\int r_t \mathrm{p}\left(r_t|\mathbf{P},\phi\right) dr_t &= \int r_t \int \mathrm{p}\left(r_t,\mathbf{V}|\mathbf{P},\phi\right) d\mathbf{V} dr_t \\
&= \int\int r_t \mathrm{p}\left(r_t|\mathbf{V},\mathbf{P},\phi\right) dr_t \mathrm{p}\left(\mathbf{V}|\mathbf{P},\phi\right) d\mathbf{V} \\
&= \int \mathrm{E}\left[r_t|\mathbf{V},\mathbf{P},\phi\right] \mathrm{p}\left(\mathbf{V}|\mathbf{P},\phi\right) d\mathbf{V} \\
&= \mathrm{E}^*\left[\mathrm{E}\left[r_t|\mathbf{V},\mathbf{P},\phi\right] \frac{\mathrm{p}\left(\mathbf{V}|\mathbf{P},\phi\right)}{\mathrm{p}^*\left(\mathbf{V}|\mathbf{P},\phi\right)}\right], \qquad (163)
\end{aligned}$$

where we have again used the importance sampling density introduced in equation (154). Given the joint invertibility of V_t and $\mathcal{P}_t$, r_t is nonstochastic given $\mathbf{V}$, $\mathbf{P}$, and ϕ, so $\mathrm{E}\left[r_t|\mathbf{V},\mathbf{P},\phi\right]$ is simply equal to the value obtained via (160). The smoothed estimate of r_t can therefore be computed via simulation. $\mathrm{E}[\mu_t^Q|\mathbf{P},\phi]$ and $\mathrm{E}[V_t|\mathbf{P},\phi]$ can be computed analogously, the latter also being relevant for both USV models.

[33] We 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.

5 Empirical 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 (61) to (64). 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. Nevertheless, 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 has many potential causes. One is that the third principal component (which explains just over 8% of yield variation) is free of observation error. Another explanation, which is supported below, is that the $A_1(3)$ model is deeply misspecified, so that the additional error term is required to offset some failing of the model.

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.[34] 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.

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.

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

5.1 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

$$\hat{e}_{1,t}^2 - \hat{e}_{2,t}^2. \tag{164}$$

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.[35]

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.

5.2 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 $\mathrm{E}[V_t|\mathbf{P}, \phi]$ 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.

[35] The 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.

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.

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 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.[36] 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 2.2. 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

[36] Litterman, 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 negative 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 .7, 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.

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 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.

5.3 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 $\mathrm{E}_t[Y_{t+1}] - \hat{Y}_t$, where $\mathrm{E}_t[Y_{t+1}]$ 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 absolute 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

$$\hat{\sigma}_{t,\tau} \equiv \sqrt{\sum_{i=1}^{N} \Delta Y(t,i,\tau)^2}, \tag{165}$$

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 8, 7, 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.

In the out of sample results, reported in Table 11, $A_1(4)$ 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 on average. 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).

5.4 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, $(Y(t,\tau) - Y(t,.5))/(\tau - .5)$.[37] 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.

6 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, and it is straightforward to show that our representation is 'maximal' (i.e., econometrically identified). Further, model-insensitive estimates of the process of the

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

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

[1] Y. Ait-Sahalia and R. Kimmel. Estimating affine multifactor term structure models using closed-form likelihood expansions. Working Paper, 2002.

[2] T. Andersen, L. Benzoni, and J. Lund. Stochastic volatility, mean-drift and jumps in the short-term diffusion: sources of steepness, level and curvature. Working paper Northwestern University, 2003.

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

[4] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.

[5] J. Y. Campbell and R. J. Shiller. Yield spreads and interest rate movements: a birds-eye view. *The Review of Economic Studies*, 58:495–514, 1991.

[6] G. Chacko. Multifactor interest rate dynamics and their implications for bond pricing. Working paper Harvard University, 1997.

[7] R. R. Chen and L. Scott. Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates. *Journal of Fixed Income*, December,Vol 3 no 3:14–31, 1993.

[8] P. Collin-Dufresne and R. S. Goldstein. Pricing swaptions in an affine framework. *Journal of Derivatives*, 10:9–26, 2002.

[9] P. Collin-Dufresne and B. Solnik. On the term structure of default premia in the swap and libor markets. *Journal of Finance*, VOL. LVI NO. 3:1095–1115, 2001.

[10] J. C. Cox, J. E. Ingersoll Jr., and S. A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53:385–407, 1985b.

[11] J.-C. Cuan and J.-G. Simonato. Estimating and testing exponential-affine term structure models by kalman filter. *Review of Quantitative Finance and Accounting*, 13:111–135, 1999.

[12] Q. Dai and K. Singleton. Fixed income pricing. Working paper Stanford University, 2002.

[13] Q. Dai and K. Singleton. Term structure dynamics in theory and reality. *The Review of Financial Studies*, 16:631–678, 2003.

[14] Q. Dai and K. J. Singleton. Specification analysis of affine term structure models. *Journal of Finance*, 55:1943–1978, 2000.

[15] F. de Jong. Time series and cross section information in affine term structure models. *Journal of Business and Economic Statistics*, 18:300–314, 2000.

[16] F. X. Diebold and R. S. Mariano. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13:253–263, 1995.

[17] J. Duarte. Evaluating an alternative risk preference in affine term structure models. Working Paper, Forthcoming Review of Financial Studies 2003.

[18] G. Duffee and R. Stanton. Estimation of dynamic term structure models. Working paper UC Berkeley, 2002.

[19] G. R. Duffee. Term premia and interest rate forecasts in affine models. *Journal of Finance*, 57 no 1, 2002.

[20] D. Duffie. *Dynamic Asset Pricing Theory*. Princeton: University Press, 1996.

[21] D. Duffie and R. Kan. A yield-factor model of interest rates. *Mathematical Finance*, 6:379–406, 1996.

[22] D. Duffie and K. Singleton. An econometric model of the term structure of interest-rate swap yields. *Journal of Finance*, LII.no.4:1287–1381, 1997.

[23] E. F. Fama and R. R. Bliss. The information in long-maturity forward rates. *Amercian Economic Review*, 77:680–692, 1987.

[24] M. Fisher and C. Gilles. Estimating exponential affine models of the term structure. Working paper, Federal Reserve Board, Washington DC, 1996.

[25] P. Glasserman and Y. Jin. Equilibrium positive interest rates: A unified view. *The Review of Financial Studies*, v14n1:187–214, 2001.

[26] B. Han. Stochastic volatilities and correlations of bond yields. Working paper Ohio-State University, 2004.

[27] M. Johannes and S. Sundaresan. Collateralized swap spreads. Working Paper, Columbia University, 2002.

[28] C. S. Jones. Bootstrapping zero coupon yields from asynchronous libor and swap quotes. working paper USC, 2003.

[29] S. J. Koopman and N. Shephard. Testing the assumptions behind the use of importance sampling. Working Paper, 2002.

[30] R. Liesenfeld and J.-F. Richard. Univariate and mutltivariate stochastic volatility models: Estimation and diagnostics. Working Paper, 2002.

[31] R. Litterman and J. Scheinkman. Common factors affecting bond returns. *Journal of Fixed Income*, june:54–61, 1991.

[32] R. Litterman, J. Scheinkman, and L. Weiss. Volatility and the yield curve. *Journal of Fixed Income*, june, 1991.

[33] J. Lund. Econometric analysis of continuous-time arbitrage-free models of the term structure of interest rates. Working Paper, 1997.

[34] D. B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59:347–370, 1991.

[35] W. K. Newey and K. D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708, 1987.

[36] W. K. Newey and K. D. West. Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies*, 61:631–653, 1994.

[37] N. D. Pearson and T.-S. Sun. Exploiting the condititional density in estimating the term structure: An application to the Cox, Ingersoll and Ross model. *Journal of Finance*, XLIX, no 4:1279–1304, 1994.

[38] G. Pennacchi. Identifying the dynamics of real interest rates and inflation evidence using survey data. *Review of Financial Studies*, 4:53–86, 1991.

[39] J.-F. Richard and W. Zhang. Accelerated monte carlo integration: An application to dynamic latent variable models. Simulation-Based Inference in Econometrics, R. Mariano, M. Weeks, and T. Schuermann, Editors, Cambridge University Press, Publisher, 1993.

[40] J.-F. Richard and W. Zhang. Econometric modelling of uk house prices using accelerated importance sampling. *Oxford Bulletin of Economics and Statistics*, 58:601–613, 1996.

[41] G. Sandmann and S. J. Koopman. Estimation of stochastic volatility models via monte carlo maximum likelihood. *Journal of Econometrics*, pages 271–301, 1998.

[42] K. J. Singleton and L. Umantsev. Pricing coupon-bond options and swaptions in affine term structure models. *Mathematical Finance*, 12, 2002.

[43] O. Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5:177–188, 1977.

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

$$dX_{i} = m_{i}^{Q}\left(\{X\},t\right)\,dt + \sum_{k=1}^{N}\sigma_{ik}\left(\{X\},t\right)\,dz_{k}^{Q}. \tag{A.1}$$

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

$$dr = r_{t}\,dt + \sum_{i=1}^{N} r_{i}\left[m_{i}^{Q}\,dt + \sum_{k=1}^{N}\sigma_{ik}\,dz_{k}^{Q}\right] + \frac{1}{2}\sum_{i,j,k=1}^{N} r_{ij}\sigma_{ik}\,\sigma_{jk}\,dt. \tag{A.2}$$

Note that this allows us to define

$$\begin{aligned}
\mu^{Q}(t) &\equiv \frac{1}{dt}\mathrm{E}_{t}^{Q}\left[dr\right] \\
&\equiv r_{t} + \sum_{i=1}^{N} r_{i}\,m_{i}^{Q} + \frac{1}{2}\sum_{i,j,k=1}^{N} r_{ij}\sigma_{ik}\,\sigma_{jk}. \qquad \text{(A.3)} \\
V(t) &\equiv \frac{1}{dt}\mathrm{Var}_{t}^{Q}\left[dr\right] \\
&\equiv \sum_{i,j,k=1}^{N} r_{i}\,r_{j}\,\sigma_{ik}\,\sigma_{jk}. \qquad \text{(A.4)}
\end{aligned}$$

Finally, from Ito's lemma we have

$$\mathrm{E}_{t}^{Q}\left[d\mu^{Q}(t)\right] = \mu_{t}^{Q}\,dt + \sum_{i=1}^{N}\mu_{i}^{Q}\,m_{i}^{Q}\,dt + \frac{1}{2}\sum_{i,\,j,\,k=1}^{N}\mu_{ij}^{Q}\,\sigma_{ik}\,\sigma_{jk}\,dt. \tag{A.5}$$

Using the relationship between yield to maturity and bond prices

$$P(t,\left(\{X\}\right),T) \equiv e^{-(T-t)\,Y(t,\{X\},T)}, \tag{A.6}$$

and similar notations as above, we find (assuming sufficient differentiability of the yield curve)

$$\begin{aligned}
P_{t} &= \left[Y - (T-t)\,Y_{t}\right]\,P \qquad \text{(A.7)} \\
P_{i} &= \left[-(T-t)\,Y_{i}\right]\,P \qquad \text{(A.8)} \\
P_{ij} &= \left[(T-t)^{2}\,Y_{i}\,Y_{j} - (T-t)\,Y_{ij}\right]\,P \qquad \text{(A.9)}
\end{aligned}$$

Bond prices satisfy the PDE

$$rP = P_{t} + \sum_{i=1}^{N} P_{i}\,m_{i}^{Q} + \frac{1}{2}\sum_{ijk=1}^{N} P_{ij}\sigma_{ik}\sigma_{jk} \tag{A.10}$$

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

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

Now we use a Taylor series expansion to write yields as

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

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

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

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

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

which is what we wished to prove. □

A.2: Proof of Proposition 2

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 (81). Indeed, in that case we have $e^{-\int_0^t r_s ds}P(t,T) = \mathrm{E}_t^Q\left[e^{-\int_0^T r_s ds}P(T,T)\right]$, which implies

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

since equations (81)-(84) 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 (81). 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

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

Thus,

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

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

□

A.3: Proof of Proposition 3

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 (81). Indeed, in that case we have $e^{-\int_0^t r_s ds} P(t,T) = \mathrm{E}_t^Q \left[e^{-\int_0^T r_s ds} P(T,T) \right]$, which implies

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

since equations (130)-(134) 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 (130). Using the fact that the functions $A(\cdot)$, $B_r(\cdot)$ and $B_\mu(\cdot)$ satisfy the system of ODE:

$$\begin{aligned}
B_r' &= a_r B_\mu + 1 && \text{(A.21)}\\
B_\mu' &= B_r + a_\mu B_\theta && \text{(A.22)}\\
B_\theta' &= B_\mu + a_\theta B_\theta && \text{(A.23)}\\
A' &= \frac{\sigma_\mu^0}{2} B_\mu^2 + \frac{\sigma_\theta^0}{2} B_\theta^2 + B_r B_\mu c_{r\mu}^0 + B_r B_\theta c_{r\theta}^0 + B_\theta B_\mu c_{\mu\theta}^0 - B_\theta, && \text{(A.24)}
\end{aligned}$$

and that, in particular, because of the restrictions on a_r, a_μ given in equations (104) and (105), 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

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

Therefore,

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

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

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

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

$$\mathrm{p}(V_t|V_{t-1},\mathbf{P},\phi) \propto \mathrm{p}\left(\mathcal{P}_{t+1},...,\mathcal{P}_T|V_t,V_{t-1},\mathcal{P}_1,...,\mathcal{P}_t,\phi\right)\mathrm{p}\left(V_t|V_{t-1},\mathcal{P}_1,...,\mathcal{P}_t,\phi\right). \tag{B.28}$$

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

$$\mathrm{p}\left(\mathcal{P}_{t+1},...,\mathcal{P}_T|V_t,\mathcal{P}_t,\phi\right)\mathrm{p}\left(V_t|V_{t-1},\mathcal{P}_{t-1},\mathcal{P}_t,\phi\right). \tag{B.29}$$

Note that within the context of the Gaussian quasi-likelihood, the second density is normal. Denote it's 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, $\mathrm{p}\left(\mathcal{P}_{t+1},...,\mathcal{P}_T|V_t,\mathcal{P}_t,\phi\right)$, 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

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

Liesenfeld and Richard (2002) note that the normalized density $\mathrm{p}_t^*\left(V_t\right)$ 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

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

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

$$\begin{aligned}
\frac{\mathrm{p}(\mathbf{P},\mathbf{V}|\phi)}{\mathrm{p}^*(\mathbf{V}|\mathbf{P},\phi)} &= \frac{\mathrm{p}\left(\mathcal{P}_1,V_1\right)}{m_1(V_1)}\times\prod_{t=2}^{T}\frac{\mathrm{p}\left(\mathcal{P}_t,V_t|\mathcal{P}_{t-1},V_{t-1}\right)}{p_t^*(V_t)} && \text{(B.32)}\\
&= \frac{\mathrm{p}\left(V_1|\mathcal{P}_1\right)\mathrm{p}\left(\mathcal{P}_1\right)}{p_1^*(V_1)}\times\prod_{t=2}^{T}\frac{\mathrm{p}\left(V_t|\mathcal{P}_{t-1},\mathcal{P}_t,V_{t-1}\right)\mathrm{p}\left(\mathcal{P}_t|\mathcal{P}_{t-1},V_{t-1}\right)}{p_t^*(V_t)} && \text{(B.33)}\\
&= \frac{\mathrm{p}\left(V_1|\mathcal{P}_1\right)\mathrm{p}\left(\mathcal{P}_1\right)\mathrm{p}\left(\mathcal{P}_2|\mathcal{P}_1,V_1\right)}{p_1^*(V_1)}\\
&\quad\times\prod_{t=2}^{T-1}\frac{\mathrm{p}\left(V_t|\mathcal{P}_{t-1},\mathcal{P}_t,V_{t-1}\right)\mathrm{p}\left(\mathcal{P}_{t+1}|\mathcal{P}_t,V_t\right)}{p_t^*(V_t)}\times\frac{\mathrm{p}\left(V_T|\mathcal{P}_{T-1},\mathcal{P}_T,V_{T-1}\right)}{p_T^*(V_T)} && \text{(B.34)}\\
&= \frac{\mathrm{p}\left(V_1|\mathcal{P}_1\right)\mathrm{p}\left(\mathcal{P}_1\right)\mathrm{p}\left(\mathcal{P}_2|\mathcal{P}_1,V_1\right)}{k_1(V_1,a_1)/\chi(a_1)}\\
&\quad\times\prod_{t=2}^{T-1}\frac{\mathrm{p}\left(V_t|\mathcal{P}_{t-1},\mathcal{P}_t,V_{t-1}\right)\mathrm{p}\left(\mathcal{P}_{t+1}|\mathcal{P}_t,V_t\right)}{k_t(V_t,a_t)/\chi(V_{t-1},a_t)}\times\frac{\mathrm{p}\left(V_T|\mathcal{P}_{T-1},\mathcal{P}_T,V_{T-1}\right)}{k_T(V_T,a_T)/\chi(V_{T-1},a_T)} && \text{(B.35)}\\
&= \mathrm{p}\left(\mathcal{P}_1\right)\chi(a_1)\frac{\mathrm{p}\left(V_1|\mathcal{P}_1\right)\mathrm{p}\left(\mathcal{P}_2|\mathcal{P}_1,V_1\right)}{k_1(V_1,a_1)/\chi(V_1,a_2)}\\
&\quad\times\prod_{t=2}^{T-1}\frac{\mathrm{p}\left(V_t|\mathcal{P}_{t-1},\mathcal{P}_t,V_{t-1}\right)\mathrm{p}\left(\mathcal{P}_{t+1}|\mathcal{P}_t,V_t\right)}{k_t(V_t,a_t)/\chi(V_t,a_{t+1})}\times\frac{\mathrm{p}\left(V_T|\mathcal{P}_{T-1},\mathcal{P}_T,V_{T-1}\right)}{k_T(V_T,a_T)} && \text{(B.36)}\\
&= \mathrm{p}\left(\mathcal{P}_1\right)\chi(a_1)\frac{\frac{1}{\sigma_{0,1}\sqrt{2\pi}}\mathrm{p}\left(\mathcal{P}_2|\mathcal{P}_1,V_1\right)\chi(V_1,a_2)}{\exp\left(a_{1,1}V_1+a_{2,1}V_1^2\right)}\\
&\quad\times\prod_{t=2}^{T-1}\frac{\frac{1}{\sigma_{0,t}\sqrt{2\pi}}\mathrm{p}\left(\mathcal{P}_{t+1}|\mathcal{P}_t,V_t\right)\chi(V_t,a_{t+1})}{\exp\left(a_{1,t}V_t+a_{2,t}V_t^2\right)}\times\frac{\frac{1}{\sigma_{0,T}\sqrt{2\pi}}}{\exp\left(a_{1,T}V_T+a_{2,T}V_T^2\right)} && \text{(B.37)}\\
&= \frac{\mathrm{p}\left(\mathcal{P}_1\right)\chi(a_1)}{\sigma_{0,1}}\frac{\frac{1}{\sigma_{0,2}\sqrt{2\pi}}\mathrm{p}\left(\mathcal{P}_2|\mathcal{P}_1,V_1\right)\chi(V_1,a_2)}{\exp\left(a_{1,1}V_1+a_{2,1}V_1^2\right)}\\
&\quad\times\prod_{t=2}^{T-1}\frac{\frac{1}{\sigma_{0,t+1}\sqrt{2\pi}}\mathrm{p}\left(\mathcal{P}_{t+1}|\mathcal{P}_t,V_t\right)\chi(V_t,a_{t+1})}{\exp\left(a_{1,t}V_t+a_{2,t}V_t^2\right)}\times\frac{\frac{1}{\sqrt{2\pi}}}{\exp\left(a_{1,T}V_T+a_{2,T}V_T^2\right)} && \text{(B.38)}\\
&= \frac{\mathrm{p}\left(\mathcal{P}_1\right)\chi(a_1)}{\sigma_{0,1}(2\pi)^{T/2}}\frac{\mathrm{p}\left(\mathcal{P}_2|\mathcal{P}_1,V_1\right)\chi(V_1,a_2)}{\sigma_{0,2}\exp\left(a_{1,1}V_1+a_{2,1}V_1^2\right)}\\
&\quad\times\prod_{t=2}^{T-1}\frac{\mathrm{p}\left(\mathcal{P}_{t+1}|\mathcal{P}_t,V_t\right)\chi(V_t,a_{t+1})}{\sigma_{0,t+1}\exp\left(a_{1,t}V_t+a_{2,t}V_t^2\right)}\times\frac{1}{\exp\left(a_{1,T}V_T+a_{2,T}V_T^2\right)} && \text{(B.39)}
\end{aligned}$$

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 $\mathrm{p}\left(V_t|V_{t-1},\mathcal{P}_{t-1},\mathcal{P}_t,\phi\right)$ 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 (B.39) 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 (B.39)

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

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

Given the simulated V_t, we do so by regressing

$$\ln \mathrm{p}\left(\mathcal{P}_{t+1}|\mathcal{P}_t, V_t\right) + \ln \chi(V_t, a_{t+1}) - \ln \sigma_{0,t+1} \tag{B.40}$$

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 $\mathrm{p}_t^*(V_t)$ 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 5. 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

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

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

From Figure 5, 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 may 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

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

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 $dx_{i,t} = \left[\kappa_i(\theta_i - x_{i,t}) - \lambda_i x_{i,t}\right] dt + \sigma_i\sqrt{x_{i,t}}dz_{i,t}^Q$, 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 it's 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	-0.074	-0.299	-0.064
	(0.292)	(0.069)	(0.286)	(0.059)
β^r	1.033	1.008	1.032	1.005
	(0.017)	(0.004)	(0.017)	(0.005)
R^2	0.999	0.999	0.998	0.997
	(0.001)	(0.000)	(0.002)	(0.002)
		Short Rate Drift		
$\alpha^\mu \times 100$	-0.042	0.024	-0.013	0.155
	(0.008)	(0.020)	(0.023)	(0.088)
β^μ	1.631	1.129	1.599	1.026
	(0.006)	(0.014)	(0.022)	(0.058)
R^2	0.996	0.980	0.976	0.890
	(0.002)	(0.010)	(0.012)	(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	0.2454	-0.1305	-0.1631
	(0.0333)	(0.0411)	(0.0501)	(0.0446)
σ_2		0	0.9898	0.9866
		NA	(0.0183)	(0.0074)
σ_3		0.9694	0.0567	0
		(0.0104)	(0.3054)	NA
ν_1		-0.0274	6.6379	0.0842
		(0.0046)	(4.1839)	(0.0230)
ν_2		-84.4499	-3.7288	-0.5160
		(2184.1561)	(0.8022)	(0.0063)
ν_3		-0.1116	3.7282	-430.3842
		(0.0019)	(3.2253)	(2377.8647)
ν_4				-76.7401
				(8.4888)
$\sigma_V \times 10^3$	0.9942	8.1198	1.8390	5.0652
	(0.1404)	(0.2545)	(0.2764)	(0.1936)
$\psi_1 \times 10^5$	0.0002	0.0112	2.3310	0.0471
	(2.1015)	(0.0357)	(1.1669)	(0.0331)
$\psi_2 \times 10^5$		-0.0004	2.2838	0.0082
		(0.0022)	(1.8648)	(0.0024)
$\psi_3 \times 10^5$		0.0002	-2.3235	0.0001
		(0.0033)	(1.8699)	(0.0001)
$\psi_4 \times 10^5$				-0.0094
				(0.0023)
κ_V^P	0.1559	0.4650	0.0071	0.2527
	(0.0980)	(0.1655)	(0.1667)	(0.0301)
$\gamma_V^P \times 10^4$	0.2694	0.4810	0.0192	0.2540
	(0.1572)	(0.0706)	(0.2275)	(0.0239)
$m_0 \times 10^2$		-0.1513	-15.3115	
		(0.0085)	(1.1706)	
$m_r \times 10^2$		-2.4899	-104.9432	
		(0.0853)	(3.6014)	
m_μ		-0.3347	-2.3793	
		(0.0057)	(0.0585)	
m_V		1	-521.8524	
		NA	(101.9732)	

Additional parameters for the $A_1(2)$ model

γ_r	κ_{rr}	κ_{rV}
0.0102	0.3297	-247.4986
(0.0068)	(0.0069)	(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	-1.5576	-714.8040	113.8364	-0.5160
(2.5943)	(0.0118)	(122.3995)	(13.4077)	(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	-0.0289	-0.0382	-0.0066
	(0.0400)	(0.0061)	(0.0364)	(0.0013)
λ_{rr}	0.3022	0.4265	0.3681	
	(0.2176)	(0.0941)	(0.2323)	
$\lambda_{r\mu}$		1.6801	-0.1711	-0.0472
		(0.2838)	(0.1816)	(0.1340)
λ_{rV}	170.7519	-244.5913	73.9945	2.5103
	(263.8621)	(35.4346)	(202.5647)	(37.9357)
$\lambda_{\mu 0}$		0.0224	0.2550	-0.0152
		(0.0078)	(0.1222)	(0.0065)
$\lambda_{\mu r}$		-0.3143	-0.3266	
		(0.1017)	(0.8028)	
$\lambda_{\mu\mu}$		-0.7552	0.3485	-0.5597
		(0.3365)	(0.6648)	(0.5148)
$\lambda_{\mu V}$		93.3392	248.4142	-1.1103
		(23.4874)	(673.2265)	(65.2360)
$\lambda_{\theta 0}$				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		0.6549	
	(0.1571)		(0.2274)	
λ_{VV}	-0.1559		-0.3672	
	(0.0980)		(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 significe 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 significe 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 logorithm 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 currend model fitted values. The table examines the bias, root mean squared error, and autocorrelation of $\hat{e}_{t+1} = (Y_{t+1} - Y_t) - (\mathrm{E}_t[Y_{t+1}] - \hat{Y}_t)$, where $\mathrm{E}_t[Y_{t+1}]$ is the model-implied expectation and $\hat{Y}_t$ is the model fitted value. * and ** denote statistical significe 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 currend model fitted values. The table examines the bias, root mean squared error, and autocorrelation of $\hat{e}_{t+1} = (Y_{t+1} - Y_t) - (\mathrm{E}_t[Y_{t+1}] - \hat{Y}_t)$, where $\mathrm{E}_t[Y_{t+1}]$ is the model-implied expectation and $\hat{Y}_t$ is the model fitted value. * and ** denote statistical significe 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 absolute yield changes ($\mathrm{E}\left[|\Delta Y|\right]$) and expected "realized volatility" ($\mathrm{E}\left[\hat{\sigma}\right]$) are calculated for .5, 1, 2, 3, 4, 5, 7, and 10-year maturities. Realized volatility is defined by $\hat{\sigma}^2_{t,\,\tau} = \sum_{i=1}^{5} \Delta Y(t,i,\tau)^2$ and is calculated using daily data. The table examines the forecast bias (actual minus forecast) and root mean squared error of $|\Delta Y|$ and $\hat{\sigma}$, where all yields are expressed in basis points. * and ** denote statistical significance at the 5% and 1% levels, respectively, where standard errors are calculated using the method of Newey and West (1987) with 8, 7, 16, and 16 lags, respectively, for the four panels of the table. For biases, statistical significance relates to the null hypothesis that the bias is zero. For RMSE, the statistical significance of the pairwise comparison of two models is reported, along with an inequality sign that reflects the direction of the rejection. The sample size is 625 weeks.

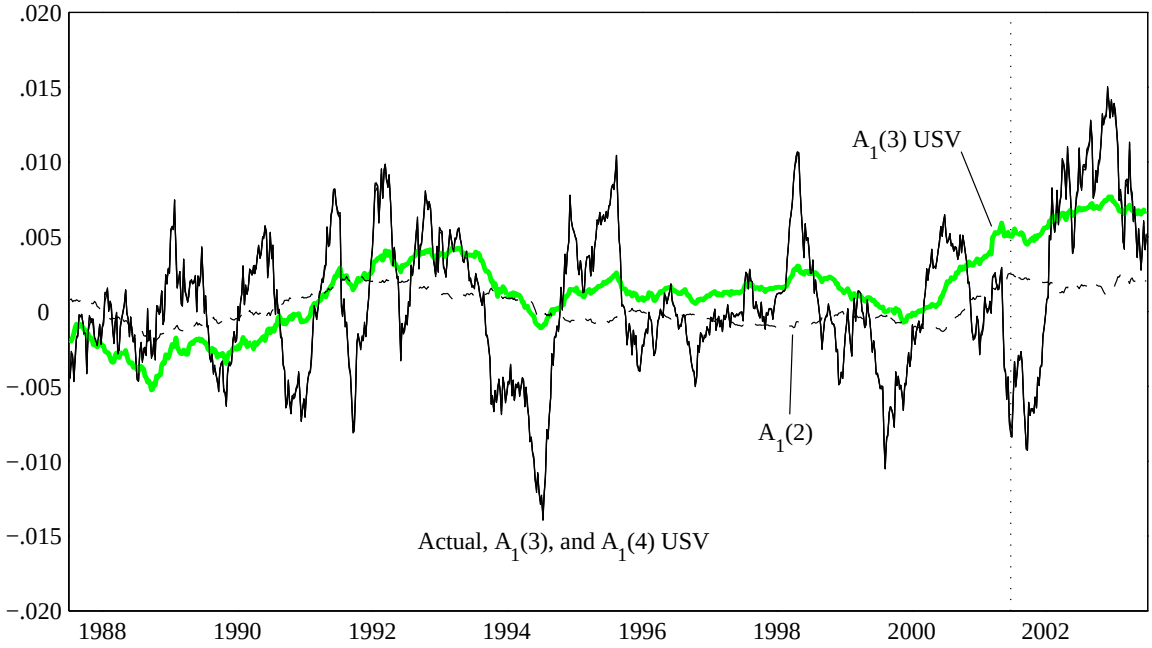
	$A_1(2)$		$A_1(3)$ USV		$A_1(3)$		$A_1(4)$ USV
				bias in weekly $\lvert\Delta Y\rvert$			
6-month	-5.39**		-1.67**		-3.05**		-2.04**
1-year	-2.26**		0.50		-1.50**		-1.18**
2-year	0.66		1.53**		-0.78		-1.07**
3-year	1.81**		1.09**		-0.32		-0.99
4-year	2.25**		0.29		-0.07		-0.84
5-year	2.41**		-0.46		0.04		-0.66
7-year	2.43**		-1.45**		0.08		-0.37
10-year	2.49**		-1.68**		0.21		0.09
				RMSE of weekly $\lvert\Delta Y\rvert$			
6-month	10.02	$>^{**}$	8.48	$<^{**}$	9.20	$>^{**}$	8.48
1-year	9.35	$>^{*}$	8.98	$<^{**}$	9.40	$>^{**}$	9.00
2-year	9.44		9.41		9.67	$>^{**}$	9.41
3-year	9.56	$>^{*}$	9.37		9.59	$>^{**}$	9.43
4-year	9.54	$>^{**}$	9.25		9.45	$>^{*}$	9.33
5-year	9.47	$>^{*}$	9.17		9.32		9.21
7-year	9.32		9.14		9.13		9.03
10-year	9.20		9.02		9.01	$>^{**}$	8.89
				bias in $\hat{\sigma}$			
6-month	-7.96**		-3.50**		-5.12**		-3.88**
1-year	-4.31**		-0.99*		-3.38**		-2.98**
2-year	-0.10		0.98*		-1.84**		-2.14**
3-year	1.66**		0.84		-0.91		-1.66**
4-year	2.32**		0.02		-0.47		-1.35**
5-year	2.57**		-0.81		-0.28		-1.07*
7-year	2.69**		-1.90**		-0.13		-0.62
10-year	2.87**		-2.07**		0.13		0.04
				RMSE of $\hat{\sigma}$			
6-month	9.59	$>^{**}$	6.49	$<^{*}$	7.66	$>^{*}$	6.56
1-year	7.27	$>^{**}$	6.18	$<^{*}$	6.93		6.54
2-year	6.61		6.68		7.07		6.89
3-year	6.91		6.74		6.97		6.89
4-year	6.90		6.52		6.69		6.63
5-year	6.79		6.39		6.43		6.37
7-year	6.58		6.34		6.12		6.02
10-year	6.53		6.21		6.09	$>^{**}$	5.88

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 absolute yield changes ($\mathrm{E}\left[|\Delta Y|\right]$) and expected "realized volatility" ($\mathrm{E}\left[\hat{\sigma}\right]$) are calculated for .5, 1, 2, 3, 4, 5, 7, and 10-year maturities. Realized volatility is defined by $\hat{\sigma}^2_{t,\,\tau} = \sum_{i=1}^{5} \Delta Y(t,i,\tau)^2$ and is calculated using daily data. The table examines the forecast bias (actual minus forecast) and root mean squared error of ${\Delta Y}^2$ 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) with 2, 2, 4, and 3 lags, respectively, for the four panels of the table. For biases, statistical significance relates to the null hypothesis that the bias is zero. For RMSE, the statistical significance of the pairwise comparison of two models is reported, along with an inequality sign that reflects the direction of the rejection. The sample size is 105 weeks.

	$A_1(2)$		$A_1(3)$ USV		$A_1(3)$		$A_1(4)$ USV
				bias in weekly $\lvert\Delta Y\rvert$			
6-month	-6.87**		-3.15**		-6.06**		-4.41**
1-year	-1.75*		1.02		-2.21**		-1.32
2-year	2.42**		3.28**		-0.31		0.19
3-year	4.34**		3.55**		0.81		1.07
4-year	5.55**		3.46**		1.75		2.00
5-year	6.19**		3.14**		2.26*		2.66*
7-year	6.47**		2.31*		2.44*		3.19**
10-year	6.07**		1.52		1.99		3.16**
				RMSE of weekly $\lvert\Delta Y\rvert$			
6-month	8.04	$>^{**}$	5.30	$<^{**}$	7.43	$>^{**}$	6.28
1-year	7.60		7.45		7.85		7.60
2-year	9.95	$<^{*}$	10.27		9.74		9.74
3-year	10.80		10.65	$>^{**}$	9.93		10.03
4-year	11.57	$>^{**}$	10.86	$>^{**}$	10.26	$<^{*}$	10.40
5-year	12.08	$>^{**}$	10.95	$>^{**}$	10.55	$<^{**}$	10.73
7-year	12.22	$>^{**}$	10.68		10.56	$<^{**}$	10.85
10-year	11.75	$>^{**}$	10.18		10.16	$<^{**}$	10.52
				bias in $\hat{\sigma}$			
6-month	-8.24**		-3.75**		-6.94**		-5.26**
1-year	-2.98**		0.40		-3.40**		-2.50**
2-year	2.50**		3.65**		-0.80		-0.23
3-year	4.98**		4.18**		0.70		1.02
4-year	6.46**		4.13**		1.85*		2.19**
5-year	7.23**		3.78**		2.49**		3.01**
7-year	7.59**		2.82**		2.75**		3.70**
10-year	7.11**		1.88*		2.23**		3.69**
				RMSE of $\hat{\sigma}$			
6-month	8.82	$>^{**}$	5.28	$<^{**}$	7.68	$>^{**}$	6.41
1-year	5.70	$>^{*}$	5.02	$<^{**}$	6.06	$>^{**}$	5.49
2-year	6.32	$<^{**}$	7.01	$>^{*}$	6.02		5.87
3-year	7.83		7.57	$>^{**}$	6.21		6.21
4-year	9.04	$>^{**}$	7.79	$>^{**}$	6.66		6.75
5-year	9.75	$>^{**}$	7.78	$>^{**}$	7.04	$<^{*}$	7.24
7-year	10.09	$>^{**}$	7.39		7.20	$<^{**}$	7.62
10-year	9.83	$>^{**}$	7.13		7.15	$<^{**}$	7.71

Figure 1

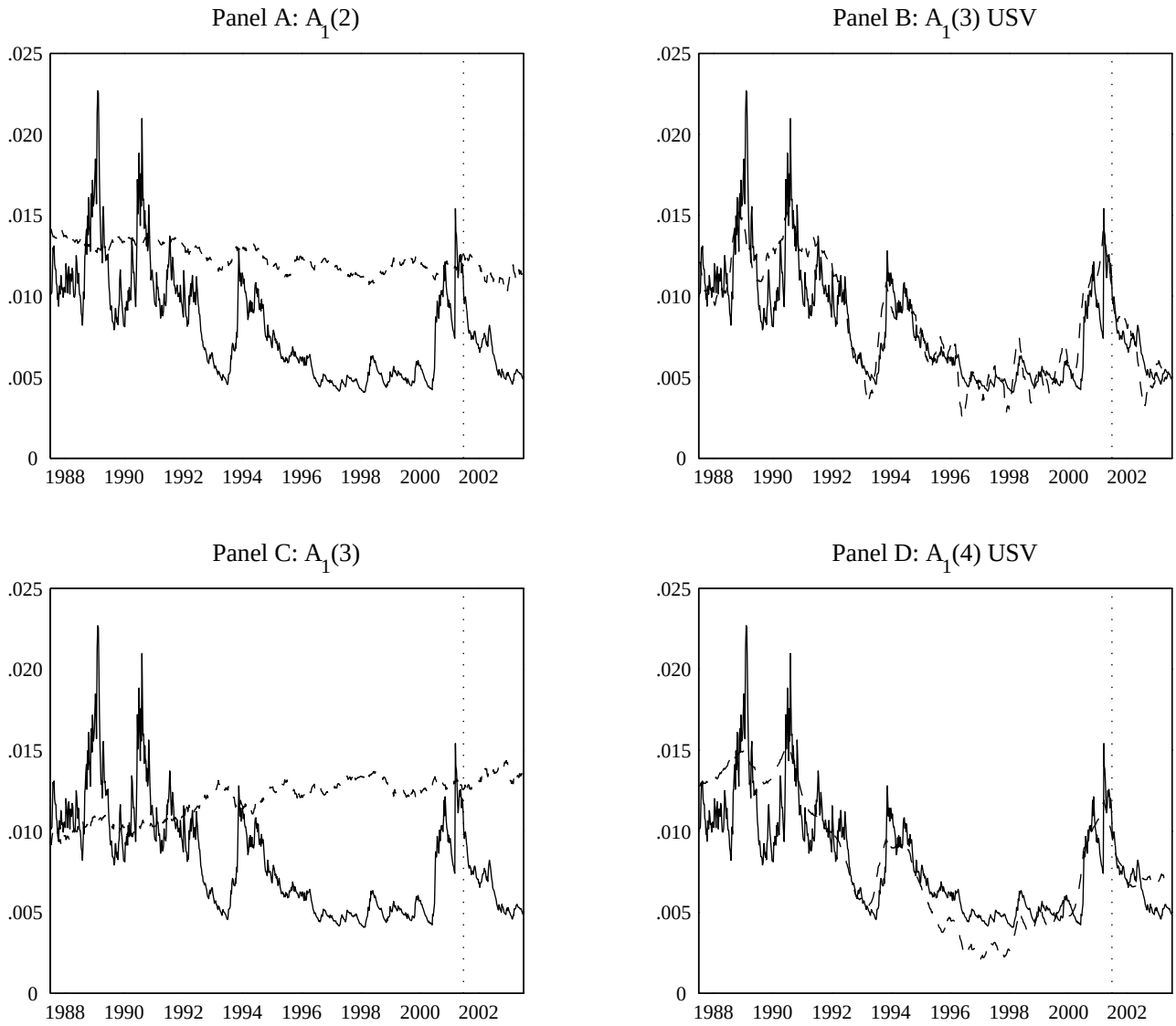
Actual and model-implied curvature



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

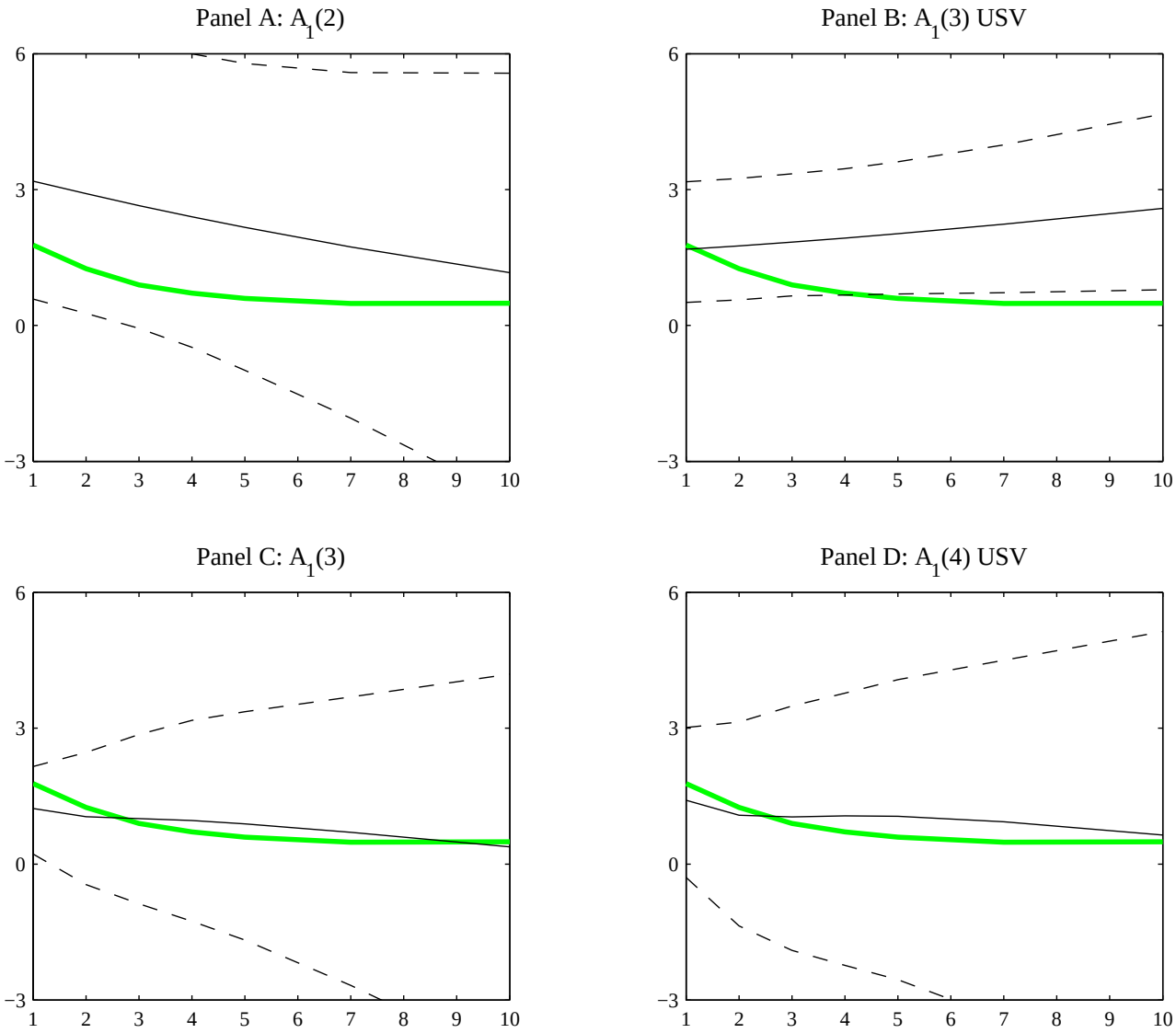
EGARCH and model-implied short rate volatility



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

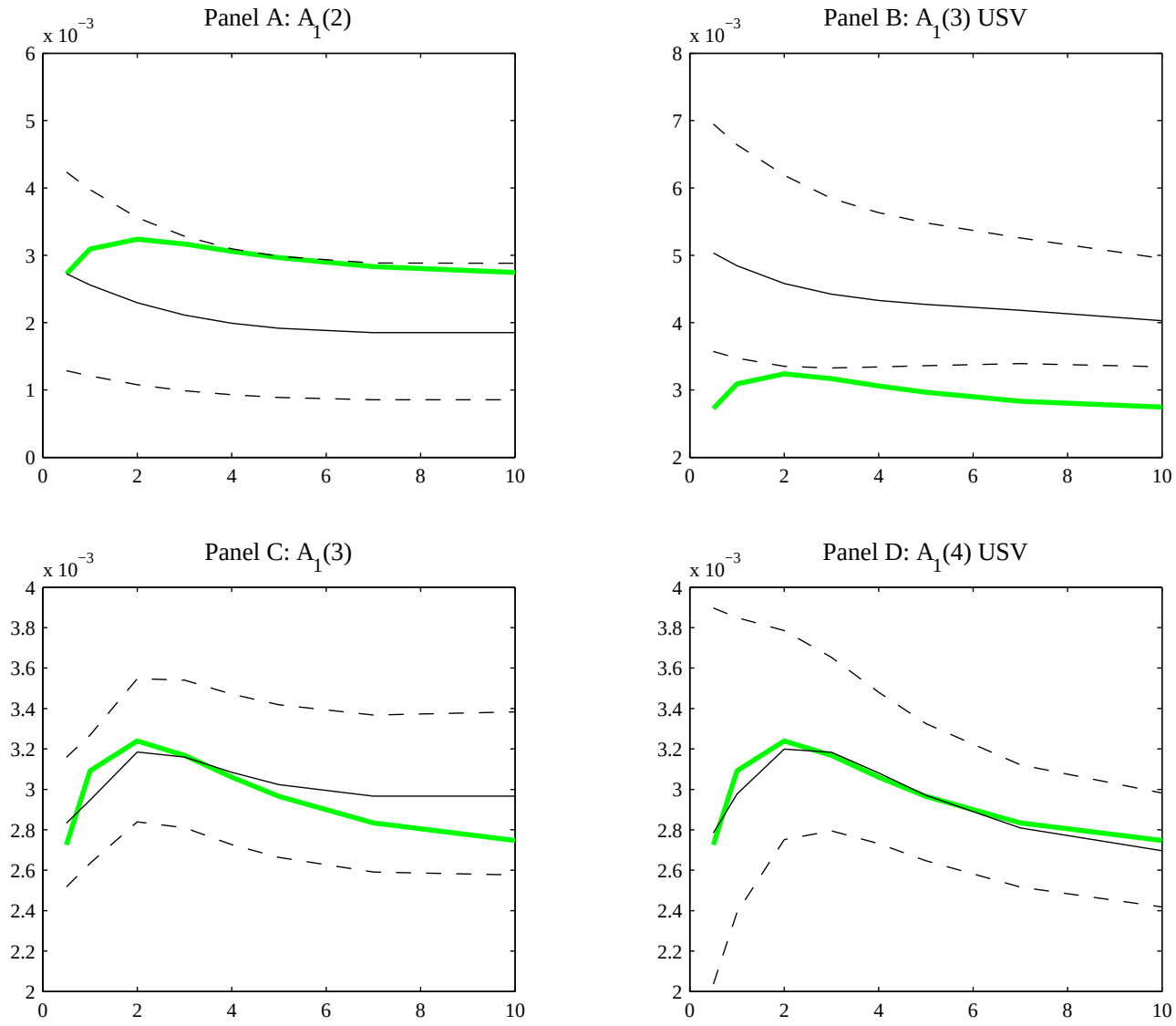
Estimated and model-implied estimates of the slope coefficient
of the regression of yield changes on lagged yield curve slope



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

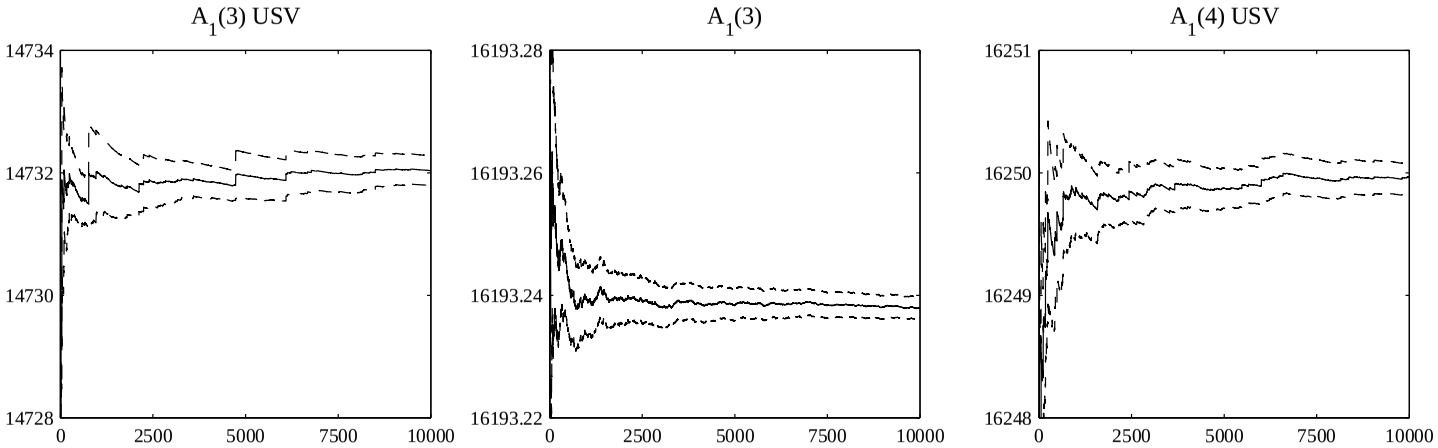
The maturity/volatility relation



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 5

Recursive log likelihoods and 95% confidence intervals



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