

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.

©2004 by Pierre Collin-Dufresne, Robert S. Goldstein, and Christopher S. Jones. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

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, Revised February 2006
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

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

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

1 Introduction

The affine class of term structure models as characterized by Duffie and Kan (DK, 1996) owes much of its popularity to its analytic 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

In this paper, we make two contributions to the affine term structure literature. First, we propose a new representation in which the elements of both the state vector and parameter vector have unique economic interpretations. In contrast, most affine yield representations are written in terms of a *latent* state vector whose elements have no economic interpretation of their own. The advantages of our representation over representations in terms of a latent state vector are discussed below.

As a second contribution, we use this representation to estimate three and four-factor stochastic volatility models. As is well known, 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 variance state variable in the $A_1(3)$ model is unable to play this dual role. As such, we investigate how well an $A_1(4)$ model exhibiting ‘unspanned stochastic volatility’ (USV) performs. USV models break the dual role that the variance state variable plays, in turn allowing it to accurately capture the time series of interest rate volatility. We now discuss these two contributions.

1.1 New representation of affine models

Typically affine term structure models are written in terms of a Markov system of latent state variables $X = \{X_1, \dots, X_n\}$ that describe the entire state of the term structure (see, e.g., Piazzesi (2004) for a survey). One problem with these latent factor models is that the parameter vector $\{\phi\}$ might not be identifiable even if a panel data set of all possible fixed income securities were available. Currently, two approaches have been proposed in the literature to deal with identification. The first approach, due to DK, is to obtain an identifiable model by rotating from the latent state variables to a set of observable zero coupon yields (with distinct finite maturities). Unfortunately, as we discuss below, their approach is often difficult to implement and therefore has not been widely used. Further, for unspanned stochastic volatility models (Collin-Dufresne and Goldstein (2002)) the rotation is not implementable.

The second approach, due to Dai and Singleton (DS, 2000), consists of performing a set of ‘invariant transformations’ that leave security prices unchanged but reduce the number of parame-

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

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

ters.³ After performing several invariant transformations, DS obtain a canonical representation for latent affine term structure models which they refer to as ‘maximal’ in the sense that no additional parameters can be identified even if the prices of all fixed income securities were available.⁴ However, a limitation of latent variable models is that neither the state variables nor the parameters have any economic meaning of their own. As such, to interpret the results of the model (beyond just goodness-of-fit), a rotation to a state vector which is economically meaningful is eventually necessary. Further, as we demonstrate below, the DS canonical representation is only *locally* and not *globally* identifiable. As such, two researchers with the same data can obtain different estimates for the state vector and parameter vector.⁵

Below, we combine insights from both DS and DK to identify an invariant transformation of latent variable affine models where the resulting representation is both tractable and is specified in terms of economically meaningful state variables. 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 economic interpretations such as level, slope and curvature, and (ii) their quadratic covariations. While our rotation is not unique, the choices it offers are intuitive and adaptable to a particular dataset and/or estimation method. Such a representation has several advantages:

First, because the state vector has a unique economic interpretation, both the state vector and the parameters are *globally identifiable*.

Second, since our representation provides simple economic interpretations for both the state variables and the parameters of the model, their values can be directly compared across different countries, sample periods, or even different models. In contrast, parameters and the state variables obtained from a latent representation cannot be compared until a rotation to an economically meaningful representation is performed.⁶

Third, our approach makes clear that the issue of identification rests mainly with the risk-neutral measure. Indeed, our approach allows us to identify how many parameters are identifiable with cross-sectional information only – a concept which we refer to as *Q-maximality*. This issue is important because the tractability of the affine class is mostly in regards to its risk-neutral dynamics.

³DS identify three such types of ‘invariant transformations’: (i) rotation of the state vector \mathcal{T}_A , (ii) Diffusion rescaling \mathcal{T}_D , (iii) Brownian motion rotation \mathcal{T}_O .

⁴As shown in Collin-Dufresne and Goldstein (2002) several ‘maximal’ models are actually not identified if one observes only bond (or yield) data. Indeed, in the presence of unspanned stochastic volatility (USV) the parameters of the drift of volatility typically cannot be identified unless one observes derivative data in addition to bond yields.

⁵Following Rothenberg (1971), a model is *globally identified* if every parameter vector implies a unique probability distribution for observable security prices, i.e. no parameter vectors are observationally equivalent. A model is only *locally identifiable* when there exist multiple parameter vectors that imply the same distribution but these parameter vectors are not “close.” A model is *unidentified* when all open sets around a given parameter vector include another vector that is observationally equivalent to it.

⁶It is often the case that state variables are highly correlated with one or more principal components, and thus researchers interpret the state variable as such. However, such interpretations are approximate at best. Furthermore, as shown by Duffee (1996) Tang and Xia (2005), the weights of such principal components change over time and across countries. Hence, attempting to compare models and/or parameters through their implied principal component dynamics is at best suggestive and likely somewhat misleading.

Indeed, some researchers (e.g., Duarte (2003)) have combined affine risk-neutral dynamics with non-affine historical measure dynamics in order to improve goodness-of-fit. Our approach makes it simple to determine the number of risk-neutral parameters that are identifiable. Once this is done and the state vector is identified, it is a trivial matter to determine which risk premia parameters are identifiable.

Our representation also has several advantages over that of DK. First, it is easy to implement. In contrast, as we discuss below, DK's yield factor representation requires solving systems of non-linear equations that are often not solvable in closed form. Second, our representation works for unspanned stochastic volatility (USV) models, for which there does not exist a one-to-one mapping between state variables and yields. Without such a mapping, the DK approach is not implementable. Third, for those models that exhibit USV, this representation isolates those parameters which are not identifiable from bond prices alone. Finally, this representation simplifies the form of the parameter constraints imposed by USV, in turn facilitating empirical investigation (we discuss this further below). One potential advantage of DK's representation is that its state vector is composed of directly-observable yields rather than just *theoretically*-observable yields whose values need to be approximated. We note, however, that even DK's approach typically requires that zero-coupon yields be approximated from coupon yields. Further, we show using simulated data that it is possible to obtain accurate estimates of our state variables that are insensitive to the method used.⁷

1.2 Identifying a failure of three-factor affine models

The second contribution of our paper adds to the growing literature that documents empirical failings of three-factor affine models. This previous literature has 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 addresses 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.⁸ 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. Furthermore, 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

⁷This may prove useful from a practical perspective in that, because we can estimate a time series for the state vector before attempting to identify parameter estimates, we can come up with a good first guess for the parameter vector, in turn simplifying the search over an often very large dimensional parameter space.

⁸See also Chacko (1997).

properties of the term structure. Here, however, the trade-off involves second-order moments.⁹ 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’) is strongly *negatively* correlated with volatilities estimated using rolling windows or a standard GARCH model applied to the time series of the 6-month rate. Furthermore, the strong in-sample fit of that model breaks down following the end of the estimation period, suggesting deep misspecification.

We interpret these findings as evidence that the $A_1(3)$ model cannot simultaneously describe the 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.¹⁰

We emphasize that our findings may have implications beyond the affine class of models. Indeed, using model-insensitive proxies for interest rate level, slope, curvature, and a GARCH estimate for volatility, we find that these four series are (unconditionally, anyway) weakly related, suggesting that there may be no three-factor model that can simultaneously capture these four features of the term structure.¹¹ Our results may therefore explain Ahn, Dittmar, and Gallant’s (2001) finding that three-factor quadratic term structure models also have difficulty reproducing yield curve volatility patterns.

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

⁹Note 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).

¹⁰Bikbov and Chernov (2004) also investigate three-factor affine models. They find that the estimated model dynamics are highly dependent on whether or not they use options data (in addition to yields) to fit their models. When both options and yield data are used, we suspect that the variance state variable will be more closely related to interest rate variance and less to the shape of the yield curve. These results, however, are unrelated to our findings that interest rate volatility is weakly correlated with the level, slope and curvature of the yield curve. As such, we suspect that their variance state vector still will not be able to play the dual role that affine models predict.

¹¹We 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.

possess. In particular, these models impose parameter constraints so that the variance cannot be determined from a linear combination of yields. 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 some standard estimation techniques, which rely on the ‘invertibility’ of the term structure with respect to the latent factors, cannot be implemented. Instead, we write term structure dynamics in nonlinear state space form and estimate the parameters of the models using Bayesian Markov chain Monte Carlo.

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. Indeed, in addition to the $A_1(3)$ model generating poor estimates for interest rate variance, it also produces out-of-sample cross sectional errors that are about twice the size as those of the $A_1(4)$ USV model. 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 (1991). 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.¹² 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.

The rest of the paper is as follows. In Section 2 we provide a general approach for deriving maximal affine models with observable state variables. In Section 3 we characterize the maximal $A_1(3)$ and $A_1(4)$ models exhibiting USV. In Section 4 we describe an estimation methodology that remains valid under USV, while Section 5 includes all empirical results. We conclude in Section 6.

2 Maximal affine models with ‘theoretically observable’ state variables

For what follows, it is important to distinguish between several related concepts: identification, identifiability, and maximality.¹³ In the applied literature, the concept of ‘identification’ deals with the issue of whether the state vector and parameter vector can be inferred from a *particular* data set. In contrast, below we will use the concept of ‘identifiability’ to deal with the issue of whether the state vector and parameter vector can be inferred from observing *all conceivable financial data*. (i.e., all possible securities, as frequently as necessary). A ‘maximal’ model, as defined by DS, is the most general model (within a class) that is identifiable given sufficiently informative data. We emphasize that maximality is a theoretical concept in that DS determine maximality by considering a series of invariant rotations that leave unaffected the fundamental PDE that security prices satisfy

¹²This 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.

¹³We thank the referee for making us aware of the distinction between these concepts.

without ever discussing what securities are available to the econometrician. Below, we follow their lead and interpret identifiability and maximality in a theoretical sense. That is, we identify which parameters are identifiable if the prices of all fixed income securities were observed as often as necessary.

It is also helpful to introduce the concept of *theoretical observable* (as opposed to ‘latent’) variable. In most econometrics, latent variables are considered to be those that are unobserved by the econometrician, regardless of the interpretation of those variables. In this paper, we instead define a latent variable to be a variable which has no intrinsic economic meaning. That is, it has no physical interpretation independent of the values of other state variables and/or parameters of the model. In contrast, a *theoretically observable* state variable is one that would be directly observable, without using a model, if all conceivable fixed income securities data were available. As such, theoretically observable variables possess an economic meaning independent of the model or its parameter values. Two important examples of variables which are theoretically observable (and not latent) are the spot rate and its volatility.

Possibly the most dangerous aspect of latent variables is that researchers sometimes attempt to attribute to them an economic interpretation when in fact they have none. A very elegant example illustrating this concern, due to Babbs and Nowman (BN, 1999), is the following. Consider the two factor Gaussian (‘maximal’ $A_0(2)$ in DS taxonomy) model:

$$dr_t = \kappa_r (\theta_t - r_t) dt + \sigma_r dz_r(t) \quad (1)$$

$$d\theta_t = \kappa_\theta (\bar{\theta} - \theta_t) dt + \sigma_\theta dz_\theta(t), \quad (2)$$

with $dz_r dz_\theta = \rho dt$. BN show that one can find an invariant transformation of the model by defining another latent variable θ_t^* by:

$$\theta_t^* = \left(1 - \frac{\kappa_r}{\kappa_\theta}\right) r_t + \frac{\kappa_r}{\kappa_\theta} \theta_t \quad (3)$$

so that the dynamics of the system become:

$$dr_t = \kappa_\theta (\theta_t^* - r_t) dt + \sigma_r dz_r(t) \quad (4)$$

$$d\theta_t^* = \kappa_r (\bar{\theta} - \theta_t^*) dt + \sigma_{\theta^*} dz_{\theta^*}(t). \quad (5)$$

Hence, even though the model is maximal in the sense of DS, two empirical researchers could estimate different parameters and state variables using the exact same data set. In particular, one cannot distinguish the short rate reverting to θ with speed κ_r from the short rate reverting to θ^* with speed κ_θ . This duplicity is especially problematic when one wants to give economic meaning to θ . For example, this variable has been previously interpreted as a long-run target rate set by the central bank (e.g., Jegadeesh and Pennacchi (1996), Backus et al. (1994)). Admittedly, complete identification can be obtained by imposing additional inequality constraints on the $\{\kappa\}$. We emphasize, however, that such constraints do not change the fact that the state variable θ has no economic meaning.

Indeed, such constraints only serve to make it more likely that some economic interpretation be incorrectly attributed to θ !

In the parlance of econometric theory (e.g. Rothenberg (1971)), maximal latent variable models are only *locally* and not *globally* identifiable. We emphasize that the insights of BN are not just relevant for Gaussian models. Indeed, the same transformation can be applied to the maximal $A_1(3)$ model of DS (2000) in its $Ar(\cdot)$ representation (equation (23), pg. 1951), to show that the ‘central tendency’ defined by DS is not uniquely determined. Further, the same issue also arises for the canonical AY representation of DS (pg. 1948).¹⁴

The example above is particularly salient because it emphasizes the difference between latent and theoretically observable state variables. In particular, the state variable r is by definition the short end of the term structure, and is therefore theoretically observable in that it cannot be changed without necessarily changing the values of some fixed income securities (in particular, those with very short, but finite maturities). In contrast, because θ is latent, its value can be replaced by θ^* and, provided the parameters are adjusted appropriately (e.g., $\kappa_\theta \leftrightarrow \kappa_r \dots$), the prices of *all* fixed income securities remain unchanged.

With these concerns in mind, we now search for a tractable affine framework where the state vector has a clear economic interpretation. Mostly following the notation of DK and DS, the risk-neutral dynamics of a Markov state vector X within an affine framework can be specified by as:

$$dX(t) = \mathcal{K}^Q \left(\Theta^Q - X(t) \right) dt + \Sigma \sqrt{S(t)} dZ^Q(t), \quad (9)$$

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

$$S_{ii}(t) = \alpha_i + \beta_i^\top X(t). \quad (10)$$

The spot rate is an affine function of X :

$$r(t) = \delta_0 + \delta_x^\top X(t), \quad (11)$$

where δ_x is an N dimensional vector. Assuming the system is admissible (i.e., that the stochastic differential equation admits a unique strong solution¹⁵), then zero coupon bond prices take the form:

$$P(t, \tau) = e^{A(\tau) - B(\tau)^\top X(t)}, \quad (12)$$

¹⁴The AY canonical $A_0(2)$ model of DS is given by:

$$r(t) = \bar{r} + \sigma_1 X_1(t) + \sigma_2 X_2(t) \quad (6)$$

$$dX_1(t) = -\kappa_{11} X_1(t) dt + dz_1(t) \quad (7)$$

$$dX_2(t) = -(\kappa_{21} X_1(t) + \kappa_{22} X_2(t)) dt + dz_2(t). \quad (8)$$

It is straightforward to show that the BN model given in equations (1) and (2) is an invariant transformation of the canonical AY model above, where, in particular, we have the relation $\kappa_{11} = \kappa_r$ and $\kappa_{22} = \kappa_\theta$. But, following the argument leading to the equivalent representation in (4) and (5) there is an equivalent AY representation with $\kappa_{22} = \kappa_r$ and $\kappa_{11} = \kappa_\theta$. This shows that the AY canonical representation is not globally identifiable.

¹⁵Sufficient conditions are given in Duffie and Kan (1996).

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

$$\frac{dA(\tau)}{d\tau} = -\Theta^{Q\top} \mathcal{K}^{Q\top} B(\tau) + \frac{1}{2} \sum_{i=1}^N \left[\Sigma^\top B(\tau) \right]_i^2 \alpha_i - \delta_0 \quad (13)$$

$$\frac{dB(\tau)}{d\tau} = -\mathcal{K}^{Q\top} B(\tau) - \frac{1}{2} \sum_{i=1}^N \left[\Sigma^\top B(\tau) \right]_i^2 \beta_i + \delta_x, \quad (14)$$

and the initial conditions:

$$A(0) = 0, \quad B(0) = 0. \quad (15)$$

Defining bond yields $Y(t, \tau)$ via $P(t, \tau) = e^{-\tau Y(t, \tau)}$, we see from equation (12) that yields are affine in the state variables:

$$Y(t, \tau) = -\frac{A(\tau)}{\tau} + \frac{B(\tau)^\top}{\tau} X(t). \quad (16)$$

DK use this observation to suggest the possibility of rotating the system from a latent state vector X to observable yields Y (with N arbitrary finite maturities). Unfortunately, such a rotation is not tractable because it involves the functions $A(\cdot), B(\cdot)$ which are, in general, not known in closed form. Further, in those cases where the model exhibits USV, such a rotation is not possible because the state variables cannot be backed out from yields alone.

Instead, we propose an alternative to DK's approach to obtain a representation in terms of theoretically observable state variables. In particular, we perform a Taylor series expansions of both the yield curve and the time-dependent coefficients $A(\tau)$ and $B(\tau)$ given in equation (16) around $\tau = 0$.¹⁶

$$Y(t, \tau) = Y(t, 0) + \tau \partial_{\tau=0} Y(t, \tau) + \left(\frac{\tau^2}{2!} \right) \partial_{\tau=0}^2 Y(t, \tau) + \dots \quad (17)$$

$$A(\tau) = A(0) + \tau \partial_{\tau=0} A(\tau) + \left(\frac{\tau^2}{2!} \right) \partial_{\tau=0}^2 A(\tau) + \dots \quad (18)$$

$$B(\tau) = B(0) + \tau \partial_{\tau=0} B(\tau) + \left(\frac{\tau^2}{2!} \right) \partial_{\tau=0}^2 B(\tau) + \dots \quad (19)$$

Using the initial conditions in equation (15), and collecting terms of the same order τ , we find from equation (16) the following relation between the terms of the expansions:

$$\begin{aligned} Y_n(t) &\equiv \partial_{\tau=0}^n Y(t, \tau) \\ &= \frac{1}{n+1} \left(-\partial_{\tau=0}^{n+1} A(\tau) + \sum_{i=1}^N \partial_{\tau=0}^{n+1} B_i(\tau) X_i(t) \right) \quad \forall n = 0, 1, 2, \dots \end{aligned} \quad (20)$$

Equation (20) 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 . Further, as we illustrate below, all the coefficients in the transformation from the vector $\mathbf{X} = \{X_1, X_2, \dots\}$ to $\mathbf{Y} = \{Y_0, Y_1, \dots\}$ can be

¹⁶To simplify notation, we define $\partial_{\tau=c}^n f(t, \tau) := \left. \frac{\partial^n f(t, \tau)}{\partial \tau^n} \right|_{\tau=c}$ for any function $f(t, \tau)$.

found explicitly by a recursion obtained by differentiating repeatedly the system of ODE's given in equations (13) and (14), and making use of the boundary conditions in equation (15).

For illustration, we provide the expression for the loadings in the definition of $\{Y_0, Y_1, Y_2\}$. From equations (13) to (15), we have:

$$\partial_{\tau=0} A(\tau) = -\delta_0 \quad (21)$$

$$\partial_{\tau=0} B(\tau) = \delta_x \quad (22)$$

$$\partial_{\tau=0}^2 A(\tau) = -\Theta^{Q\top} \mathcal{K}^{Q\top} \delta_x \quad (23)$$

$$\partial_{\tau=0}^2 B(\tau) = -\mathcal{K}^{Q\top} \delta_x \quad (24)$$

$$\partial_{\tau=0}^3 A(\tau) = \Theta^{Q\top} \mathcal{K}^{Q\top} \mathcal{K}^{Q\top} \delta_x + \sum_{i=1}^N [\Sigma^\top \delta_x]_i^2 \alpha_i \quad (25)$$

$$\partial_{\tau=0}^3 B(\tau) = \mathcal{K}^{Q\top} \mathcal{K}^{Q\top} \delta_x - \sum_{i=1}^N [\Sigma^\top \delta_x]_i^2 \beta_i. \quad (26)$$

Plugging these into equation (20) and identifying the terms we find:

$$\begin{aligned} Y_0(t) &= \delta_0 + \delta_x^\top X(t) \\ &\equiv r(t) \end{aligned} \quad (27)$$

$$\begin{aligned} Y_1(t) &= \frac{1}{2} \delta_x^\top \mathcal{K}^Q (\Theta^Q - X(t)) \\ &= \frac{1}{2} \frac{d}{dt} E_t^Q [dr(t)] \\ &\equiv \frac{1}{2} \mu^Q(t) \end{aligned} \quad (28)$$

$$\begin{aligned} Y_2(t) &= \frac{1}{3} \left(-\delta_x^\top \mathcal{K}^Q \mathcal{K}^Q (\Theta^Q - X(t)) - \sum_{i=1}^N [\Sigma^\top \delta_x]_i^2 (\alpha_i + \beta_i^\top X(t)) \right) \\ &= \frac{1}{3} \frac{d}{dt} \left(E_t^Q [d\mu^Q(t)] - (dr(t))^2 \right) \\ &\equiv \frac{1}{3} \left(\frac{1}{dt} E_t^Q [d\mu^Q(t)] - V(t) \right). \end{aligned} \quad (29)$$

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

The above suggests a natural transformation from the latent variables \mathbf{X} to the theoretically observable state vector \mathbf{Y} (or a subset of it). We emphasize that the latter is only theoretically observable because it is the vector of Taylor expansion coefficients of the term structure at zero. That is, the state vector consists of yields and sums of yields of infinitesimal maturity. Of course, in practice only a finite maturity bonds are actually observable. However, from the point of view of theoretical identification, this is not an issue. In fact, this is similar to DK where continuously

compounded yields on zero-coupon bonds may not be observable in practice either, but instead must be estimated using some interpolation scheme from coupon bonds or from swap quotes.

A key advantage of this representation is that by construction it is globally identifiable. In particular, given the prices of all fixed income securities, all state variables and their risk-neutral parameters are uniquely identified. Hence, this model does not possess multiple solutions as do the latent variable representations. Furthermore, our simulation results in the next section show that it is possible to get very accurate estimates of the \mathbf{Y} state variables that are extremely insensitive to the model used to extract them. This suggests the rotation may have practical advantages in addition to the property of being globally identifiable.¹⁷

We emphasize that the choice of representation is not unique. Any invariant transformation from a maximal latent variable model to a theoretically observable model will yield a globally identifiable model. In particular, for the investigation of stochastic volatility models which we pursue below, we find it useful to combine elements of the state vector \mathbf{Y} and a subset of the quadratic (co-)variations of some of the \mathbf{Y} state variables (which we call \mathbf{V}), rather than to choose the entire state vector from \mathbf{Y} alone. Because of the properties of the quadratic-covariation process in continuous time, this preserves the observability of the state vector. We can thus provide a definition of an observable representation:

Definition 1 *A theoretically observable Q -Representation is an invariant transformation¹⁸ of the latent state vector X given in equation (9) to a N -dimensional state vector $\mathbf{H} \equiv [\hat{\mathbf{Y}}, \hat{\mathbf{V}}]$ that combines elements of \mathbf{Y} and \mathbf{V} and contains the short rate $Y_0 = r$.*

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

- By definition of the vectors \mathbf{Y} and \mathbf{V} , the state variables in this representation are theoretically 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.¹⁹
- Since \mathbf{H} is an invariant transformation of X it has jointly Markov affine dynamics.
- The representation is not unique: there are many ‘invariant transformations’ that preserve ‘observability.’

¹⁷One could think of working directly with the model-independent estimates of the state variables. This could be especially useful for econometric work involving physical measure dynamics, such as forecasting or hedging. We leave this for future research.

¹⁸The notion of invariant transformation is defined in DS (2000). See also footnote 3.

¹⁹Knowledge of the risk-neutral short rate process is sufficient to describe prices of all fixed-income derivatives. See, e.g., Glasserman and Jin (2001).

- The representation is independent of the risk-premium structure. Hence, this representation emphasizes that *theoretical observability* is intimately tied to the risk-neutral dynamics of the state variables.

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 $\theta^Q(t) \equiv E^Q[d\mu^Q(t)]/dt$ 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 (27)-(29) 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 from $\mathbf{H}(t) = (r(t), \mu^Q(t), \theta^Q(t))$ to $\mathbf{H} = (r, \mu^Q, V)$, as we demonstrate in Section 3.1 below. Note that the variance state variable is theoretically observable as well in that it has a physical interpretation independent of the model's parameter values, and in particular, can be estimated from the quadratic variation of the time series of the short rate.

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

Conceptually, the rotation of the state vector to theoretically observable variables is similar to the original idea of DK (1996), who rotate a latent state vector to an observable state vector defined in terms of yields of finite maturities. However, there are several cases for which their approach is difficult or even impossible to implement. First, for the subset of models exhibiting USV, the rotation proposed by DS fails since 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 Riccati equations, which are not generally known in closed-form.

To illustrate the difficulties in implementing the DK approach, here we consider a two-factor Gaussian (i.e., non-USV) model of the short rate r and a latent variable x :

$$dr_t = (\alpha_r + \beta_{rr} r_t + \beta_{rx} x_t) dt + \sigma_r dZ_r^Q(t) \quad (30)$$

$$dx_t = (\alpha_x + \beta_{xr} r_t + \beta_{xx} x_t) dt + \sigma_x dZ_x^Q(t), \quad (31)$$

where $dZ_r^Q(t) dZ_x^Q(t) = \rho_{r,x} dt$. 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 securities are available. Consistent with the insights of DS and DK,

we show below that only 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.:

$$dr_t = \left(\hat{\alpha}_r + \hat{\beta}_{rr} r_t + \hat{\beta}_{ry} Y(t, \hat{\tau}) \right) dt + \sigma_r dZ_{r,t}^Q \quad (32)$$

$$dY(t, \hat{\tau}) = \left(\hat{\alpha}_y + \hat{\beta}_{yr} r_t + \hat{\beta}_{yy} Y(t, \hat{\tau}) \right) dt + \sigma_y dZ_{y,t}^Q, \quad (33)$$

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:

$$\hat{A}(\hat{\tau}) = 0, \quad \hat{B}_r(\hat{\tau}) = 0 \quad \text{and} \quad \hat{B}_y(\hat{\tau}) = \hat{\tau}. \quad (34)$$

Although these constraints are non-linear, one would (correctly) suspect that they will lead three restrictions on the parameters in equations (32)-(33). Hence, while the latent state vector representation (equations (30) and (31)) seems to suggest that there are nine free risk-neutral parameters, by rotating to an observable vector, we see that there are only six. Unfortunately, this ‘yield-based’ approach proposed by Duffie and Kan (1996) is often intractable, because the coefficients $A(\tau)$ and $B(\tau)$ are generally not known in closed-form²⁰, making it difficult to impose the constraints implied in equation (34).

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 $\{Y_0, Y_1 \dots\}$. 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. In the example above, our representation 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 (30), we find $\mu^Q = \alpha_r + \beta_{rr} r_t + \beta_{rx} x_t$. Hence, the dynamics of the system become:

$$dr(t) = \mu^Q(t)dt + \sigma_r dZ_r^Q(t) \quad (35)$$

$$d\mu^Q(t) = (\beta_0 + \beta_1 r_t + \beta_2 \mu^Q(t))dt + \sigma_m dZ_m^Q(t), \quad (36)$$

²⁰We 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.

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:

$$\beta_0 = \beta_{rx}\alpha_x - \beta_{xx}\alpha_r \quad (37)$$

$$\beta_1 = \beta_{rx}\beta_{xr} - \beta_{xx}\beta_{rr} \quad (38)$$

$$\beta_2 = \beta_{rr} + \beta_{xx} \quad (39)$$

$$\sigma_m^2 = \beta_{rx}^2\sigma_x^2 + \beta_{rr}^2\sigma_r^2 + 2\rho_{r,x}\beta_{rx}\sigma_x\beta_{rr}\sigma_r \quad (40)$$

$$\rho_{r,m}\sigma_m\sigma_r = \rho_{r,x}\beta_{rx}\sigma_r\sigma_x + \beta_{rr}\sigma_r^2. \quad (41)$$

Note that with ‘no effort’ our representation demonstrates that only 6 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 (30)-(31) that leaves the left hand side of equations (37)-(41) unchanged generates a short rate process which is path-by-path identical to that of model (35)-(36). Consequently, both models are observationally equivalent conditional on observing all possible fixed-income securities. In other words, only the left hand side of equations (37)-(41) are separately identifiable from fixed-income derivatives data.

2.2 A constructive proof showing that a theoretically identifiable state vector guarantees parameter identification

DS consider a set of invariant rotations that reduce the number of parameters while not affecting the prices of fixed income securities. Further, they conjecture, but do not prove, that no additional invariant rotations exist. Our representation in terms of theoretical observables demonstrates that their conjecture is correct. Indeed, we can isolate a set of infinitesimal-maturity fixed income securities whose prices will uniquely identify both the state vector and all risk-neutral parameters.

In order to demonstrate that a theoretical observable state vector guarantees risk-neutral parameter identification, here we consider the example in equations (30)-(36) above. As noted in equations (27)-(28), the state vector $\{r, \mu^Q\}$ is identifiable from observation of the yield curve at infinitesimal maturities. Further, since all agents agree on these variables, contracts could, in theory, be written on the short rate, its volatility, on the slope of the yield curve, and on its volatility. If such contracts were traded, then agents could observe the following ‘futures prices’ with infinitesimal maturities: $E_t^Q [(\Delta r_t)^2]$, $E_t^Q [(\Delta \mu_t^Q)^2]$, $E_t^Q [\Delta r_t \Delta \mu_t^Q]$. These contracts would directly identify $\sigma_r, \sigma_m, \rho_{r,m}$. Finally, the futures prices

$$\begin{aligned} F(t, \Delta) &= E_t^Q[\Delta \mu^Q(t)] \\ &= (\beta_0 + \beta_1 r(t) + \beta_2 \mu^Q(t)) \Delta t \end{aligned} \quad (42)$$

from three sets of (distinct) observations $\hat{F} \equiv \{F(t_1), F(t_2), F(t_3)\}$ allow us to infer the parameters $(\beta_0, \beta_1, \beta_2)$.²¹

²¹We emphasize that there is no time series information here. In particular, we do not need to know the temporal ordering of these three observations.

Of course, both the availability of derivative data and observability of the state variables are crucial to our argument. The specific claims (i.e., the infinitesimal maturity futures prices), whose existence we postulate, help make the argument transparent. In practice, we find that finite maturity bonds alone are sufficient for identifying all risk-neutral parameters when the model does not exhibit USV. When the model exhibits USV, bonds and simple fixed income derivatives such as caps are sufficient.

Note further that our discussion is valid irrespective of the risk-premium specification chosen. Starting from such an observable representation, any risk-premium specification that is ‘reasonable’ in the sense that it leads to a P -measure state variable process identifiable based on its observed time series data (e.g., using vector auto-regression if the P -dynamics are affine as well), will lead to a model that is maximal in the sense of DS (2000). Therefore, our approach can also be used for models with more general, non-affine, risk-premium specifications (e.g., Duarte (2004)) or for the case of models with jump diffusion. In contrast, the DS approach to identification is based on the idea that a model is identifiable when the Jacobian of the likelihood function is non-singular. That is, when the likelihood function possesses local maxima. It thus seems inherently tied to the risk-premium specification.

The next section documents via simulation that the model-independent ‘observability’ of our state variables may also be of practical interest.

2.3 Model-insensitive estimation of the state variables

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

In practice, however, we rarely observe the entire (continuous) term structure of zero-coupon yields. Rather, we only observe discrete points along the curve. Further, there may be some noise resulting from, e.g., bid-ask spread and non-synchronous trading. To investigate how this would affect the model-independent recovery of the state variables, we perform the following experiment. We simulate a two factor $A_2(2)$ model using the estimates of Duffie and Singleton (1997). We sample 10 years of weekly data and use a set of maturities typical of those used in the term structure literature, namely $\{0.5, 1, 2, 5, 7, 10\}$ years. Then we add i.i.d. noise with either 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:

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

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

We emphasize that we have made no particular effort to find an appropriate interpolation procedure. Rather, we have used the simplest available procedures, and did not try any others. These first results thus seem very promising. The first state variable can be recovered very accurately without much effort from available data. The second state variable can be recovered quite accurately with an appropriate interpolation/extrapolation procedure.²² 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 Stochastic volatility

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 misspecified 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 allow for only one factor to 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.

²²We 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.

3.1 Observable $A_1(3)$ 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:

$$dv = (\gamma_v - \kappa_v v) dt + \sigma_v \sqrt{v} dZ_3^Q \quad (43)$$

$$d\theta = [\gamma_\theta - \kappa_\theta \theta - \kappa_{\theta r} r - \kappa_{\theta v} v] dt + \sigma_{\theta r} \sqrt{\alpha_r + \alpha_v v} dZ_1^Q + \sqrt{\beta_\theta + \beta_v v} dZ_2^Q + \sigma_{\theta v} \sqrt{v} dZ_3^Q \quad (44)$$

$$dr = [\gamma_r - \kappa_r r - \kappa_{r\theta} \theta - \kappa_{rv} v] dt + \sqrt{\alpha_r + \alpha_v v} dZ_1^Q + \sigma_{r\theta} \sqrt{\beta_\theta + \beta_v v} dZ_2^Q + \sigma_{rv} \sqrt{v} dZ_3^Q. \quad (45)$$

Further, since we are interested in models where the short rate displays stochastic volatility we assume that at least one of the terms $(\alpha_v, \sigma_{r\theta}^2 \beta_v, \sigma_{rv}^2)$ is positive. DS demonstrate that this model is not identifiable, 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 a latent state vector (r, θ, v) to the theoretically observable state vector (r, μ^Q, V) defined by:²³

$$\mu^Q = \gamma_r - \kappa_r r - \kappa_{r\theta} \theta - \kappa_{rv} v \quad (46)$$

$$V = \alpha_r + \sigma_{r\theta}^2 \beta_\theta + (\alpha_v + \sigma_{r\theta}^2 \beta_v + \sigma_{rv}^2) v \quad (47)$$

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 the definitions in equations (46) and (47) with the original dynamics of (43)-(45), to obtain:

$$dV_t = (\gamma_V - \kappa_V V_t) dt + \sigma_V \sqrt{V_t - \psi_1} dZ_1^Q(t) \quad (48)$$

$$dr_t = \mu_t^Q dt + \sigma_1 \sqrt{V_t - \psi_1} dZ_1^Q(t) + \sqrt{\sigma_2^2 V_t - \psi_2} dZ_2^Q(t) + \sqrt{\sigma_3^2 V_t - \psi_3} dZ_3^Q(t) \quad (49)$$

$$d\mu_t^Q = (m_0 + m_r r_t + m_\mu \mu_t^Q + m_V V_t) dt + \nu_1 \sqrt{V_t - \psi_1} dZ_1^Q(t) + \nu_2 \sqrt{\sigma_2^2 V_t - \psi_2} dZ_2^Q(t) + \nu_3 \sqrt{\sigma_3^2 V_t - \psi_3} dZ_3^Q(t), \quad (50)$$

where by definition of V_t we have:

$$\sigma_1^2 + \sigma_2^2 + \sigma_3^2 = 1 \quad (51)$$

$$\sigma_1^2 \psi_1 + \psi_2 + \psi_3 = 0. \quad (52)$$

The model is admissible if²⁴

$$\gamma_V \geq \kappa_V \psi_1 \quad (53)$$

$$\psi_1 \geq \max\left(\frac{\psi_2}{\sigma_2^2}, \frac{\psi_3}{\sigma_3^2}\right). \quad (54)$$

²³Since we have restricted ourselves to model where the short rate displays stochastic volatility (i.e., at least one of $(\alpha_v, \sigma_{r\theta}^2 \beta_v, \sigma_{rv}^2)$ is positive), 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.

²⁴Note 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.

Note that all the parameters in equations (48)-(50) above can be expressed as combinations of parameters appearing in equations (43)-(45) above that are not separately identifiable. For example, for the volatility dynamics we have:²⁵

$$\psi_1 = \alpha_r + \beta_\theta \sigma_{r\theta}^2 \quad (55)$$

$$\gamma_V = (\alpha_r + \beta_\theta \sigma_{r\theta}^2) \kappa_v + (\alpha_v + \beta_v \sigma_{r\theta}^2 + \sigma_{rv}^2) \gamma_v \quad (56)$$

$$\kappa_V = \kappa_v \quad (57)$$

$$\sigma_V = \sigma_v \sqrt{\alpha_v + \beta_v \sigma_{r\theta}^2 + \sigma_{rv}^2}. \quad (58)$$

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 (48)-(50). 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:

$$\begin{bmatrix} \frac{1}{dt} \mathbf{E}^Q [dr] \\ \frac{1}{dt} \mathbf{E}^Q [d\mu^Q] \\ \frac{1}{dt} \mathbf{E}^Q [dV] \end{bmatrix} = \begin{bmatrix} \mu^Q \\ m_0 + m_r r + m_\mu \mu^Q + m_V V \\ \gamma_V - \kappa_V V \end{bmatrix} \quad (59)$$

$$\Sigma^2 = \begin{bmatrix} V & c_0 + c_V V & \sigma_1 \sigma_V (V - \psi_1) \\ c_0 + c_V V & \sigma_0^\mu + \sigma_V^\mu V & \nu_1 \sigma_V (V - \psi_1) \\ \sigma_1 \sigma_V (V - \psi_1) & \nu_1 \sigma_V (V - \psi_1) & \sigma_V^2 (V - \psi_1) \end{bmatrix}, \quad (60)$$

where, by definition

$$c_V = \sigma_1 \nu_1 + \sigma_2^2 \nu_2 + \sigma_3^2 \nu_3 \quad (61)$$

$$\sigma_V^\mu = \nu_1^2 + \nu_2^2 \sigma_2^2 + \nu_3^2 \sigma_3^2 \quad (62)$$

$$\sigma_0^\mu = -(\nu_1^2 \psi_1 + \nu_2^2 \psi_2 + \nu_3^2 \psi_3) \quad (63)$$

$$c_0 = -(\sigma_1 \nu_1 \psi_1 + \nu_2 \psi_2 + \nu_3 \psi_3). \quad (64)$$

We note that while the SDE notation of equations (48)-(50) is useful, from an economic point of view only the left hand side of equations (61)-(64) are truly separately identifiable. For example, ν_2 and ν_3 are perfectly symmetric. This simply emphasizes that only the instantaneous variance covariance matrix is ‘theoretically observable,’ i.e., has economic meaning independent of the model.

We note that in total 14 risk-neutral parameters are identifiable: six from the drift and eight from the variance-covariance matrix. Below, we will use both of these representations to simplify the notation.

²⁵We omit similar identities for the other parameters for the sake of brevity.

3.2 USV in the observable $A_1(3)$ model

In this section we propose a full characterization of the ‘observable’ $A_1(3)$ 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:

$$m_r = -2c_V^2 \quad (65)$$

$$m_\mu = 3c_V \quad (66)$$

$$m_V = 1 \quad (67)$$

$$\sigma_V^\mu = c_V^2. \quad (68)$$

Interestingly, note that our representation leads naturally to the condition in equation (67). Indeed, equation (29) 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 free 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 gets reduced further to nine. Indeed, admissibility requires that the model satisfy both the USV conditions given in equations (65)-(68) and the admissibility conditions given in equations (53) and (54).

Combining the USV conditions (66) and (68) 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 (61) and (62), it follows that the parameters must satisfy the following system of equations:

$$\begin{cases} \sigma_1\nu_1 + \sigma_2^2\nu_2 + \sigma_3^2\nu_3 &= -\sqrt{\nu_1^2 + \nu_2^2\sigma_2^2 + \nu_3^2\sigma_3^2} \\ \sigma_1^2 + \sigma_2^2 + \sigma_3^2 &= 1. \end{cases} \quad (69)$$

If we can find parameters that satisfy these equations, then the three USV conditions (65)-(67) can be satisfied by appropriately choosing the parameter values for m_r , m_μ , m_V . Further, the admissibility conditions (53) and (54) 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 (69), note that if we define the two vectors $\{u, v\}$ in \mathbb{R}^3 by their coordinates $u = [\sigma_1, \sigma_2, \sigma_3]$ and $v = [\nu_1, \nu_2\sigma_2, \nu_3\sigma_3]$, then the system can be rewritten as:

$$\begin{cases} \|u\| &= 1 \\ u \cdot \frac{v}{\|v\|} &= -1. \end{cases} \quad (70)$$

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

$$\sigma_1 = -\frac{\nu_1}{\sqrt{\nu_1^2 + \nu_2^2 \sigma_2^2 + \nu_3^2 \sigma_3^2}} \quad (71)$$

$$\sigma_2 = -\frac{\nu_2 \sigma_2}{\sqrt{\nu_1^2 + \nu_2^2 \sigma_2^2 + \nu_3^2 \sigma_3^2}} \quad (72)$$

$$\sigma_3 = -\frac{\nu_3 \sigma_3}{\sqrt{\nu_1^2 + \nu_2^2 \sigma_2^2 + \nu_3^2 \sigma_3^2}}. \quad (73)$$

There are three possible solutions to this system:

- Case 1: $\sigma_2, \sigma_3 \neq 0$
- Case 2: $\sigma_3 = 0$
- 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 nested in Case 2. Hence, we focus our attention on Case 2.

In this case equation (73) holds for any value of ν_3 . Equations (71) and (72) further imply that

$$\nu_1 = \sigma_1 c_V \quad (74)$$

$$\nu_2 = c_V. \quad (75)$$

Thus the system of equations becomes:

$$dV_t = (\gamma_V - \kappa_V V_t) dt + \sigma_V \sqrt{V_t - \psi_1} dZ_1^Q(t) \quad (76)$$

$$dr_t = \mu_t^Q dt + \sigma_1 \sqrt{V_t - \psi_1} dZ_1^Q(t) + \sqrt{(1 - \sigma_1^2)V_t + \sigma_1^2 \psi_1 + \psi_3} dZ_2^Q(t) + \sqrt{-\psi_3} dZ_3^Q(t) \quad (77)$$

$$d\mu_t^Q = (m_0 - 2c_V^2 r_t + 3c_V \mu_t^Q + V_t) dt + c_V \left(\sigma_1 \sqrt{V_t - \psi_1} dZ_1^Q(t) + \sqrt{(1 - \sigma_1^2)V_t + \sigma_1^2 \psi_1 + \psi_3} dZ_2^Q(t) \right) + \nu_3 \sqrt{-\psi_3} dZ_3^Q(t), \quad (78)$$

with the following conditions:

$$\kappa_V > 0 \quad \text{for stationarity} \quad (79)$$

$$c_V < 0 \quad \text{for stationarity} \quad (80)$$

$$\gamma_V - \kappa_V \psi_1 > 0 \quad \text{for admissibility} \quad (81)$$

$$-\psi_3 > 0 \quad \text{for admissibility} \quad (82)$$

$$1 > \sigma_1^2 \quad \text{for admissibility} \quad (83)$$

$$\psi_1 + \psi_3 > 0 \quad \text{for admissibility.} \quad (84)$$

Thus the model has 9 parameters under the Q measure:

$$\gamma_V, \kappa_V, \sigma_V, \psi_1, \sigma_1, c_V, \psi_3, m_0, \nu_3.$$

Finally, we verify that the short rate process given by equations (76)-(78) above exhibits USV in that the zero-coupon bond price is not a function of the volatility state variable V_t :

Proposition 1 *If the short rate process follows the three-factor Markov process given by equations (76), (77) and (78), where the parameters satisfy the admissibility conditions (79)-(84), then zero-coupon bond prices are given by:*

$$P(t, T) = \exp \left(A(T-t) - B_r(T-t)r_t - B_\mu(T-t)\mu_t^\varrho \right), \quad (85)$$

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

$$B_r(\tau) = \frac{-3 + 4e^{c_V\tau} - e^{2c_V\tau}}{2c_V} \quad (86)$$

$$B_\mu(\tau) = \frac{(1 - e^{c_V\tau})^2}{2c_V^2} \quad (87)$$

$$A(\tau) = \int_0^\tau \left(\frac{1}{2}\sigma_0^\mu B_\mu^2(s) - m_0 B_\mu(s) + c_0 B_r(s) B_\mu(s) \right) ds \quad (88)$$

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

$$c_0 = \psi_3 (c_V - \nu_3) \quad (89)$$

$$\sigma_0^\mu = \psi_3 (c_V^2 - \nu_3^2). \quad (90)$$

Proof See Appendix A \square

From equations (86)-(88) it is clear that only four parameters, $\{m_0, \psi_3, \nu_3, c_V\}$, are identifiable from the cross-section of bond prices. Further, from observing a time series of bond prices we can determine both the volatility state variable and the three diffusion parameters $\{\sigma_V, \sigma_1, \psi_1\}$. However, using only panel data on bond prices, we cannot determine the risk-neutral drift parameters (γ_V, κ_V) of V .²⁶ Rather, prices of other fixed income derivatives (e.g., caps) must be used to infer these risk-neutral parameters.

Note that bond prices would retain their exponential-affine form in the above model for *any* specification of the process for V_t . Indeed, the proof of Proposition 1 does not depend in on the specific process followed by the variance of the short rate.²⁷ In other words, bond prices can be exponential-affine even if the state vector is not! This could prove helpful in estimating more general models for the volatility dynamics while retaining the analytical tractability of affine models for bond prices. 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 equation (88) has an analytic solution, but to simplify notation we leave it in integral form.

²⁶This statement assumes that the risk premia are general enough so that the risk-neutral parameters (γ_V, κ_V) are distinct from their physical-measure counterparts.

²⁷The only condition is that the volatility process be sufficiently regular for the stochastic integral in equation (A.20) to be a martingale.

3.3 Observable $A_1(4)$ USV model

In this section we identify an $A_1(4)$ model which exhibits USV. For identification purposes, we see from equation (29) that is natural to use $\{Y_0, Y_1, Y_2, V\}$, or equivalently $\{r_t, \mu_t^Q, \theta_t^Q, V_t\}$ as the state vector, where θ_t^Q is defined by $\theta_t^Q \equiv 3Y_{2,t} = \left(\frac{1}{dt}E_t^Q[d\mu_t^Q] - V_t\right)$. The maximal $A_1(4)$ model is given by:

$$dV_t = (\gamma_V - \kappa_V V_t) dt + \sigma_V \sqrt{V_t - \psi_1} dZ_1^Q(t) \quad (91)$$

$$dr_t = \mu_t^Q dt + \sigma_1 \sqrt{V_t - \psi_1} dZ_1^Q(t) + \sum_{i=2}^4 \sqrt{\sigma_i^2 V_t - \psi_i} dZ_i^Q(t) \quad (92)$$

$$d\mu_t^Q = (\theta_t^Q + V_t) dt + \nu_1 \sqrt{V_t - \psi_1} dZ_1^Q(t) + \sum_{i=2}^4 \nu_i \sqrt{\sigma_i^2 V_t - \psi_i} dZ_i^Q(t) \quad (93)$$

$$d\theta_t^Q = (a_0 + a_r r_t + a_\mu \mu_t^Q + a_\theta \theta_t^Q + a_V V_t) dt + \eta_1 \sqrt{V_t - \psi_1} dZ_1^Q(t) + \sum_{i=2}^4 \eta_i \sqrt{\sigma_i^2 V_t - \psi_i} dZ_i^Q(t), \quad (94)$$

where by definition of V_t we have:

$$\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2 = 1 \quad (95)$$

$$\sigma_1^2 \psi_1 + \psi_2 + \psi_3 + \psi_4 = 0. \quad (96)$$

The model is admissible if

$$\gamma_V \geq \kappa_V \psi_1 \quad (97)$$

$$\psi_1 \geq \max \frac{\psi_i}{\sigma_i^2} \quad \forall i = 2, 3, 4 \quad \text{s.t. } \sigma_i \neq 0. \quad (98)$$

Note that the maximal unrestricted $A_1(4)$ model has a total of 22 free risk-neutral parameters $(\gamma_V, \kappa_V, \sigma_V, \{\psi_i, \nu_i, \eta_i, \sigma_i\}_{i=1}^4, a_0, a_r, a_\theta, a_\mu, a_V)$, and two restrictions from equations (95)-(96).

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

$$\sigma \equiv (\sigma_1, \sigma_2, \sigma_3, \sigma_4) \quad (99)$$

$$\nu \equiv (\nu_1, \nu_2 \sigma_2, \nu_3 \sigma_3, \nu_4 \sigma_4) \quad (100)$$

$$\eta \equiv (\eta_1, \eta_2 \sigma_2, \eta_3 \sigma_3, \eta_4 \sigma_4). \quad (101)$$

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

$$\Sigma^2 = \begin{bmatrix} V & c_{r\mu} V + c_{r\mu}^0 & c_{r\theta} V + c_{r\theta}^0 \\ c_{r\mu} V + c_{r\mu}^0 & \sigma_\mu^2 V + \sigma_\mu^0 & c_{\mu\theta} V + c_{\mu\theta}^0 \\ c_{r\theta} V + c_{r\theta}^0 & c_{\mu\theta} V + c_{\mu\theta}^0 & \sigma_\theta^2 V + \sigma_\theta^0 \end{bmatrix}, \quad (102)$$

where, by definition

$$c_{r\mu} \equiv \sigma \cdot \nu \quad (103)$$

$$c_{r\theta} \equiv \sigma \cdot \eta \quad (104)$$

$$c_{\mu\theta} \equiv \nu \cdot \eta \quad (105)$$

$$\sigma_\theta^2 \equiv \|\eta\|^2 \quad (106)$$

$$\sigma_\mu^2 \equiv \|\nu\|^2. \quad (107)$$

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

$$a_r = -2c_{r\mu}^2(3c_{r\mu} - a_\theta) \quad (108)$$

$$a_\mu = 7c_{r\mu}^2 - 3c_{r\mu}a_\theta \quad (109)$$

$$a_V = 3c_{r\mu} \quad (110)$$

$$\sigma_\mu = -c_{r\mu} \quad (111)$$

$$\sigma_\theta = c_{r\theta} \quad (112)$$

$$c_{r\theta} = c_{r\mu}^2 \quad (113)$$

$$c_{\mu\theta} = c_{r\mu}^3. \quad (114)$$

Analogous to the $A_1(3)$ USV model, there is a natural geometric interpretation for the restrictions on the variance covariance matrix. For example, equations (95), (103), (107), and (111) imply that the vectors σ and ν are collinear but pointing in opposite directions. Similarly, equations (95), (104), (106) and (112) imply that the vectors σ and η are collinear and pointing in the same direction. Combining these results with the implications from equations (103), (104) and (113) we conclude that

$$\sigma = \frac{\eta}{\|\eta\|} = -\frac{\nu}{\|\nu\|} \quad (115)$$

$$\|\eta\| = \|\nu\|^2. \quad (116)$$

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 (115) implies that $\eta_2 = \eta_3 = \eta_4$ and $\nu_2 = \nu_3 = \nu_4$. Therefore, we can define the Brownian motion B_t^Q by $\sqrt{\sigma^2 V - \bar{\psi}} dB^Q(t) \equiv \sum_{i=2}^4 \sqrt{\sigma_i^2 V - \psi_i} dZ_i^Q(t)$, where $\bar{\sigma}^2 = \sum_{i=2}^4 \sigma_i^2$ and $\bar{\psi} = \sum_{i=2}^4 \psi_i$. It follows that the dynamics of the state vector is then adapted to the natural filtration generated by the two Brownian motions (Z_1^Q, B^Q) . That is, this case reduces to a two-factor model as claimed.

- Case 2: $\sigma_i = 0$ for some $i \in [2, 3, 4]$ and $\sigma_j \neq 0$ for all $j \in [2, 3, 4]$ such that $j \neq i$. Analogous to the previous case, we can show that this case reduces to a three factor model.
- Case 3: $\sigma_j = \sigma_i = 0$ for some $i \neq j \in [2, 3, 4]$. Without loss of generality, assume $\sigma_3 = \sigma_4 = 0$. Then equations (115) and (116) imply:

$$\nu_1 = c_{r\mu} \sigma_1 \quad (117)$$

$$\nu_2 = c_{r\mu} \quad (118)$$

$$\eta_1 = c_{r\mu}^2 \sigma_1 \quad (119)$$

$$\eta_2 = c_{r\mu}^2. \quad (120)$$

Further, from equations (95) and (96) we have:

$$\sigma_1^2 + \sigma_2^2 = 1 \quad (121)$$

$$\sigma_1^2 \psi_1 + \psi_2 + \psi_3 + \psi_4 = 0. \quad (122)$$

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

$$dV_t = (\gamma_V - \kappa_V V_t) dt + \sigma_V \sqrt{V_t - \psi_1} dZ_1^Q(t) \quad (123)$$

$$\begin{aligned} dr_t = & \mu_t^Q dt + \sigma_1 \sqrt{V_t - \psi_1} dZ_1^Q(t) + \sqrt{(1 - \sigma_1^2)V_t + \sigma_1^2 \psi_1 + \psi_3 + \psi_4} dZ_2^Q(t) \\ & + \sqrt{-\psi_3} dZ_3^Q(t) + \sqrt{-\psi_4} dZ_4^Q(t) \end{aligned} \quad (124)$$

$$\begin{aligned} d\mu_t^Q = & (\theta_t^Q + V_t) dt + c_{r\mu} \sigma_1 \sqrt{V_t - \psi_1} dZ_1^Q(t) + c_{r\mu} \sqrt{(1 - \sigma_1^2)V_t + \sigma_1^2 \psi_1 + \psi_3 + \psi_4} dZ_2^Q(t) \\ & + \nu_3 \sqrt{-\psi_3} dZ_3^Q(t) + \nu_4 \sqrt{-\psi_4} dZ_4^Q(t) \end{aligned} \quad (125)$$

$$\begin{aligned} d\theta_t^Q = & \left(a_0 - 2c_{r\mu}^2 (3c_{r\mu} - a_\theta) r_t + (7c_{r\mu}^2 - 3c_{r\mu} a_\theta) \mu_t^Q + a_\theta \theta^Q + 3c_{r\mu} V_t \right) dt \\ & + c_{r\mu}^2 \sigma_1 \sqrt{V_t - \psi_1} dZ_1^Q(t) + c_{r\mu}^2 \sqrt{(1 - \sigma_1^2)V_t + \sigma_1^2 \psi_1 + \psi_3 + \psi_4} dZ_2^Q(t) \\ & + \eta_3 \sqrt{-\psi_3} dZ_3^Q(t) + \eta_4 \sqrt{-\psi_4} dZ_4^Q(t). \end{aligned} \quad (126)$$

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.²⁸

The admissibility restrictions are:

$$\kappa_V > 0 \quad \text{for stationarity} \quad (127)$$

$$c_{r\mu} < 0 \quad \text{for stationarity: see equation (111)} \quad (128)$$

²⁸Note that the two restrictions $\sigma_3 = \sigma_4 = 0$ makes one of the seven restrictions from equations (108)-(114) redundant, leading to eight total restrictions, and thus $22 - 8 = 14$ parameters.

$$a_\theta - 3c_{r\mu} < 0 \quad \text{for stationarity: see, for example, equation (135)} \quad (129)$$

$$\gamma_V - \kappa_V \psi_1 > 0 \quad \text{for admissibility} \quad (130)$$

$$\psi_3, \psi_4 < 0 \quad \text{for admissibility} \quad (131)$$

$$1 > \sigma_1^2 \quad \text{for admissibility: see equation (121)} \quad (132)$$

$$\psi_1 + \psi_3 + \psi_4 > 0 \quad \text{for admissibility.} \quad (133)$$

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 2 *If the short rate process follows a four-factor Markov process given by equations (123)-(126) where the parameters satisfy the admissibility conditions (127)-(133) then zero-coupon bond prices are given by:*

$$P(t, T) = \exp \left(A(T-t) - B_r(T-t) r_t - B_\mu(T-t) \mu_t^Q - B_\theta(T-t) \theta_t^Q \right), \quad (134)$$

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

$$B_r(\tau) = \frac{e^{c_{r\mu}\tau} (6c_{r\mu} - 2a_\theta)}{4c_{r\mu}^2 - c_{r\mu}a_\theta} + \frac{e^{2c_{r\mu}\tau} (3c_{r\mu} - a_\theta)}{-10c_{r\mu}^2 + 2c_{r\mu}a_\theta} + \frac{7c_{r\mu} - 3a_\theta}{-6c_{r\mu}^2 + 2c_{r\mu}a_\theta} - \frac{2c_{r\mu}^2 e^{(-3c_{r\mu}+a_\theta)\tau}}{\Gamma} \quad (135)$$

$$B_\mu(\tau) = \frac{a_\theta}{2c_{r\mu}^2 (-3c_{r\mu} + a_\theta)} + \frac{e^{2c_{r\mu}\tau} (2c_{r\mu} - a_\theta)}{10c_{r\mu}^3 - 2c_{r\mu}^2 a_\theta} + \frac{e^{c_{r\mu}\tau} (c_{r\mu} - a_\theta)}{-4c_{r\mu}^3 + c_{r\mu}^2 a_\theta} + \frac{3c_{r\mu} e^{(-3c_{r\mu}+a_\theta)\tau}}{\Gamma} \quad (136)$$

$$B_\theta(\tau) = \frac{e^{c_{r\mu}\tau}}{c_{r\mu}^2 (-4c_{r\mu} + a_\theta)} + \frac{1}{6c_{r\mu}^3 - 2c_{r\mu}^2 a_\theta} + \frac{e^{2c_{r\mu}\tau}}{10c_{r\mu}^3 - 2c_{r\mu}^2 a_\theta} - \frac{e^{(-3c_{r\mu}+a_\theta)\tau}}{\Gamma} \quad (137)$$

$$A(\tau) = \int_0^\tau \left(\frac{\sigma_\mu^0}{2} B_\mu(s)^2 + \frac{\sigma_\theta^0}{2} B_\theta(s)^2 + B_r(s) B_\mu(s) c_{r\mu}^0 + B_r(s) B_\theta(s) c_{r\theta}^0 + B_\theta(s) B_\mu(s) c_{\mu\theta}^0 - B_\theta(s) a_0 \right) ds, \quad (138)$$

and where

$$\Gamma = (3c_{r\mu} - a_\theta) (4c_{r\mu} - a_\theta) (5c_{r\mu} - a_\theta). \quad (139)$$

Proof See Appendix A \square

As before, we note that the integral in (138) has an analytic expression, but to simplify notation we leave it in integral form. We now turn to the estimation of three and four-factor 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:

$$dZ_i^P(t) = dZ_i^Q(t) - \lambda_i(t) dt, \quad \forall i = 1, 2, 3, \quad (140)$$

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

$$\lambda_1(t) = \frac{\lambda_{10} + \lambda_{13}V_t}{\sqrt{V_t - \psi_1}} \quad (141)$$

$$\lambda_2(t) = \frac{\lambda_{20} + \lambda_{21}r_t + \lambda_{22}\mu_t^Q + \lambda_{23}V_t}{\sqrt{\sigma_2^2V_t - \psi_2}} \quad (142)$$

$$\lambda_3(t) = \frac{\lambda_{30} + \lambda_{31}r_t + \lambda_{32}\mu_t^Q + \lambda_{33}V_t}{\sqrt{\sigma_3^2V_t - \psi_3}}. \quad (143)$$

By including a term in eq. (141) 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.²⁹ We therefore impose the Feller condition for both measures

²⁹Cheridito, Filipovic, and Kimmel (2004) recently offer an alternative proof of this result in the context of affine models. Liptser and Shiryaev's result applies to any process of the 'diffusion type' (see their definition 7 p. 118).

as a constraint in our analysis.³⁰ Combined with condition (54), the Feller condition implies that the Radon-Nikodym density defined by the risk-premia specification in eq. (140) 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:

$$dV_t = \left(\gamma_V + \lambda_{V0} - (\kappa_V - \lambda_{VV}) V_t \right) dt + \sigma_V \sqrt{V_t - \psi_1} dZ_1(t) \quad (144)$$

$$dr_t = \left(\lambda_{r0} + \lambda_{rr} r_t + (1 + \lambda_{r\mu}) \mu_t^Q + \lambda_{rV} V_t \right) dt + \sigma_1 \sqrt{V_t - \psi_1} dZ_1(t) + \sqrt{\sigma_2^2 V_t - \psi_2} dZ_2(t) + \sqrt{\sigma_3^2 V_t - \psi_3} dZ_3(t) \quad (145)$$

$$d\mu_t^Q = \left(m_0 + \lambda_{\mu 0} + (m_r + \lambda_{\mu r}) r_t + (m_\mu + \lambda_{\mu\mu}) \mu_t^Q + (m_V + \lambda_{\mu V}) V_t \right) dt + \nu_1 \sqrt{V_t - \psi_1} dZ_1(t) + \nu_2 \sqrt{\sigma_2^2 V_t - \psi_2} dZ_2(t) + \nu_3 \sqrt{\sigma_3^2 V_t - \psi_3} dZ_3(t). \quad (146)$$

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 the $A_1(4)$ USV model, essentially affine risk premia are defined similarly:

$$\lambda_1(t) = \frac{\lambda_{10} + \lambda_{13} V_t}{\sqrt{V_t - \psi_1}} \quad (147)$$

$$\lambda_2(t) = \frac{\lambda_{20} + \lambda_{21} r_t + \lambda_{22} \mu_t^Q + \lambda_{23} V_t + \lambda_{24} \theta_t}{\sqrt{(1 - \sigma_1^2) V_t + \sigma_1^2 \psi_1 + \psi_3 + \psi_4}} \quad (148)$$

$$\lambda_3(t) = \frac{\lambda_{30} + \lambda_{31} r_t + \lambda_{32} \mu_t^Q + \lambda_{33} V_t + \lambda_{34} \theta_t}{\sqrt{-\psi_3}} \quad (149)$$

$$\lambda_4(t) = \frac{\lambda_{40} + \lambda_{41} r_t + \lambda_{42} \mu_t^Q + \lambda_{43} V_t + \lambda_{44} \theta_t}{\sqrt{-\psi_4}}. \quad (150)$$

Note, however, that λ_{10} and λ_{13} are not identifiable because of USV, so that only 15 of the 17 risk premia parameters can be identified (along with 14 risk-neutral parameters).

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

$$dV_t = (\gamma_V - \kappa_V V_t) dt + \sigma_V \sqrt{V_t - \psi_1} dZ_1^Q(t) \quad (151)$$

$$dr_t = (\gamma_r - \kappa_{rr} r_t - \kappa_{rV} V_t) dt + \sigma_1 \sqrt{V_t - \psi_1} dZ_1^Q(t) + \sqrt{(1 - \sigma_1^2) V_t + \sigma_1^2 \psi_1} dZ_2^Q(t). \quad (152)$$

³⁰The 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.

Generalized essentially affine risk premia for this model are

$$\lambda_1(t) = \frac{\lambda_{10} + \lambda_{12}V_t}{\sqrt{V_t - \psi_1}} \quad (153)$$

$$\lambda_2(t) = \frac{\lambda_{20} + \lambda_{21}r_t + \lambda_{22}V_t}{\sqrt{(1 - \sigma_1^2)V_t + \sigma_1^2\psi_1}}, \quad (154)$$

yielding a total of 13 parameters (8 risk neutral plus 5 risk premia).

4.2 Data

We use weekly LIBOR and swap rate data from Datastream from January 6, 1988, to December 29, 2004. 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.³¹ 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.³² To overcome this problem, following Jones (2003a), we estimate the ‘synchronized’ values of the LIBOR rates. The procedure is essentially a smoothing algorithm that exploits the extremely high correlations between changes in LIBOR and swap rates of similar maturities. Jones (2003a) shows that the errors of the procedure are typically on the order of one basis point, or roughly one third the magnitude of the errors one would make by using either the same morning’s LIBOR quote, the next morning’s quote, or the average of the two.

From the bootstrapped yield curves we extract yields with maturities of 0.5, 1, 2, 3, 4, 5, 7, and 10 years. We choose these eight maturities because on each day in the sample there is some 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 unobserved (but theoretically observable) variables. We proceed with these methodological caveats in place.

³¹If 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)).

³²LIBOR 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.

4.3 Posterior sampler

We estimate all models using a Bayesian approach that combines elements of Jones (2003b), Lamoureux and Witte (2002), Bester (2004), Sanford and Martin (2003), and Polson, Stroud, and Müller (2001). In each of these papers, data augmentation and a Gibbs-like posterior sampler are used to simplify the computation of posterior distributions of the model parameters. As in all Bayesian analysis, we seek to describe the posterior density of the model parameters conditional on the yield data, which is decomposed using Bayes rule as the product of the likelihood function and the prior. As is common in continuous time finance models, evaluating the likelihood function is difficult because of the presence of imperfectly observed state variables and the fact that transition densities are not usually known in closed form.

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

As in Pennacchi (1991), Brandt and He (2002), and Bester (2004), we assume that all yields are observed with error, thereby avoiding the arbitrariness of assuming that some yields (or linear combinations of yields) are measured with error, while some are not. Somewhat differently from these papers, however, we treat the *principal components* of yields, rather than yields themselves, as our raw data. We do so because it is unlikely that the yield errors are cross-sectionally uncorrelated, yet specifying a completely flexible covariance matrix of yield errors creates a rather large number of additional parameters to estimate (e.g. 36 parameters for 8 yields). Using principal components, which are by construction unconditionally orthogonal, should lead to model errors that are approximately uncorrelated, allowing us to use a more parsimonious diagonal error covariance matrix.³³ The principal component loadings and percentages of variance explained appear in Table 2. As in Litterman and Scheinkman (1991), the first three principal components explain most variation in yields, and can roughly be interpreted as level, slope, and curvature.

While the posterior simulator is described in much more detail in Appendix B, we briefly outline our approach here and compare it to others in the literature. Letting $\mathcal{P} = \{\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_T\}$ denote the time series of principal component vectors and ϕ the vector of model parameters, we seek to

³³An appealing alternative approach Brandt and He (2002) is to parameterize the covariances as parsimonious functions of bond maturities and a few free parameters.

compute

$$p(\phi|\mathcal{P}) \propto p(\mathcal{P}|\phi)p(\phi),$$

where the first term on the right is the likelihood and the second is the prior. The intractability of the likelihood makes a direct calculation of the posterior impossible, so we proceed using several techniques from the Markov chain-Monte Carlo literature.

Following earlier approaches, we augment the observable data \mathcal{P} with the term structure factor data $X = \{X_1, X_2, \dots, X_T\}$. While these factors are theoretically observable in the sense that they have interpretations independent of the model being considered, they are not directly observed in practice given the availability of only finite maturity yields. We therefore integrate out the uncertainty in these state variables using a Gibbs-like posterior simulator that alternates between performing draws from $p(\phi|\mathcal{P}, X)$ and $p(X|\mathcal{P}, \phi)$. Under very weak conditions, the resulting sequence of draws of ϕ converges in distribution to our target, the posterior distribution $p(\phi|\mathcal{P})$.

Unlike Lamoureux and Witte (2002) and Bester (2005), we approximate the true dynamics,

$$dX_t = \mathcal{K}(\Theta - X_t) dt + \Sigma \sqrt{S_t} dZ_t, \quad (155)$$

using the Euler scheme

$$X_{t+h} - X_t = h\mathcal{K}(\Theta - X_t) + \sqrt{h}\Sigma \sqrt{S_t} \epsilon_{t+h}, \quad (156)$$

where $\epsilon_{t+h} \sim N(0, I)$. Thus we do not use the true noncentral chi-squared distribution for the square root factors,³⁴ but instead choose h to be small enough to approximate that distribution.

The likelihood function is completed by specifying the relation between the data and the state vector. Given the linearity of bond yields in state variables (16) and the linear relation between principal components and yields,

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

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

$$\mathcal{P}_t = K + LX_t + e_t, \quad (157)$$

our “measurement equation”.

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

³⁴Further, note that the transition density is only known in closed form for one (i.e., volatility) of the three state variables.

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

$$X_t = \begin{bmatrix} V_t \\ X_t^o \end{bmatrix}, \quad (158)$$

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

By far the most difficult component of the algorithm is the draw of $\phi^Q \equiv \{\mathcal{K}^Q, \Theta^Q, \Sigma, \alpha, \text{ and } \beta\}$, the parameters that drive state variable dynamics under Q . If we were to follow the standard recipe for a Gibbs sampler, we would attempt to draw ϕ^Q conditional on the data and all the other unknowns in the model, i.e. from

$$p(\phi^Q | \mathcal{P}, X^o, V, \phi^\Omega, \phi^\lambda). \quad (159)$$

The problem is that ϕ^Q influences bond prices in a highly nonlinear manner, in part through a differential equation whose solution is often not known analytically. Direct sampling from this density is therefore impossible.

Even more numerical methods, such as the Metropolis-Hastings algorithm, are difficult to apply. This is because the success of Metropolis-Hastings requires finding a distribution of ϕ^Q (the ‘‘candidate generator’’) that is at least roughly consistent with both \mathcal{P} and X^o (in addition to the other conditioning arguments). The problem is that the link between yield data, represented by \mathcal{P} , and latent factors such as X^o is extremely tight, highly nonlinear, and completely parameter-dependent. For these reasons it is very hard to find a density that comes close to approximating (159).

We therefore find relatively large gains in efficiency from performing the draws of ϕ^Q and X^o *jointly* from

$$p(\phi^Q, X^o | \mathcal{P}, V, \phi^\Omega, \phi^\lambda) = p(\phi^Q | \mathcal{P}, V, \phi^\Omega, \phi^\lambda) p(X^o | \phi^Q, \mathcal{P}, V, \phi^\Omega, \phi^\lambda). \quad (160)$$

To satisfy this density, a candidate ϕ^Q must be consistent with \mathcal{P} only, which is a much easier criterion. The result is a Metropolis-Hastings algorithm that traverses the parameter space much more quickly than one based on (159). Furthermore, evaluating this density, at least up to a constant of proportionality, is feasible because of the linear Gaussian state space representation that holds after conditioning on V . This enables the use of the textbook Kalman filter³⁵ to compute $p(\phi^Q | \mathcal{P}, V, \phi^\Omega, \phi^\lambda)$. The second component, $p(X^o | \phi^Q, \mathcal{P}, V, \phi^\Omega, \phi^\lambda)$, can also be computed in closed form. More importantly, as noted above, the draw of X^o , consisting of all state variables

³⁵The time variation in means and covariances requires a version of the Kalman filter that is more general than some simple versions. See de Jong and Shephard (1995) for one implementation.

save V over all time periods, can be made in a single “block” using the simulation smoother of de Jong and Shephard (1995). These features distinguish our algorithm from prior work.

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

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

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

Finally, as in all Bayesian analysis we must specify prior distributions. In all of our analysis, priors are completely flat, or exactly proportional to a constant, with three exceptions. First, the prior density for each measurement error standard deviation $\sqrt{\Omega_{i,i}}$ is proportional to $1/\sqrt{\Omega_{i,i}}$, as is standard. Second, all regions of the parameter space in which the model is nonstationary, inadmissible, or in which the Feller condition does not hold are assumed to have zero prior probability. This means that any parameter draw that violates any of these conditions is immediately rejected. Finally, all parameter draws that generate covariance degeneracies are assumed to have zero probability. This means, for instance, that the Σ matrix must have full rank and that each $\Omega_{i,i}$ is positive.

4.4 Point estimates and confidence intervals

For each parameter, we report point estimates along with 95% confidence intervals computed from the 2.5 and 97.5 percentiles of the MCMC draws from the posterior distribution. Let ϕ_i denote the i^{th} MCMC draw of the parameter vector from its posterior distribution and $\tilde{\phi}_i$ be the same vector normalized by the posterior standard deviations. We then find the draw i that minimizes

$$\sum_j \left| \tilde{\phi}_j - \tilde{\phi}_i \right|.$$

The resulting ϕ_i is the L_1 center of the normalized posterior distribution, a version of the multivariate posterior median, and is taken as our vector of point estimates.

Raftery (1996) reports that the unnormalized version often provides an accurate approximation of the posterior mode. Because of the vastly different scales of some of the model parameters, we felt that normalization would be more natural. Using multivariate rather than univariate medians or means is appropriate because univariate medians or means are not necessarily located in regions of high posterior probability.³⁶

From MCMC output, it is possible to compute posteriors of functions of parameters in addition to the parameters themselves. We therefore report posterior statistics on restricted parameters, such as the parameter σ_2 in the $A_1(3)$ USV model, which is restricted to equal $\sqrt{1 - \sigma_1^2}$. As before, confidence intervals are calculated from the 2.5 and 97.5 percentiles of the posterior draws, while the restricted point estimates are computed directly from the point estimates of the unrestricted parameters.³⁷

In some cases, point estimates are quite far from the midpoint of the 95% confidence interval. Figure 1 gives a preview of a few univariate posterior distributions for the $A_1(3)$ USV model. The first panel illustrates a case in which the posterior is approximately Gaussian, the multivariate posterior median (the solid vertical line) is at the univariate mode, and the 95% confidence interval (the two dashed vertical lines) are symmetric about the mode. The second panel is much different, as the posterior of ψ_1 is seen to be highly non-Gaussian with a posterior mode at zero, the lowest admissible value. In this case, the multivariate posterior median of $.134 \times 10^{-5}$ (the solid line) is somewhat higher than the univariate median of $.106 \times 10^{-5}$, and the 95% confidence interval is highly skewed. Finally, the third panel shows that multivariate posterior medians may differ from univariate modes even for parameters with relatively Gaussian posteriors, though these differences tend to be minor.

³⁶As an extreme example, consider a hypothetical posterior distribution for x and y in which $x \sim U[-5, 5]$ and $y = x^2 + u$, where $u \sim U[-1, 1]$. In this case the posterior means of x and y are zero and 8.3, respectively, yet the pair $\{x = 0, y = 8.3\}$ is not even in the support of the joint distribution since y has a maximum value of 1 when $x = 0$.

³⁷In one atypical case below there is a restricted parameter whose point estimate is outside its 95% confidence interval. To see how this can happen, consider the case in which $x \sim U[-1, 1]$ is the only free parameter and $y = \sqrt{1 - x^2}$ is a restricted parameter. The mean and median of x are both clearly 0, which would imply a point estimate of 0 for x and $\sqrt{1 - 0^2} = 1$ for y . Yet, $y = 1$ is clearly above the 97.5 percentile of the distribution of y .

5 Empirical results

5.1 Posterior summaries

Tables 3A and 3B present posterior distributions of the parameters of four models: unrestricted $A_1(2)$, $A_1(3)$ with USV restrictions, unrestricted $A_1(3)$, and $A_1(4)$ with USV restrictions. All models were estimated using weekly data from January 1988 to December 2002, thus leaving data from 2003-2004 for out-of-sample analysis. Finally, posteriors were computed by setting the discretization parameter h equal to 1, as posteriors under $h = .2$ (summarized in Appendix B) resulted in no appreciable differences but considerably greater computational effort.

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

The complexity of term structure models sometimes makes interpretation of the parameter estimates difficult. Our rotation of the state vector reduces this difficulty, because by pinning down the interpretation of the state vector it is much easier to compare parameter values across different specifications. For instance, in all four specifications the parameter ψ_1 alone determines the lower bound of the short rate variance, and the covariance between shocks to the short rate and its variance is always equal to $\sigma_1 \sigma_V (V_t - \psi_1)$. While it is entirely feasible to quantify these moments under the Dai and Singleton (2000) rotation, we emphasize that they are not nearly as obvious. The reason is that in the DS rotation the short rate variance is not a state variable, and so to compute any moment involving the short rate variance we must first compute the relation between that variance and the the state vector, a mapping that is different for each specification. This is straightforward but represents an additional step that is not necessary in our analysis.

It is therefore easy to see from the κ_V row of Table 3A that a major difference between USV and non-USV specifications is that the speed of mean reversion in short rate volatility, determined by κ_V^P , is much faster under the two USV specifications. Presumably, this is to offset the fact that short rate volatility is substantially more volatile under USV, as evidenced by the higher values of σ_V . A typical short rate volatility can be calculated by taking the square root of the long-run mean of short rate variance, or $\sqrt{\gamma_V^P / \kappa_V^P}$, which results in a fairly reasonable range of .0077 (for $A_1(4)$ USV) to .0133 (for unrestricted $A_1(3)$) per year, where .01 would indicate a yield volatility of one percentage point per year.

In comparing the unrestricted $A_1(3)$ model with the restricted USV version, we see numerous other differences. Covariance parameters (the η_i and ψ_i) bear little resemblance across the two specifications, and the parameters that govern the drift of μ^Q (m_0 , m_V , m_r , and m_μ) are also different. In particular, the USV restriction that $m_V = 1$ is strongly violated by the unrestricted model posterior, a finding that suggests misspecification of the $A_1(3)$ USV model. However, our results below suggest that both versions of the $A_1(3)$ model are deeply misspecified, making this

finding somewhat difficult to interpret.

These two specifications also imply vastly different risk premia, as seen in Table 3B. In fact, one of the only similarities in the table is that the risk premia parameters that are common to the two USV models show substantial agreement in sign and approximate magnitude. Furthermore, most of the additional risk premia parameters introduced in the four factor specification are indistinguishable from zero, indicating that the generalized essentially affine risk premia are likely overparameterized for this model. Again, we emphasize that comparing parameters across specifications is only feasible when the interpretation of the state vector is fixed.

5.2 Specification analysis

As in Eraker (2001), specification analysis can be conducted by looking at the residuals of the Euler approximation. Given each posterior draw of the parameter vector and the term structure factors, we may use (156) to “invert” a time series of residuals for each of the state variables. These residuals, constructed by subtracting the conditional mean and dividing by the conditional standard deviation, will be i.i.d. standard normals under a correct specification. Since these residuals will be different for each posterior draw, we form posterior distributions of functions of these residuals by evaluating the functions at each step of the sampler. For the purpose of model diagnostics, these functions will consist of the sample mean, standard deviation, skewness, kurtosis, as well as autocorrelations in levels, absolute values, and squares. Table 4 contains posterior medians of each of these statistics. If more than 95% of the posterior mass is above or below zero, the statistic is marked with an asterisk.

While it is immediately clear from the table that all four models display failings, several general patterns emerge. First, only the $A_1(4)$ USV model appears to capture the mean, standard deviation, and skewness of each of the state variables. Second, positive excess residual kurtosis is apparent in almost every case, though both USV specifications do markedly better since the degree of kurtosis in the short rate residuals is substantially lower. Finally, the non-USV specifications display large autocorrelations in the absolute values and squares of the residuals of the short rate factor. While the $A_1(4)$ USV model gets short rate dynamics mostly correct, it suffers from nonzero autocorrelations for the μ^Q and θ^Q factor errors.

For the two non-USV models, the combination of positive autocorrelation in squared and absolute residuals and the high degree of kurtosis in the short-rate residuals suggest a simple explanation, which will be supported by further analysis. Specifically, the conditional short rate volatility of both these models must be severely misspecified. This explanation also accounts for the negative skewness in the short rate factor residuals for the same models. Over our sample period, most of the unconditional skewness in short-term yield changes comes from the Federal Reserve’s sudden provision of liquidity in response to the terrorist attacks of September 2001. This happened to be a period of high conditional volatility, however, so the normalized residual for that observation would not be particularly negative for a model with a correctly specified volatility. In contrast, that single observation would have far more impact if that period was incorrectly characterized as having a low

or even average level of volatility.

5.3 Yield curve fit

Rather than relying only on statistical evaluations, 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 models' in-sample fits of the yield curve, both in terms of bias and root mean squared error. Below, we also investigate out-of-sample performance.

In this section and those that follow, we evaluate model performance using the point estimates from Table 3. While it would be preferable to integrate over the entire posterior distribution, this turns out to be very computationally demanding for most of the analyses we perform. For computing fitted yields, we therefore re-run our posterior sampler where only the state variables are sampled and the parameter values are held fixed. In a few places, such as Table 5, where we can compute results both by integrating over the posterior and by conditioning on the multivariate medians in Table 3, we find extremely minor differences.

Statistical tests for biases in fitted yields are relatively standard. Errors are defined as actual yields minus fitted yields, where fitted yields are computed via eq. (16). 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 6, 8, 9, 10, 11, and 12. 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 affect the ability of the model to fit the yield curve cross section and time-series?
- Unrestricted $A_1(3)$ versus $A_1(4)$ USV: Does the addition of a 'free' volatility state variable allow the remaining three factors to better explain yields?

For each maturity, these three pairwise comparisons are made using the method of Diebold and Mariano (1995). For each model, we compute forecast errors, say $\hat{e}_{1,t}$ and $\hat{e}_{2,t}$, and calculate t-

statistics for the difference in squared forecast errors

$$\hat{e}_{1,t}^2 - \hat{e}_{2,t}^2. \quad (161)$$

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 21 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 models other than $A_1(3)$ USV imply reasonably unbiased fits of individual yields with no rejections of zero mean errors, while the $A_1(3)$ USV model displays some clear failings. Root mean squared errors are clear in their preference for models with three factors in the yield curve ($A_1(3)$ and $A_1(4)$ USV) rather than two ($A_1(2)$ and $A_1(3)$ USV). Between the unrestricted $A_1(3)$ and the $A_1(4)$ USV model, the former offers modest but significant reductions in RMSE. Finally, the errors from all four models are highly autocorrelated, an indication that all the models are misspecified.

Given the superior performance of the unrestricted $A_1(3)$ model in sample relative to the $A_1(4)$ USV model, it is surprising that the reverse holds in Table 6, which reports out-of-sample yield fits using data from January 2003 to December 2004. While both models display significant biases, they are far smaller for the $A_1(4)$ USV specification. In addition, these deviations are much larger in terms of RMSE for $A_1(3)$ than they are for $A_1(4)$ USV. In some cases, the unrestricted $A_1(3)$ model underperforms its USV version as well, while the $A_1(3)$ USV model is clearly superior to $A_1(2)$. The dramatic out-of-sample breakdown of the unrestricted $A_1(3)$ model is consistent with some type of serious misspecification, the form of which we identify below.

In contrast, the reason for the high RMSEs of the $A_1(2)$ and $A_1(3)$ USV models, both in and out of sample, is obvious. Figure 2 plots time series of actual and model-implied curvature of yields, where curvature here is defined as $Y_{10y} - 2Y_{3y} + Y_{6m}$. A vertical dotted line denotes the end of the estimation period. It is perhaps not too surprising that actual curvature is virtually indistinguishable from that implied by the $A_1(4)$ USV model, since that model's θ^Q variable is essentially a curvature factor. While the unrestricted $A_1(3)$ model has also three factors that affect yields, none of those factors explicitly determines curvature, so the performance of that model was somewhat less predictable. In contrast, 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. Given Litterman and Scheinkman's (1991) finding that curvature essentially represents the third principal component in yields, it should be expected that the inability of our two simplest models to match actual curvature has substantial consequences.

5.4 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 again estimated by running the posterior sampler with parameters held fixed at the point estimates from Table 3. The resulting draws of the state variables are then averaged to get smoothed estimates.

Figure 3 shows the resulting time series of $E[V_t|\mathcal{P}, \hat{\phi}]$ for each specification along with 30-day trailing-window volatilities estimated from daily 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 trailing-window volatilities track each other closely, but for the non-USV specifications there appears to be little or no relation and hardly any movement at all.

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

Model-implied volatilities for both USV specifications are highly related to volatilities estimated both using trailing windows and Bollerslev's (1986) GARCH(1,1) model fitted to demeaned daily changes in the 6-month yield. Volatility from the unrestricted $A_1(3)$ and $A_1(2)$ models are actually negatively correlated with both the trailing-window and GARCH volatilities over the full sample period.

The bottom panel Table 7 also reports correlations of model volatilities with one-year at-the-money cap/floor option-implied volatilities, 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) forward LIBOR rates. As such, we 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 positively correlated with implied volatility, and also that implied volatility is about as closely related to the trailing window and GARCH series. On the other hand, Table 7 shows that volatilities from the non-USV specifications are negatively correlated to 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 2. In general, this relationship is weak for the USV specifications and for the GARCH and implied volatilities as well. For the $A_1(2)$ and $A_1(3)$ models, 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.

Finally, Table 7 reports correlations between model-implied estimates of the state variables underlying yields and the values obtained from the ‘interpolation’ scheme of section 2.3. In general, interpolated state variables are highly related to the values obtained through estimation of the model. For the $A_1(4)$ USV model, in particular, these correlations are all above .97 in each of the three sample periods. We take this as strong empirical evidence of the practical, not just theoretical, observability of the state vector under our model rotations.

These results highlight the dual role that volatility plays in an unrestricted affine model, as it affects both the cross section of bond prices as well as the time series properties of the short rate. The estimation of such a model therefore presents a tradeoff between choosing volatility dynamics that are more consistent with either role, and in the present data set it seems that the tradeoff is heavily tilted towards explaining the cross section. The result is that volatilities imputed from the two models without USV restrictions are essentially nonsensical, being unrelated to most 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.

5.5 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 3, so the bulk of our forecasts are in-sample. After using two years of data to initialize the forecasts, we are left with a 677-week in-sample period. With our hold-out sample from 2003 and 2004, we perform a 104-week out-of-sample validation of those results.

To construct a forecast, we first estimate the value of the current state variables. These are

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

Results for in-sample forecasts of yield changes are reported in Table 8. Out-of-sample forecasts appear in Table 9. In both tables, forecast errors are defined as the actual yield change minus $E_t[Y_{t+1}] - \hat{Y}_t$, where $E_t[Y_{t+1}]$ is the model-implied expectation and \hat{Y}_t is the model's current fitted value. The statistical significance of biases, which are averages of these errors, is assessed using Newey-West standard errors. Pairwise comparison tests for root mean squared errors are tested with the method of Diebold and Mariano (1995), also with Newey-West standard errors. The Newey-West lag length selection, following the procedure outlined for Tables 5 and 6, results in lags of 12 and 4 for the two panels of Table 8. For Table 9, 5 lags are used for both panels.

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. Out of sample forecasts are slightly more informative, perhaps because misspecified models tend to break down, and Table 9 shows small but significant advantages of $A_1(4)$ USV over $A_1(3)$ in forecasting short term yields, though all models display bias.

For volatility forecasting, we consider two alternative proxies for realized volatility. 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

$$\hat{\sigma}_{t,\tau} \equiv \sqrt{\sum_{i=1}^N \Delta Y(t, i, \tau)^2}, \quad (162)$$

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 9, 12, 16, and 17, 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 a significant bias in its forecasts of the daily realized volatility of the six-month yield. While the two USV specifications are roughly equivalent in their ability to forecast the volatility at the short end of the term structure, the $A_1(3)$ USV model clearly fails to describe the volatilities of longer yields. This may be an artifact of having a bad cross-sectional fit.

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 its long yield volatilities are somewhat biased.

The $A_1(3)$ USV model performs even worse out of sample in predicting long yield volatility and is dominated by the unrestricted $A_1(3)$ model.

This particular success of the $A_1(3)$ model is revealing of a more generally positive aspect of that model, namely the ability to match the term structure of unconditional yield volatility. In their review article, Dai and Singleton (2003) identify several empirical observations that have each proven somewhat of a challenge for affine term structure models. One of them is the fact that there is a “hump” in the plot of 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 1988-2002 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, which at least in part explains some of their poor performance in Tables 10 and 11. Both models come close to matching short rate volatilities but fail to match anything else – the hump generated by the $A_1(2)$ model is inverted and the model-implied volatilities for the $A_1(3)$ USV model are essentially flat. In contrast, both the unrestricted $A_1(3)$ and $A_1(4)$ USV models generate the right shape, at least approximately.

5.6 A closer look at short rate volatilities

Because of the special position occupied by the short rate in our model, we provide further evidence on the optimality of short rate volatility forecasts and the relation between that volatility and the shape of the yield curve. Specifically, we ask whether there is any possibility that short rate volatility is spanned by the yield curve, and more generally whether it is information from the cross section or the time series that is more useful in predicting volatility.

Table 12 contains the results of a variety of forecasting regressions. In the upper panel, the dependent variable is the absolute weekly difference in the six-month yield. In the lower panel, it is the realized volatility of the six-month yield, constructed from daily data. Both measures are the same as those used in previous tables. The forecasting variables include the first three principal components of the yield curve, a GARCH(1,1) volatility computed, as before, from daily six-month yield changes, and the same model forecasts from the unrestricted $A_1(3)$ and $A_1(4)$ USV models that were used previously. Because of deficiencies already identified, we exclude the $A_1(2)$ and $A_1(3)$ USV forecasts. Newey-West standard errors are reported throughout the table using six lags for the top panel and 11 for the bottom. Finally, in unreported results we have used raw yields instead of principal components, and the results are virtually identical.

In the first specification in the top panel, levels of the three principal components are used to forecast absolute yield changes. The result is an adjusted R-squared of 0.09, with the second and third principal components both displaying statistically significant slope coefficients. Thus, there appears to be information in the yield curve that is relevant for future volatility. The analogous

regression in the bottom panel results in the same finding, except that the adjusted R-squared of realized volatility is somewhat higher, at around 0.15.

We emphasize that these empirical results, namely, that cross sectional information has predictive power for volatility, is perfectly consistent with USV models. In particular, USV does *not* imply that changes in volatility are independent of changes in yields.³⁸ The correlation between shocks to the short rate and its variance is unrestricted in both the $A_1(3)$ USV and $A_1(4)$ USV models. We investigate this correlation structure in table 13, which contains two sets of regression estimates. In both cases, the dependent variable is the variance of the short rate and the explanatory variables are levels of the first three principal components of the yield curve. The first set of coefficients are implied by the posterior distribution of the parameters of the $A_1(4)$ USV model. This model, because it is stationary, implies an unconditional covariance matrix of the state vector. Since principal components are linear functions of the state vector we may therefore derive the joint covariance matrix of V_t and the first three principal components. This unconditional covariance matrix implies a set of “betas” in the regression of V_t on those principal components. Since the covariance matrix is parameter-dependent, we compute it for the $A_1(4)$ USV point estimates reported in Table 3 and for all other draws from the posterior distribution. Using the other draws we are able to calculate posterior standard deviations of these model-implied regression coefficients. The second set of regression coefficients come from a simpler “model-free” approach. Here we simply regress a GARCH(1,1) proxy of V_t on the three principal components observed in the sample. The table reports OLS coefficient estimates and Newey-West standard errors using 20 lags. Though we know of no statistically valid method to formally compare the two sets of coefficients, they are broadly consistent. Under the $A_1(4)$ USV model, volatility is weakly positively related to the first principal component, strongly negatively related to the second, and indeterminately related to the third. Using GARCH variance and the sample PCs, the relation between short rate variance and the third principal component is positive, but it has a magnitude that is consistent with the posterior distribution of the $A_1(4)$ USV model. Returning to Table 12, we see the same patterns. It is therefore consistent with the $A_1(4)$ USV model that the shape of the yield curve alone can be useful in forecasting volatility, as regressions 1 and 1' both show.

For absolute yield changes, GARCH-based forecasts are about as informative as PC-based forecasts, but for predicting realized daily volatility GARCH is far superior. Presumably, this has something to do with the fact that the GARCH forecasts, like the realized volatilities, are based on daily data. Combining GARCH and the three principal components, regression 3 shows that there is little incremental value in including both predictors. The coefficients on the second and third principal components decline in magnitude but remain significant, while the coefficient on GARCH volatility also declines but maintains a t-statistic of around 4.6. Stronger results obtain for the realized volatility regression 3', where the inclusion of the three PCs offers little benefit over the GARCH

³⁸USV models only predict that changes in volatility cannot be *perfectly* hedged by changes in yields – a property which is inconsistent with non-USV models.

measure alone.

Embedding forecasts from the unrestricted $A_1(3)$ model into the regression models 4, 5, 4', and 5' of table 12 results in strikingly negative coefficients, whether or not principal component levels are also included in the regression. These coefficients, which should equal one if the forecasts are conditionally unbiased, are instead significantly negative, providing further evidence on the misspecification of the model. This result is clearly related to the finding in Table 7 of a negative correlation between GARCH volatility and the volatility inferred under the $A_1(3)$ model.

Finally, forecasts from the $A_1(4)$ USV model perform similarly to those from the GARCH model. This is actually quite striking given that the $A_1(4)$ USV model was estimated from weekly data, while the GARCH model has the advantage of being fitted from daily data. In fact, the $A_1(4)$ USV forecasts of volatility are even stronger than GARCH in the sense that they drive out the significance of the information in the cross-section of the yield curve, i.e., in the three principal components, in 7'. On the other hand, even for the best performing $A_1(4)$ USV model the coefficient in the predictive regression is 0.655 and statistically different from one, which indicates that the forecast is not unbiased.³⁹

The similarities between GARCH and USV volatilities highlight the importance of inferring volatility from time series and not cross sectional information, since only time series information can be used to compute volatility under USV, as it is with GARCH models.

Overall, the results in Table 12 demonstrate that time series-based volatility proxies contain the vast majority of information relevant for predicting future volatilities, and that the information contained in the yield curve alone is insufficient for producing accurate forecasts. It seems therefore unambiguous that short rate volatility is not “spanned” by the yield curve.

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 identifiable). Further, model-insensitive estimates of the process of the state variable are readily available, which simplifies the empirical estimation of the model and makes direct comparisons of parameters and state variables across models and data sets.

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

³⁹We note that there are at least two possibilities for this. First, as shown in Ahn et al (2003) multiple *volatility* factors might actually be necessary for explaining different dynamics in short and long maturity yield volatilities. Second, the forecasts from the $A_1(4)$ USV model were constructed sub-optimally. Specifically, we constructed forecasts by conditioning on the point estimates in Table 3 rather than integrating across the entire posterior distribution. Unfortunately, the alternative is computationally unfeasible, and we see no way to measure the potential impact of this simplification.

dual role played by the volatility state variable in the unrestricted affine model: it is both a linear combination of yields (i.e., it affects the cross-section of the term structure) and the quadratic variation of the short rate (i.e., it impacts the time series of the term structure). Bayesian estimation results in more weight 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] D.-H. Ahn, R. F. Dittmar, and A. R. Gallant. Quadratic term structure models: Theory and evidence. *The Review of Financial Studies*, v15:243–288, 2001.
- [2] D.-H. Ahn, R. F. Dittmar, A. R. Gallant, and B. Gao. Purebred or hybrid?: Reproducing the volatility in term structure dynamics. *Journal of Econometrics*, v116:147–180, 2003.
- [3] T. G. Andersen, L. Benzoni, and J. Lund. An empirical investigation of continuous-time equity return models. *Journal of Finance*, VLII:1239–1284, 2002.
- [4] P. Balduzzi, S. R. Das, and S. Foresi. The central tendency: A second factor in bond yields. *Review of Economics and Statistics*, 80:62–72, 1998.
- [5] C. A. Bester. Random field and affine models for interest rates: An empirical comparison. Working Paper, 2004.
- [6] R. Bikbov and M. Chernov. Term structure and volatility: Lessons from the eurodollar market. Working Paper, 2004.
- [7] R. Bliss. Testing term structure estimation methods. *Advances in Futures and Options Research*, 9:197–231, 1997.
- [8] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.
- [9] M. Brandt and P. He. Simulated maximum likelihood estimation of affine models. Working paper, Wharton School, 2002.
- [10] G. Chacko. Multifactor interest rate dynamics and their implications for bond pricing. Working paper Harvard University, 1997.
- [11] P. Collin-Dufresne and R. S. Goldstein. Efficient pricing of swaptions in the affine framework. forthcoming in *The Journal of Derivatives*, 2001.
- [12] P. Collin-Dufresne and B. Solnik. On the term structure of default premia in the swap and LIBOR markets. *Journal of Finance*, 56:1095–1115, 2001.
- [13] J. C. Cox, J. E. Ingersoll Jr., and S. A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53:385–407, 1985.
- [14] Q. Dai and K. Singleton. Term structure dynamics in theory and reality. *The Review of Financial Studies*, 16:631–678, 2003.
- [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.-C. Duan 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.
- [18] J. Duarte. Evaluating an alternative risk preference in affine term structure models. Working Paper, Forthcoming Review of Financial Studies 2003.
- [19] G. R. Duffee. Idiosyncratic variation of treasury bill yields. *Journal of Finance*, v51:527–551, 1996.
- [20] G. R. Duffee. Term premia and interest rate forecasts in affine models. *Journal of Finance*, 57, 2002.
- [21] G. R. Duffee and R. Stanton. Estimation of dynamic term structure models. Working paper, UC Berkeley, 2002.
- [22] D. Duffie. *Dynamic Asset Pricing Theory*. Princeton: University Press, 1996.
- [23] D. Duffie and R. Kan. A yield-factor model of interest rates. *Mathematical Finance*, 6:379–406, 1996.
- [24] D. Duffie and K. Singleton. An econometric model of the term structure of interest rate swap yields. *Journal of Finance*, 52:1287–1381, 1997.
- [25] O. Elerian, S. Chib, and N. Shephard. Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*, v69 n4:959–993, 2001.
- [26] B. Eraker. Mcmc analysis of diffusion models with application to finance. *Journal of Business and Economic Statistics*, v19 n2:177–191, 2001.
- [27] P. Glasserman and Y. Jin. Equilibrium positive interest rates: a unified view. *The Review of Financial Studies*, v14:187–214, 2001.
- [28] N. Jegadeesh and G. G. Pennacchi. The behavior of interest rates implied by the term structure of eurodollar futures. *Journal of Money, Credit and Banking*, 28:426–446, 1996.
- [29] M. Johannes and S. Sundaresan. Collateralized swap spreads. Working paper, Columbia University, 2002.
- [30] C. S. Jones. Bootstrapping zero coupon yields from asynchronous LIBOR and swap quotes. Working paper, Univeristy of Southern California, 2003.
- [31] C. S. Jones. Nonlinear mean reversion in the short-term interest rate review of financial studies. *The Review of Financial Studies*, v16 n3:793–843, 2003.

- [32] C. G. Lamoureux and H. D. Witte. Empirical analysis of the yield curve: The information in the data viewed through the window of cox, ingsoll, and ross. *Journal of Finance*, v57 n3:1479–1520, 2002.
- [33] R. Litterman and J. Scheinkman. Common factors affecting bond returns. *Journal of Fixed Income*, june:54–61, 1991.
- [34] R. Litterman, J. Scheinkman, and L. Weiss. Volatility and the yield curve. *Journal of Fixed Income*, june, 1991.
- [35] J. Lund. Econometric analysis of continuous-time arbitrage-free models of the term structure of interest rates. Working Paper, 1997.
- [36] W. K. Newey and K. D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708, 1987.
- [37] W. K. Newey and K. D. West. Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies*, 61:631–653, 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] N. G. Polson, J. R. Stroud, and P. Müller. Affine state-dependent variance models. Working Paper, 2001.
- [40] A. E. Raftery. *Markov Chain Monte Carlo in Practice*, chapter Hypothesis Testing and Model Selection. Chapman and Hall, Boca Raton, Florida, 1996.
- [41] T. J. Rothenberg. Identification in parametric models. *Econometrica*, 39:577–591, 1971.
- [42] K. J. Singleton and L. Umantsev. Pricing coupon-bond options and swaptions in affine term structure models. *Mathematical Finance*, 12, 2002.
- [43] H. Tang and Y. Xia. An international examination of affine term structure models and the expectations hypothesis. Working Paper, 2005.
- [44] 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 (27), (28) and (29)

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

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

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

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

Note that this allows us to define

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

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

Finally, from Ito's lemma we have

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

Using the relationship between yield to maturity and bond prices

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

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

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

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

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

Bond prices satisfy the PDE

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

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

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

Now we use a Taylor series expansion to write yields as

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

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

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

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

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

which is what we wished to prove. \square

A.2: Proof of Proposition 1

To prove the proposition note that it is sufficient to show that $e^{-\int_0^t r_s ds} P(t, T)$ is a Q -martingale for P as defined in equation (85). Indeed, in that case we have $e^{-\int_0^t r_s ds} P(t, T) = \mathbb{E}_t^Q \left[e^{-\int_0^T r_s ds} P(T, T) \right]$, which implies

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

since equations (85)-(88) 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 (85). 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

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

Thus,

$$\begin{aligned} e^{-\int_0^t r_s ds} P(t, T) &= - \int_0^t \sqrt{2B_\mu(s)} \left(\sigma_1 \sqrt{V_s - \psi_1} dZ_1^Q(s) + \sqrt{(1 - \sigma_1^2)V_s + \sigma_1^2 \psi_1 + \psi_3} dZ_2^Q(s) \right) \\ &\quad - \int_0^t (B_r(s) + \nu_3 B_\mu(s)) \sqrt{-\psi_3} dZ_3^Q(s). \end{aligned} \quad (\text{A.20})$$

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

□

A.3: Proof of Proposition 2

To prove the proposition it is sufficient to show that $e^{-\int_0^t r_s ds} P(t, T)$ is a Q -martingale for P as defined in equation (85). Indeed, in that case we have $e^{-\int_0^t r_s ds} P(t, T) = \mathbb{E}_t^Q \left[e^{-\int_0^T r_s ds} P(T, T) \right]$, which implies

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

since equations (134)-(138) 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 (134). Using the fact that the functions $A(\cdot)$, $B_r(\cdot)$ and $B_\mu(\cdot)$ satisfy the system of ODE:

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

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

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

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

and that, in particular, because of the restrictions on a_r, a_μ given in equations (108) and (109), we have:

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

we find that

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

Therefore,

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

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

□

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

B Details of the MCMC procedure

As summarized in the text, our MCMC algorithm is similar to a Gibbs sampler that alternates between drawing unobservable state variables and model parameters. We further decompose the

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

- $p(V_t | \mathcal{P}, V_1, \dots, V_{t-h}, V_{t+h}, \dots, V_T, X^o, \phi)$ for each $t \in \{1, 1+h, 1+2h, \dots, T\}$.
- $p(\phi^\Omega | \mathcal{P}, V, X^o, \phi^\lambda, \phi^Q)$
- $p(\phi^\lambda | \mathcal{P}, V, X^o, \phi^\Omega, \phi^Q)$
- $p(\phi^Q, X^o | \mathcal{P}, V, \phi^\Omega, \phi^\lambda)$

In all blocks, draws are rejected if they violate stationarity, admissibility, or Feller conditions, or if they imply degenerate covariance matrices.

B.1: The linear state space representation

In all the models we consider, only V_t enters the conditional variance of the state vector X_{t+h} , so we may rewrite (156) as

$$X_{t+h} \sim N(h\mathcal{K}\Theta + (I - h\mathcal{K})X_t, U_t), \quad (\text{B.28})$$

where $U_t = h\Sigma S_t \Sigma'$ is a function of V_t only. Partitioning the state vector as

$$X_t = \begin{bmatrix} V_t \\ X_t^o \end{bmatrix}, \quad (\text{B.29})$$

standard properties of the multivariate normal distribution imply that the mean of X_{t+h}^o is linear in X_t^o with a variance that is time-varying but only as a function of V_t . Thus, the conditional dynamics of X_t^o given the full path of V (the ‘‘state equation’’) are both linear and Gaussian (at a time horizon of length h).

At the same time, the observed data are linear in the unobserved vector X_t^o . Since the affine structure implies that yields are linear in state variables, and because principal components are simple linear combinations of yields, there exist a vector K and a matrix L such that in the absence of measurement error we would have

$$\mathcal{P}_t = K + LX_t \quad (\text{B.30})$$

After adding a Gaussian error vector $e_t \sim N(0, \Omega)$ and breaking up X_t , we are left with a Gaussian ‘‘measurement equation’’ that is linear in X_t^o ,

$$\mathcal{P}_t = K + L^v V_t + L^o X_t^o + e_t, \quad (\text{B.31})$$

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

Thus, conditional on the path of V , we have both state and measurement equations that are Gaussian and linear in the state variable X_t^o . This enables the use of the standard Kalman filter to compute $p(\mathcal{P}|\phi, V)$. One small complication is that our state equation defines transitions over a unit of time of length h , while the measurement equation is only applicable for observation times $t \in \{1, 2, \dots, T\}$. To resolve this asymmetry, consider the equivalent situation where yields were instead observed at every length- h interval, but that the measurement error variance for non-integer t was infinitely large. For non-integer t , the Kalman “gain” matrix is then zero, meaning that the observed data has no effect on the conditional distribution of the state vector. Thus, we can apply the Kalman filter in its textbook form simply by zeroing out the Kalman gain matrix when $t \notin \{1, 2, \dots, T\}$.

B.2: Drawing V_t

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

$$\begin{aligned} & p(V_t|\mathcal{P}, V_1, \dots, V_{t-h}, V_{t+h}, \dots, V_T, X^o, \phi) \\ & \propto p(V_t|\mathcal{P}_t, V_{t-h}, V_{t+h}, X_{t-h}^o, X_t^o, X_{t+h}^o, \phi) \\ & \propto p(V_{t+h}, X_{t+h}^o|\mathcal{P}_t, V_{t-h}, V_t, X_{t-h}^o, X_t^o, \phi)p(V_t|\mathcal{P}_t, V_{t-h}, X_{t-h}^o, X_t^o, \phi) \\ & \propto p(V_{t+h}, X_{t+h}^o|V_t, X_t^o, \phi)p(V_t|\mathcal{P}_t, V_{t-h}, X_{t-h}^o, X_t^o, \phi) \end{aligned}$$

The first and third steps both use the Markov property along with the fact that \mathcal{P}_s is an irrelevant conditioning variable when we are also conditioning on X_s . The middle step simply uses Bayes rule.

Note that $p(V_t|\mathcal{P}_t, V_{t-h}, X_{t-h}^o, X_t^o, \phi)$ is a Gaussian density for V_t . This is because the joint density of $\{V_t, X_t^o, \mathcal{P}_t\}$ given $\{V_{t-h}, X_{t-h}^o\}$ is multivariate Gaussian (from the Euler approximation and the Gaussian measurement errors), and $p(V_t|\mathcal{P}_t, V_{t-h}, X_{t-h}^o, X_t^o, \phi)$ is simply a conditional version of that density. Unfortunately, this is not our target density, as it ignores information from time $t + h$. We therefore use this density as the candidate generator for a Metropolis-Hastings draw, forming the acceptance factor from the omitted component of the target density. We therefore replace the current draw V_t with the candidate V_t^* with probability

$$\min \left\{ \frac{p(V_{t+h}, X_{t+h}^o|V_t^*, X_t^o, \phi)}{p(V_{t+h}, X_{t+h}^o|V_t, X_t^o, \phi)}, 1 \right\}.$$

This acceptance probability, as the ratio of multivariate Gaussians, is straightforward to evaluate. This produces a draw from the target density, $p(V_t|\mathcal{P}, V_1, \dots, V_{t-h}, V_{t+h}, \dots, V_T, X^o, \phi)$, as desired.

For $t \notin \{1, 2, \dots, T\}$, the draw is almost identical, but the target density is instead

$$p(V_{t+h}, X_{t+h}^o|V_t, X_t^o, \phi)p(V_t|V_{t-h}, X_{t-h}^o, X_t^o, \phi).$$

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

B.3: Drawing ϕ^Ω

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

B.4: Drawing ϕ^λ

To draw ϕ^λ we write the Euler approximation (156) as

$$X_{t+h} - X_t = h(a + bX_t) + \sqrt{h}\Sigma\sqrt{S_t}\epsilon_{t+h}.$$

Since the drift is linear and the covariance matrix is known (as a function of V and ϕ^Q), we can directly apply the seeming unrelated regression approach of Chib and Greenberg (1996) to draw the a vector and the b matrix. Since we are using the ‘generalized essentially affine’ risk premia of Cheridito et al (2004), each nonzero element of a and b is effectively a free parameter (subject to stationarity and admissibility conditions). Thus, no linkage between the P and Q drift parameters need be imposed and we can simply back out risk premia according to

$$\lambda_0 + \lambda_1 X_t = a + bX_t - a^Q - b^Q X_t,$$

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

B.5: Drawing ϕ^Q and X^o

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

$$p(\phi^Q | \mathcal{P}, V, \phi^\Omega, \phi^\lambda) p(X^o | \mathcal{P}, V, \phi^Q, \phi^\Omega, \phi^\lambda).$$

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

$$\begin{aligned} p(\phi^Q | \mathcal{P}, V, \phi^\Omega, \phi^\lambda) &\propto p(\mathcal{P}, V | \phi^Q, \phi^\Omega, \phi^\lambda) \\ &\propto p(\mathcal{P} | V, \phi^Q, \phi^\Omega, \phi^\lambda) p(V | \phi^Q, \phi^\Omega, \phi^\lambda) \end{aligned}$$

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

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

$$\min \left\{ \frac{p(\mathcal{P}|V, \phi^{Q*}, \phi^\Omega, \phi^\lambda)p(V|\phi^{Q*}, \phi^\Omega, \phi^\lambda)}{p(\mathcal{P}|V, \phi^Q, \phi^\Omega, \phi^\lambda)p(V|\phi^Q, \phi^\Omega, \phi^\lambda)}, 1 \right\},$$

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

Following Bester (2004), we alternate between usually drawing the entire ϕ^Q vector at once using a multivariate candidate generator and occasionally (once every ten iterations) drawing each element of ϕ^Q individually. The covariance matrix of the candidate generator is chosen by running a long preliminary chain and computing the sample covariance matrix of the draws of ϕ^Q from the chain. A second chain is run in which a scaling parameter is chosen adaptively to set the Metropolis acceptance rate approximately equal to .4. A final third chain is run to generate the posteriors reported.

B.6: Sensitivity to the choice of h

The results of Jones (2003b) and Eraker (2001) suggest that even nonlinear term structure models do not suffer from appreciable discretization bias when the discretization interval is set equal to one day. Given our use of weekly data, this suggests that at a minimum we should investigate values of h as small as .2 (one fifth of the weekly observation interval). In this section we compare the results presented in the paper, which were computed using $h = 1$, to the case in which $h = .2$. To reduce computation time, we calculated results only for the $A_1(3)$ USV model.

Figure B1 displays posterior histograms for the two cases. In each pair of histograms, the top panel represents the $h = 1$ case while the bottom case represents $h = .2$. Differences between the two are indistinguishable with the exception of η_2 , whose posterior shape (though not location or dispersion) is slightly different between the two panels. Figure B2 overlays the fitted state variables that result from the two procedures. Very small differences are noticeable for estimated short rate volatility $\sqrt{V_t}$, but not for r_t or μ_t^Q . Finally, Table B1 reports in- and out-of-sample yield fits. Again, differences are very minor and are clearly insufficient to change any of the conclusions of the paper.

Table 1: Observability of state variables

The table contains output from the regressions

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

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

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

Table 2: Principal component loadings

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

	Principal Component					
	1	2	3	4	5	6
6-month	0.08	2.37	1.98	14.08	15.35	216.68
1-year	0.11	1.93	0.06	-16.81	-37.29	-696.92
2-year	0.14	0.99	-1.26	-11.12	20.30	1231.63
3-year	0.14	0.19	-1.15	4.12	22.47	-387.75
4-year	0.14	-0.42	-0.72	11.72	1.23	-878.34
5-year	0.14	-0.87	-0.22	12.53	-18.09	-274.75
7-year	0.13	-1.42	0.69	3.49	-26.74	1351.94
10-year	0.12	-1.77	1.61	-17.01	23.77	-561.48
% explained	64.82	17.74	8.04	5.37	2.73	1.04
Total % explained by first six principal components: 99.73						

Table 3A: Posterior distributions of model parameters

Posterior distributions are calculated from weekly bootstrapped yields from January 1988 to December 2002. Both free parameters and restricted parameters are included, where restricted parameters are functions of the free parameters and are displayed in italics. For free parameters, the point estimates displayed are multivariate posterior medians. Point estimates for restricted parameters are computed as functions of the free parameter point estimates. Confidence interval bounds, in parentheses, are equal to the 2.5 and 97.5 percentiles of the posterior distribution.

	$A_1(2)$	$A_1(3)$ USV	$A_1(3)$	$A_1(4)$ USV
σ_1	0.007 (-0.039, 0.029)	0.062 (-0.109, 0.316)	-0.248 (-0.330, -0.144)	0.176 (0.019, 0.377)
σ_2		0.998 (0.949, 1.000)	0.955 (0.939, 0.987)	0.984 (0.926, 1.000)
σ_3		0.000	0.164 (0.012, 0.155)	0.000
ν_1		-0.008 (-0.039, 0.014)	-0.687 (-1.264, -0.316)	0.001 (0.000, 0.003)
ν_2		-0.123 (-0.129, -0.119)	-4.412 (-5.067, -3.817)	0.008 (0.007, 0.009)
ν_3		-30.613 (-42.621, -13.160)	5.581 (3.081, 9.525)	0.016 (0.009, 0.021)
ν_4				0.011 (0.000, 0.018)
$\sigma_V \times 10^3$	0.076 (0.036, 0.111)	1.196 (0.894, 1.525)	0.195 (0.172, 0.210)	1.094 (0.923, 1.356)
$\psi_1 \times 10^5$	6.520 (0.094, 8.327)	0.134 (0.007, 0.502)	2.162 (1.251, 3.128)	0.192 (0.095, 0.428)
$\psi_2 \times 10^5$		0.001 (-0.022, 0.009)	1.254 (0.650, 2.439)	0.178 (0.067, 0.362)
$\psi_3 \times 10^5$		-0.002 (-0.011, -0.001)	-1.388 (-2.630, -0.769)	-0.142 (-0.251, -0.015)
$\psi_4 \times 10^5$				-0.042 (-0.238, -0.015)
κ_V^P	0.301 (0.062, 0.537)	1.003 (0.352, 2.332)	0.297 (0.057, 0.746)	1.134 (0.267, 1.886)
$\gamma_V^P \times 10^4$	0.294 (0.072, 0.502)	0.835 (0.407, 1.603)	0.526 (0.145, 1.064)	0.668 (0.350, 1.016)
$m_0 \times 10^2$		0.256 (0.240, 0.281)	20.008 (18.382, 22.782)	
m_V		1.000	-765.952 (-934.757, -683.993)	
m_r		-0.030 (-0.033, -0.028)	-1.270 (-1.381, -1.166)	
m_μ		-0.370 (-0.386, -0.358)	-2.827 (-2.949, -2.646)	
Additional parameters for the $A_1(2)$ model				
	γ_r	κ_{rr}	κ_{rV}	
	0.167 (0.112, 0.296)	0.501 (0.486, 0.523)	1383.344 (775.024, 2825.008)	
Additional parameters for the $A_1(4)$ USV model				
$a_0 \times 10^3$	a_θ	η_3	η_4	$c_{r\mu}$
1.364 (1.147, 1.696)	-1.443 (-1.639, -1.316)	-0.020 (-0.031, -0.009)	-0.019 (-0.029, -0.003)	-0.089 (-0.093, -0.084)

Table 3B: Posterior distributions of risk premia parameters

Posterior distributions are calculated from weekly bootstrapped yields from January 1988 to December 2002. Point estimates displayed are multivariate posterior medians. Confidence interval bounds, in parentheses, are equal to the 2.5 and 97.5 percentiles of the posterior distribution.

	$A_1(2)$	$A_1(3)$ USV	$A_1(3)$	$A_1(4)$ USV
$\lambda_{V_0} \times 10^4$	0.264 (0.040, 0.469)		0.500 (0.121, 1.040)	
λ_{VV}	-0.264 (-0.496, -0.022)		-0.260 (-0.711, -0.021)	
λ_{r_0}	-0.097 (-0.195, 0.020)	-0.040 (-0.059, -0.017)	0.002 (-0.055, 0.050)	-0.031 (-0.051, -0.007)
$\lambda_{rV} \times 10^{-2}$	6.674 (-5.468, 17.139)	-3.238 (-4.444, -1.487)	-0.887 (-3.390, 2.248)	-2.024 (-2.821, -0.754)
λ_{rr}	0.399 (0.141, 0.497)	0.672 (0.314, 1.009)	0.023 (-0.343, 0.347)	0.458 (0.084, 0.767)
$\lambda_{r\mu}$		1.618 (0.400, 2.722)	-0.357 (-0.603, -0.135)	0.700 (-1.144, 1.736)
$\lambda_{r\theta}$				0.588 (-0.817, 1.277)
λ_{μ_0}		0.024 (0.015, 0.040)	-0.096 (-0.339, 0.099)	0.080 (0.019, 0.165)
$\lambda_{\mu V} \times 10^{-2}$		1.761 (1.050, 2.454)	4.390 (-6.335, 18.414)	3.454 (0.709, 6.503)
$\lambda_{\mu r}$		-0.444 (-0.715, -0.302)	0.329 (-0.866, 1.898)	-1.462 (-2.630, -0.447)
$\lambda_{\mu\mu}$		-1.307 (-2.160, -0.760)	1.030 (-0.200, 1.902)	-2.265 (-8.472, 1.773)
$\lambda_{\mu\theta}$				-0.796 (-5.468, 1.956)
λ_{θ_0}				-0.080 (-0.204, 0.005)
$\lambda_{\theta V} \times 10^{-2}$				-3.324 (-8.069, 0.360)
$\lambda_{\theta r}$				1.544 (0.129, 3.288)
$\lambda_{\theta\mu}$				0.869 (-4.557, 10.022)
$\lambda_{\theta\theta}$				-0.596 (-4.216, 6.229)

Table 4: Specification analysis

This table reports properties of realized normalized model errors. For a each posterior draw of the matrix of state variables, we define the normalized residuals

$$\hat{\epsilon}_t = (\Delta X_t - E_{t-1}[\Delta X_t]) / SD_{t-1}(\Delta X_t)$$

where X corresponds to either V , r , μ^Q , or θ . Moments are computed using the corresponding posterior draw of the parameter vector and the lagged state vector. A variety of statistics are calculated for each time series of residuals for each posterior draw. The posterior medians of these statistics are reported in the table. Statistics that have more than 95% posterior probability of taking the reported sign are marked with an asterisk.

	$A_1(2)$			$A_1(3)$ USV			$A_1(3)$			$A_1(4)$ USV			θ
	V	r	μ^Q	V	r	μ^Q	V	r	μ^Q	V	r	μ^Q	
Mean $\hat{\epsilon}$	-0.02	0.02*	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.00
SD of $\hat{\epsilon}$	1.00	0.98	1.09*	0.99	0.93*	0.96	0.99	0.97	0.97	0.99	0.97	0.99	0.99
Skewness of $\hat{\epsilon}$	-0.11*	-0.67*	0.04	-0.03	-0.34*	0.06	-0.05	0.01	0.03	0.06	0.01	-0.02	0.01
Kurtosis of $\hat{\epsilon}$	3.44*	7.29*	2.94	3.75*	7.24*	4.18*	4.21*	4.44*	4.01*	2.95	4.44*	4.02*	4.04*
1st order AC of $\hat{\epsilon}$	0.10*	0.19*	0.00	0.12*	0.05*	0.00	-0.11*	0.03*	0.04*	0.01	0.03*	0.02*	0.02*
1st order AC of $ \hat{\epsilon} $	-0.04*	0.15*	0.00	0.02	0.23*	-0.02	0.06*	0.01	0.11*	0.00	0.01	0.10*	0.09*
1st order AC of $\hat{\epsilon}^2$	-0.03*	0.06*	0.00	0.01	0.17*	-0.02	0.07*	0.00	0.08*	0.00	0.00	0.08*	0.08*

Table 5: In-sample yield fits

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

	$A_1(2)$	$A_1(3)$ USV	$A_1(3)$	$A_1(4)$ USV
mean \hat{e} (basis points)				
6-month	-0.01	-5.24**	-0.18	0.69
1-year	0.01	-1.04	0.24	-0.75
2-year	0.03	3.28**	0.09	-0.74
3-year	0.04	4.20**	-0.06	0.17
4-year	-0.02	3.27**	-0.12	0.70
5-year	-0.08	1.53*	-0.11	0.76
7-year	-0.11	-2.22**	-0.04	0.10
10-year	0.15	-5.62**	0.19	-0.90
RMSE (basis points)				
6-month	14.44	14.61	>** 3.42	<** 4.82
1-year	6.53	6.58	>** 4.51	<** 6.01
2-year	11.27	10.36	>** 2.56	<** 3.71
3-year	9.84	10.05	>** 1.54	<** 1.67
4-year	7.40	7.95	>** 2.64	<** 3.84
5-year	5.55	5.50	>** 2.93	<** 4.27
7-year	5.80	5.55	>** 1.65	<** 1.88
10-year	11.08	14.11	>** 4.03	<** 5.89
autocorrelation of \hat{e}				
6-month	0.94	0.96	0.91	0.95
1-year	0.49	0.72	0.89	0.95
2-year	0.74	0.80	0.81	0.93
3-year	0.67	0.79	0.69	0.90
4-year	0.46	0.71	0.84	0.95
5-year	0.11	0.53	0.85	0.95
7-year	0.27	0.57	0.66	0.88
10-year	0.80	0.92	0.89	0.95

Table 6: Out-of-sample yield fits

This table contains statistics on the in-sample fits of zero coupon yields (Y). For each model, fitted yields (\hat{Y}_t) are calculated for .5, 1, 2, 3, 4, 5, 7, and 10-year maturities. The table examines the bias, root mean squared error, and autocorrelation of $\hat{e}_t = Y_t - \hat{Y}_t$, where \hat{Y}_t denotes the model fitted value. * and ** denote statistical significance at the 5% and 1% levels, respectively, where standard errors are calculated using the method of Newey and West (1987) with 8 lags for bias tests and 7 for RMSE tests. 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 104 weeks.

	$A_1(2)$	$A_1(3)$ USV	$A_1(3)$	$A_1(4)$ USV
mean \hat{e} (basis points)				
6-month	13.56**	-2.74	-5.98**	1.43**
1-year	2.50**	-2.13**	6.57**	-1.60*
2-year	-7.53**	1.29	5.67**	-1.14**
3-year	-10.60**	2.74	-1.22**	0.11
4-year	-9.62**	2.81*	-5.38**	1.02**
5-year	-6.24**	2.11**	-6.32**	1.49**
7-year	3.59**	-0.51	-2.19**	1.14**
10-year	17.99**	-6.27*	8.35**	-2.34**
RMSE (basis points)				
6-month	18.72	12.48	>** 6.44	>** 2.69
1-year	4.59	4.25	<** 7.09	>** 3.41
2-year	12.05	8.87	>* 6.12	>** 1.88
3-year	14.23	>* 9.00	>** 1.46	>** 0.85
4-year	12.26	>** 7.20	5.83	>** 1.96
5-year	8.19	>** 4.99	<* 6.82	>** 2.32
7-year	5.93	4.58	>** 2.44	>** 1.29
10-year	21.07	>* 11.39	8.94	>** 3.41
autocorrelation of \hat{e}				
6-month	0.96	0.96	0.76	0.78
1-year	0.15	0.21	0.77	0.82
2-year	0.73	0.71	0.84	0.67
3-year	0.72	0.68	0.43	0.84
4-year	0.59	0.52	0.73	0.73
5-year	0.27	0.17	0.75	0.71
7-year	0.41	0.55	0.56	0.75
10-year	0.92	0.92	0.82	0.73

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. Rolling 30-day window and GARCH(1,1) volatilities are calculated from demeaned changes in the six-month rate. Interpolated r , μ^Q , and θ are calculated using a third-order polynomial regression of yields on maturity. The Implied Volatility series, obtained from one-year cap and floor contracts, is an average of Black-Scholes implied volatilities on the logarithm of the one-year LIBOR rate times the level of the one-year rate. 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	Rolling	Other volatility measures	
					GARCH	Implied	
Full Sample (1988 to 2002)							
Actual vs. model average yield	1.000	1.000	1.000	1.000			
Actual vs. model slope	0.998	0.997	0.998	0.998			
Actual vs. model curvature	0.332	0.379	0.998	0.997			
Rolling vs. model volatility	-0.573	0.758	-0.595	0.777			
GARCH vs. model volatility	-0.558	0.749	-0.580	0.773	0.949		
Interpolated vs. model r	0.996	0.996	0.981	0.998			
Interpolated vs. model μ^Q		0.880	0.513	0.974			
Interpolated vs. model θ				0.976			
Actual curvature vs. model volatility	0.415	-0.077	0.275	-0.102	-0.058	-0.050	
Actual curvature vs. model variance	0.419	-0.065	0.287	-0.088	-0.028	-0.019	
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	0.262	0.401	0.998	0.997			
Rolling vs. model volatility	-0.465	0.583	-0.460	0.633			
GARCH vs. model volatility	-0.455	0.566	-0.451	0.626	0.935		
Interpolated vs. model r	0.997	0.997	0.985	0.999			
Interpolated vs. model μ^Q		0.862	0.454	0.975			
Interpolated vs. model θ				0.979			
Actual curvature vs. model volatility	0.457	-0.151	0.276	-0.246	-0.126	-0.132	
Actual curvature vs. model variance	0.458	-0.093	0.282	-0.176	-0.048	-0.057	
1995 to 2002							
Actual vs. model average yield	0.999	1.000	1.000	1.000			
Actual vs. model slope	0.998	0.997	0.998	0.998			
Actual vs. model curvature	0.634	0.478	0.999	0.998			
Rolling vs. model volatility	0.006	0.726	-0.027	0.719			
GARCH vs. model volatility	0.030	0.739	-0.004	0.736	0.920		
Implied vs. model volatility	-0.192	0.340	-0.289	0.309	0.546	0.490	
Interpolated vs. model r	0.996	0.995	0.975	0.997			
Interpolated vs. model μ^Q		0.922	0.665	0.985			
Interpolated vs. model θ				0.992			
Actual curvature vs. model volatility	0.798	0.023	0.657	0.055	0.064	0.094	0.105
Actual curvature vs. model variance	0.798	0.008	0.657	0.052	0.063	0.096	0.102

Table 8: In-sample yield forecasts

This table contains statistics on the in-sample one-week forecasts of zero coupon yield changes. For each model, expected yield changes are calculated for .5, 1, 2, 3, 4, 5, 7, and 10-year maturities as differences between the model expectations of future yields and the current model fitted values. The table examines the bias, root mean squared error, and autocorrelation of $\hat{e}_{t+1} = (Y_{t+1} - Y_t) - (E_t[Y_{t+1}] - \hat{Y}_t)$, where $E_t[Y_{t+1}]$ is the model-implied expectation and \hat{Y}_t is the model fitted value. * and ** denote statistical significance at the 5% and 1% levels, respectively, where standard errors are calculated using the method of Newey and West (1987) with 12 lags for bias tests and 5 for RMSE tests. For biases, statistical significance relates to the null hypothesis that the bias is zero. For RMSE, the statistical significance of the pairwise comparison of two models is reported, along with an inequality sign that reflects the direction of the rejection. Forecasts begin at the beginning of 1990, allowing for a sample size of 677 weeks.

	$A_1(2)$	$A_1(3)$ USV	$A_1(3)$	$A_1(4)$ USV
	mean \hat{e} (basis points)			
6-month	-0.55	0.18	-0.34	-0.25
1-year	-0.54	0.15	-0.18	-0.10
2-year	-0.49	0.11	0.01	0.09
3-year	-0.44	0.08	0.11	0.18
4-year	-0.39	0.06	0.19	0.23
5-year	-0.35	0.04	0.24	0.26
7-year	-0.29	0.02	0.33	0.28
10-year	-0.20	-0.01	0.41	0.28
	RMSE (basis points)			
6-month	11.08	>**	10.84	10.96
1-year	13.37		13.26	13.34
2-year	14.97		14.93	14.99
3-year	15.04		15.02	15.07
4-year	14.87		14.86	14.91
5-year	14.71		14.69	14.74
7-year	14.44		14.44	14.46
10-year	14.24		14.24	14.25
	autocorrelation of \hat{e}			
6-month	0.04	-0.02	0.01	0.00
1-year	-0.01	-0.04	-0.02	-0.01
2-year	-0.01	-0.02	-0.01	0.00
3-year	-0.02	-0.02	-0.02	0.00
4-year	-0.04	-0.04	-0.04	-0.02
5-year	-0.06	-0.06	-0.06	-0.04
7-year	-0.08	-0.08	-0.08	-0.07
10-year	-0.10	-0.10	-0.10	-0.09

Table 9: Out-of-sample yield forecasts

This table contains statistics on the in-sample one-week forecasts of zero coupon yield changes. For each model, expected yield changes are calculated for .5, 1, 2, 3, 4, 5, 7, and 10-year maturities as differences between the model expectations of future yields and the current model fitted values. The table examines the bias, root mean squared error, and autocorrelation of $\hat{e}_{t+1} = (Y_{t+1} - Y_t) - (E_t[Y_{t+1}] - \hat{Y}_t)$, where $E_t[Y_{t+1}]$ is the model-implied expectation and \hat{Y}_t is the model fitted value. * and ** denote statistical significance at the 5% and 1% levels, respectively, where standard errors are calculated using the method of Newey and West (1987) with 5 lags. 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 104 weeks.

	$A_1(2)$	$A_1(3)$ USV	$A_1(3)$	$A_1(4)$ USV
mean \hat{e} (basis points)				
6-month	2.05**	2.51**	3.93**	1.65**
1-year	2.04*	2.49*	3.06**	0.90
2-year	1.67	2.08	2.12	-0.19
3-year	1.29	1.65	1.66	-0.76
4-year	0.97	1.28	1.35	-1.06
5-year	0.70	0.96	1.13	-1.22
7-year	0.29	0.46	0.79	-1.35
10-year	-0.11	-0.02	0.46	-1.34
RMSE (basis points)				
6-month	5.63	5.98	6.34	>** 5.43
1-year	8.83	9.08	9.16	>* 8.67
2-year	12.75	<* 12.89	12.86	12.61
3-year	14.63	14.72	14.71	14.52
4-year	15.65	15.71	15.71	15.56
5-year	16.10	16.13	16.15	16.02
7-year	16.08	16.08	16.12	16.01
10-year	15.35	15.33	15.38	15.29
autocorrelation of \hat{e}				
6-month	0.18	0.24	0.07	0.16
1-year	0.07	0.09	0.04	0.07
2-year	0.07	0.07	0.06	0.08
3-year	0.09	0.09	0.09	0.10
4-year	0.11	0.11	0.11	0.12
5-year	0.13	0.13	0.13	0.13
7-year	0.14	0.14	0.14	0.15
10-year	0.14	0.14	0.14	0.15

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 ($E[|\Delta Y|]$) and expected “realized volatility” ($E[\hat{\sigma}]$) are calculated for .5, 1, 2, 3, 4, 5, 7, and 10-year maturities. Realized volatility is defined by $\hat{\sigma}_{t,\tau}^2 = \sum_{i=1}^{\tau} \Delta Y(t, i, \tau)^2$ and is calculated using daily data. The table examines the forecast bias (actual minus forecast) and root mean squared error of $|\Delta Y|$ and $\hat{\sigma}$, where all yields are expressed in basis points. * and ** denote statistical significance at the 5% and 1% levels, where standard errors are calculated using the method of Newey and West (1987) with 9, 13, 16, and 17 lags, respectively, for the four panels of the table. For biases, statistical significance relates to the null hypothesis that the bias is zero. For RMSE, the statistical significance of the pairwise comparison of two models is reported, along with an inequality sign that reflects the direction of the rejection. Forecasts begin at the beginning of 1990, allowing for a sample size of 677 weeks.

	$A_1(2)$		$A_1(3)$ USV		$A_1(3)$		$A_1(4)$ USV
bias in weekly $ \Delta Y $							
6-month	-2.26**		-0.72*		-2.95**		-0.51
1-year	1.03*		1.82**		-1.63**		0.62*
2-year	3.35**		3.22**		-1.00*		0.45
3-year	3.57**		3.04**		-0.80		-0.07
4-year	3.29**		2.63**		-0.71		-0.31
5-year	2.93**		2.27**		-0.70		-0.36
7-year	2.34**		1.86**		-0.73		-0.25
10-year	2.02**		1.98**		-0.61		0.17
RMSE of weekly $ \Delta Y $							
6-month	8.68	>**	7.97	<**	8.93	>**	7.84
1-year	9.16		8.96		9.29	>**	8.74
2-year	10.25		9.99		9.80	>**	9.45
3-year	10.29	>*	9.98		9.76	>*	9.50
4-year	10.11	>*	9.82		9.64		9.46
5-year	9.91	>*	9.68		9.55		9.41
7-year	9.61		9.49		9.40		9.31
10-year	9.29		9.24		9.14		9.03
bias in $\hat{\sigma}$							
6-month	-3.61**		-1.79**		-4.45**		-1.53**
1-year	0.20		1.14**		-2.96**		-0.29
2-year	3.24**		3.08**		-1.95**		-0.22
3-year	3.73**		3.10**		-1.47**		-0.61
4-year	3.56**		2.77**		-1.22*		-0.74
5-year	3.25**		2.47**		-1.08*		-0.68
7-year	2.70**		2.12**		-0.97*		-0.40
10-year	2.28**		2.24**		-0.85		0.07
RMSE of $\hat{\sigma}$							
6-month	6.63	>**	5.29	<**	7.18	>**	5.18
1-year	6.18		5.90	<**	6.91	>**	5.82
2-year	7.61		7.38		7.23		6.74
3-year	7.79		7.46		7.06		6.84
4-year	7.54		7.22		6.82		6.74
5-year	7.24		6.96		6.62		6.58
7-year	6.78		6.59		6.36		6.30
10-year	6.47		6.37		6.20		6.01

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 ($E[|\Delta Y|]$) and expected “realized volatility” ($E[\hat{\sigma}]$) are calculated for .5, 1, 2, 3, 4, 5, 7, and 10-year maturities. Realized volatility is defined by $\hat{\sigma}_{t, \tau}^2 = \sum_{i=1}^5 \Delta Y(t, i, \tau)^2$ and is calculated using daily data. The table examines the forecast bias (actual minus forecast) and root mean squared error of ΔY^2 and $\hat{\sigma}$, where all yields are expressed in basis points. * and ** denote statistical significance at the 5% and 1% levels, where standard errors are calculated using the method of Newey and West (1987) with 3, 4, 4, and 4 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 104 weeks.

	$A_1(2)$		$A_1(3)$ USV		$A_1(3)$		$A_1(4)$ USV
bias in weekly $ \Delta Y $							
6-month	-5.93**		-0.85*		-6.78**		-1.67**
1-year	-2.33**		1.84**		-5.17**		-0.50
2-year	1.38*		4.35**		-3.33**		0.50
3-year	2.93**		5.31**		-1.81*		1.32
4-year	3.54**		5.64**		-0.82		2.00*
5-year	3.65**		5.62**		-0.33		2.43*
7-year	3.10**		5.04**		-0.29		2.59**
10-year	2.00*		4.16**		-0.92		2.17*
RMSE of weekly $ \Delta Y $							
6-month	6.90	>**	3.65	<**	7.65	>**	4.05
1-year	5.69		5.62	<**	7.34	>**	5.33
2-year	7.81	<**	8.96		8.39	>*	7.79
3-year	9.22	<**	10.32	>*	8.94		8.91
4-year	9.92	<**	10.93	>**	9.33		9.53
5-year	10.15	<**	11.08	>**	9.51		9.82
7-year	10.00	<**	10.82	>*	9.54		9.88
10-year	9.49	<**	10.22		9.34		9.55
bias in $\hat{\sigma}$							
6-month	-7.25**		-1.19**		-8.27**		-2.17**
1-year	-2.70**		2.27**		-6.09**		-0.52
2-year	2.15**		5.69**		-3.46**		1.11*
3-year	3.88**		6.71**		-1.77**		1.95**
4-year	4.48**		6.97**		-0.73		2.63**
5-year	4.50**		6.85**		-0.24		3.06**
7-year	3.92**		6.23**		-0.13		3.31**
10-year	2.91**		5.48**		-0.57		3.10**
RMSE of $\hat{\sigma}$							
6-month	7.80	>**	3.30	<**	8.76	>**	3.82
1-year	4.72		4.64	<**	7.23	>**	4.03
2-year	6.01	<**	8.04	>*	6.62	>*	5.76
3-year	7.55	<**	9.35	>**	6.76		6.78
4-year	8.21	<**	9.81	>**	6.98		7.36
5-year	8.37	<**	9.84	>**	7.12		7.67
7-year	8.02	<**	9.37	>**	7.06		7.71
10-year	7.19	<**	8.57	>**	6.67		7.24

Table 12: Short rate volatility forecast regressions

This table contains coefficients and standard errors from regressions of two volatility proxies computed from six-month yield changes on different forecasting variables. Newey-West standard errors are calculated using 6 lags for the top panel and 11 lags for the bottom. Realized volatility ($\hat{\sigma}$) is defined by $\hat{\sigma}_{t,\tau}^2 = \sum_{i=1}^5 \Delta Y(t, i, \tau)^2$ and is calculated using daily data. GARCH volatilities are the same fitted values used in Table 7, while the $A_1(3)$ and $A_1(4)$ USV forecasts are the same as those used in Table 10. Regressions are estimated using data from 1990 to 2002, allowing for a sample size of 677 weeks.

Specification Number	Intercept*	GARCH volatility	$A_1(3)$ forecast	$A_1(4)$ USV forecast	1st PC*	2nd PC*	3rd PC*	Adjusted R-Squared
dependent variable: $ \Delta Y $								
1	-0.084 (0.026)				0.740 (0.496)	-0.479 (0.077)	1.686 (0.589)	0.090
2	0.000 (0.000)	0.623 (0.090)						0.086
3	-0.072 (0.022)	0.403 (0.086)			0.322 (0.409)	-0.299 (0.076)	1.353 (0.503)	0.115
4	0.006 (0.001)		-5.567 (0.982)					0.077
5	8.913 (4.640)		-69.606 (35.921)		-20.227 (10.822)	3.973 (2.304)	-5.751 (3.852)	0.093
6	0.000 (0.000)			0.835 (0.093)				0.115
7	-0.032 (0.022)			0.710 (0.116)	-0.495 (0.485)	-0.099 (0.090)	1.249 (0.493)	0.122
dependent variable: $\hat{\sigma}$								
1'	-0.053 (0.022)				0.799 (0.504)	-0.417 (0.074)	1.233 (0.539)	0.153
2'	0.000 (0.000)	0.676 (0.076)						0.240
3'	-0.037 (0.012)	0.545 (0.083)			0.233 (0.323)	-0.174 (0.062)	0.783 (0.306)	0.263
4'	0.006 (0.001)		-3.991 (0.714)					0.132
5'	4.992 (4.205)		-32.729 (27.284)		-10.958 (9.820)	2.079 (2.097)	-2.937 (3.509)	0.154
6'	0.000 (0.000)			0.655 (0.067)				0.237
7'	0.003 (0.017)			0.641 (0.098)	-0.533 (0.468)	-0.008 (0.094)	0.763 (0.398)	0.243

* denotes a coefficient that has been multiplied by 100

Table 13: The relation between short rate volatility and the yield curve

This table contains the results of regressing the short rate variance on the first three principal components. The first set of coefficients are indirect estimates implied by the estimated parameters of the $A_1(4)$ USV model. These are computed from the covariance matrix of the state vector using the fact that the principal components are linear in that vector. Point estimates are implied by the parameter values in Table 3, while numbers in parentheses denote posterior standard deviations. The second set of regression coefficients come regressing the GARCH(1,1) variance computed from six-month yields on the three principal components over the 1988-2002 sample. The values in parentheses are Newey-West standard errors calculated using 20 lags.

	Intercept	1st PC*	2nd PC*	3rd PC*	R-Squared
$A_1(4)$ USV model-implied population regression coefficients					
Point Estimate	-0.125	2.738	-0.789	0.563	0.397
Posterior Std. Dev.	(0.065)	(1.509)	(0.259)	(1.417)	(0.190)
regression coefficients from GARCH and sample principal components					
OLS Estimate	-0.137	0.933	-0.277	2.106	0.262
Newey-West Std. Err.	(0.039)	(0.593)	(0.135)	(0.811)	

* denotes a coefficient that has been multiplied by 1000

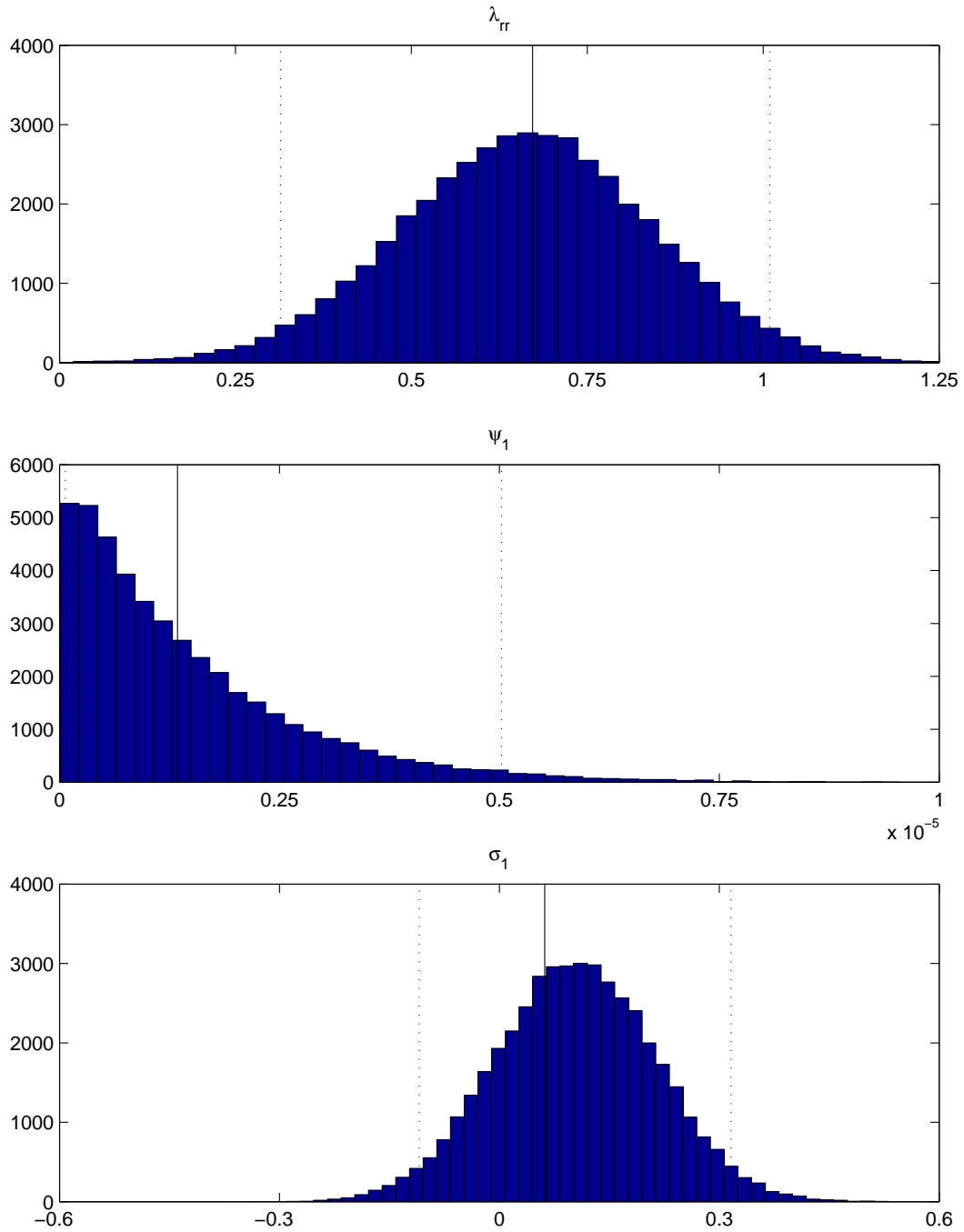
Table B1: Yield fits for A₁(3)USV with and without high frequency data augmentation

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 $\hat{\epsilon}_t = Y_t - \hat{Y}_t$, where \hat{Y}_t denotes the model fitted value. The sample size is 782 weeks for in-sample statistics and 104 weeks for out-of-sample statistics.

	----- In sample -----		----- Out of sample -----	
	Without HFDA	With HFDA	Without HFDA	With HFDA
	mean $\hat{\epsilon}$ (basis points)			
6-month	-5.24	-5.41	-2.74	-2.94
1-year	-1.04	-1.10	-2.13	-2.20
2-year	3.28	3.33	1.29	1.38
3-year	4.20	4.31	2.74	2.90
4-year	3.27	3.37	2.81	2.96
5-year	1.53	1.59	2.11	2.21
7-year	-2.22	-2.28	-0.51	-0.56
10-year	-5.62	-5.87	-6.27	-6.59
	RMSE (basis points)			
6-month	14.61	14.69	12.48	12.57
1-year	6.58	6.43	4.25	4.07
2-year	10.36	10.21	8.87	8.64
3-year	10.05	9.93	9.00	8.77
4-year	7.95	7.81	7.20	6.95
5-year	5.50	5.32	4.99	4.68
7-year	5.55	5.45	4.58	4.44
10-year	14.11	14.17	11.39	11.58
	autocorrelation of $\hat{\epsilon}$			
6-month	0.96	0.96	0.96	0.97
1-year	0.72	0.76	0.21	0.23
2-year	0.80	0.83	0.71	0.74
3-year	0.79	0.82	0.68	0.71
4-year	0.71	0.75	0.52	0.55
5-year	0.53	0.60	0.17	0.17
7-year	0.57	0.66	0.55	0.62
10-year	0.92	0.94	0.92	0.93

Figure 1

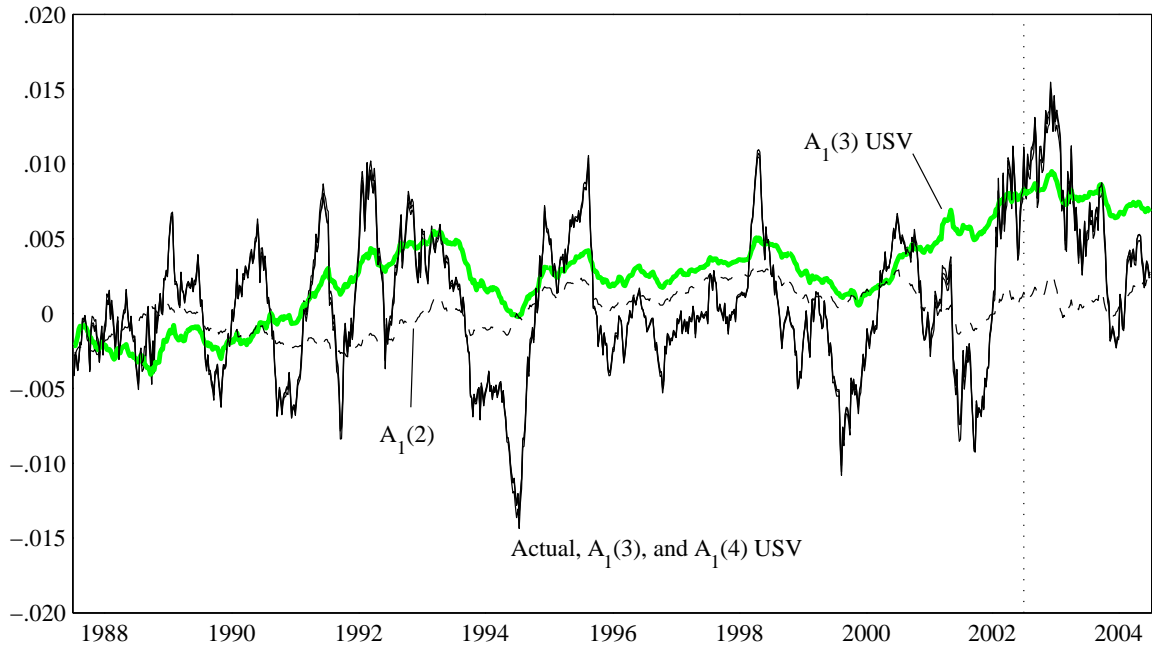
Posterior histograms of the $A_1(3)$ USV model



This figure depicts histograms of the MCMC draws from the posterior distribution of several parameters of the $A_1(3)$ USV model. Solid vertical lines denote the multivariate posterior median reported in Table 3, while dashed lines denote 2.5 and 97.5 percentiles of the posterior.

Figure 2

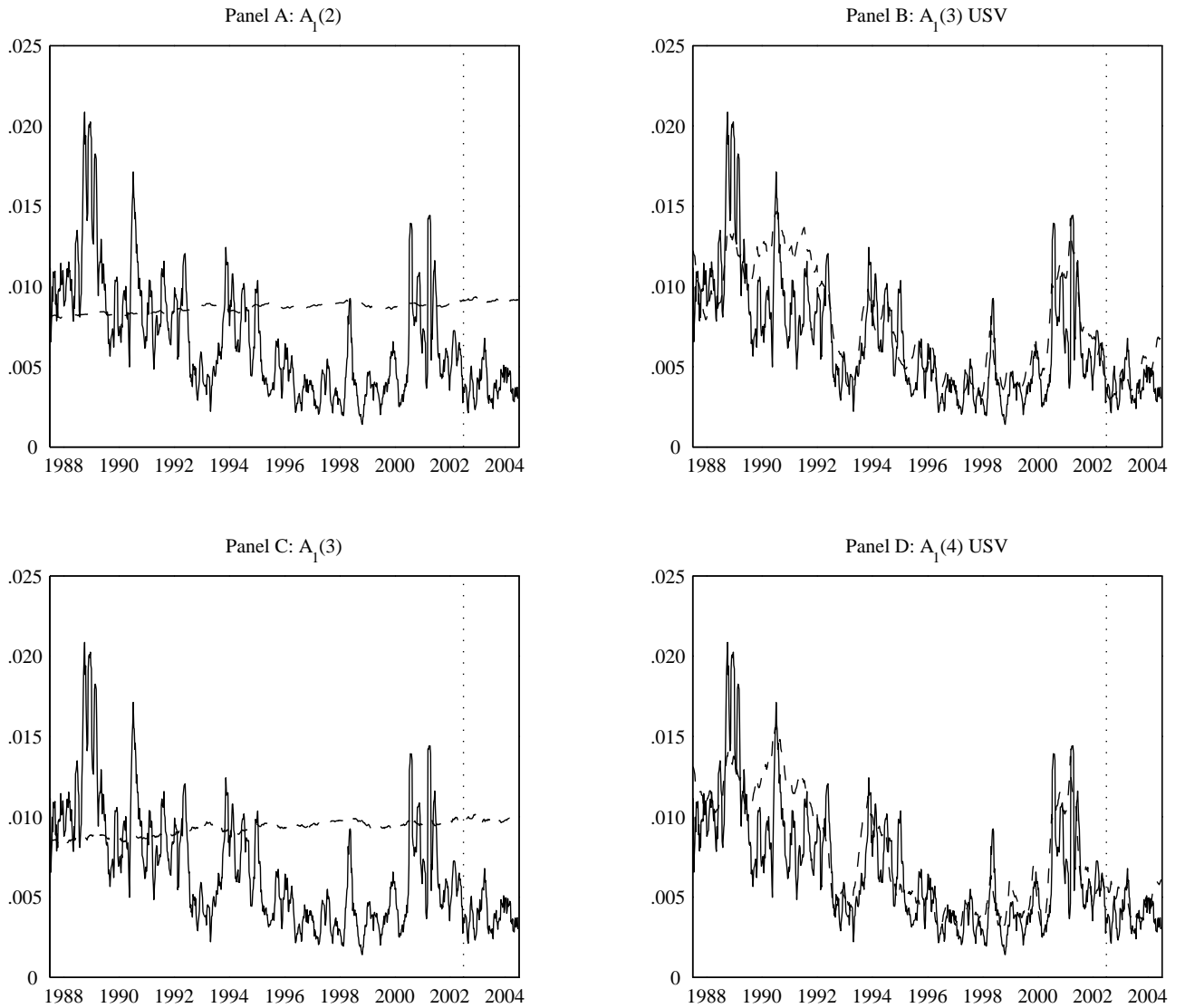
Actual and model-implied curvature



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 3

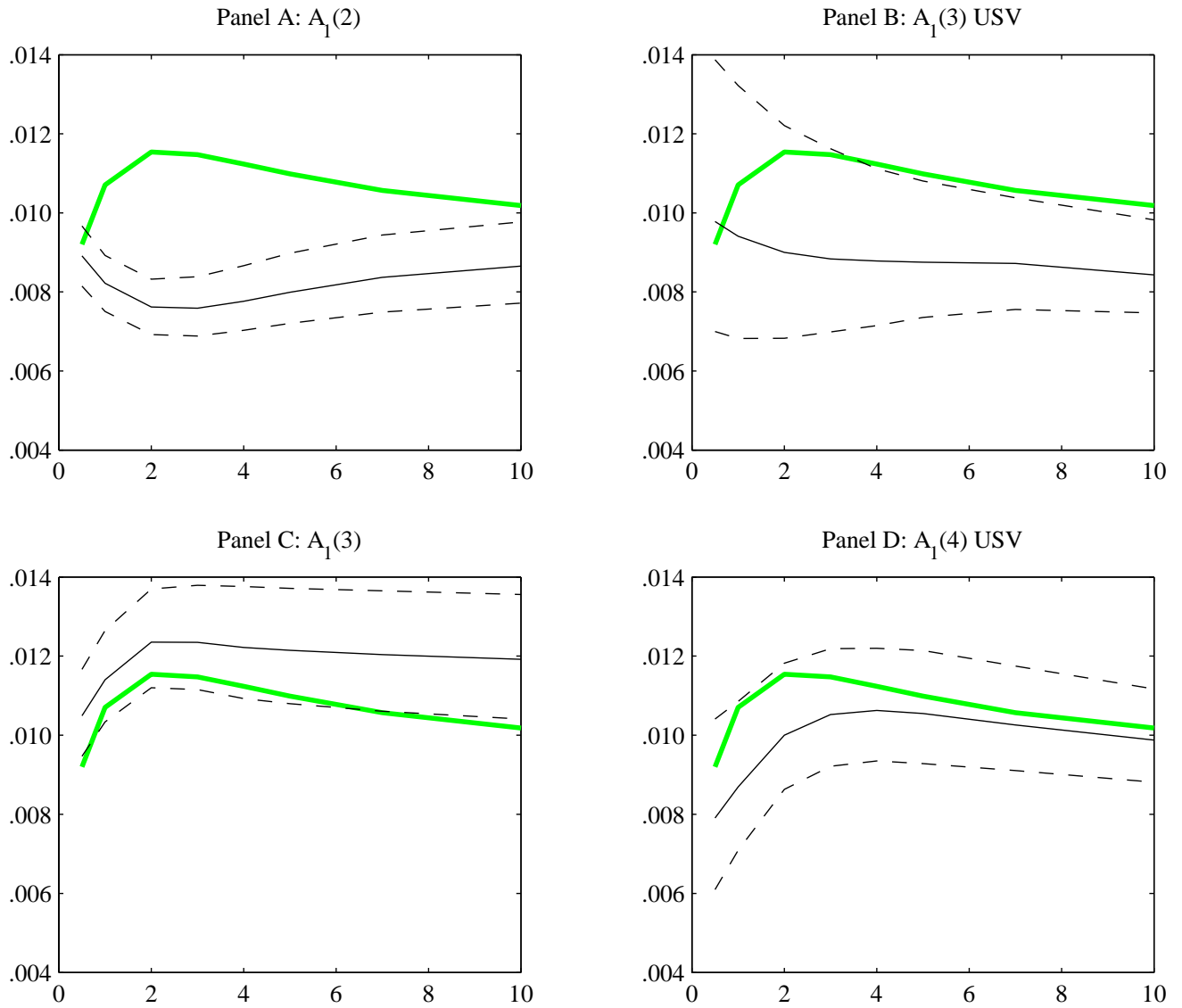
Rolling window and model-implied short rate volatility



In each panel, the solid line depicts the fitted path of the volatility of the 6-month yield that constructed from rolling 30-day windows. The dashed lines correspond to smoothed estimates of instantaneous volatility implied by each of the affine specifications considered. The vertical dotted lines denote the end of the estimation period.

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 3. The means and 95% confidence intervals of these distributions are depicted by solid and dashed black lines, respectively.

Figure B1

Posterior distributions for $A_1(3)USV$ with and without high frequency data augmentation

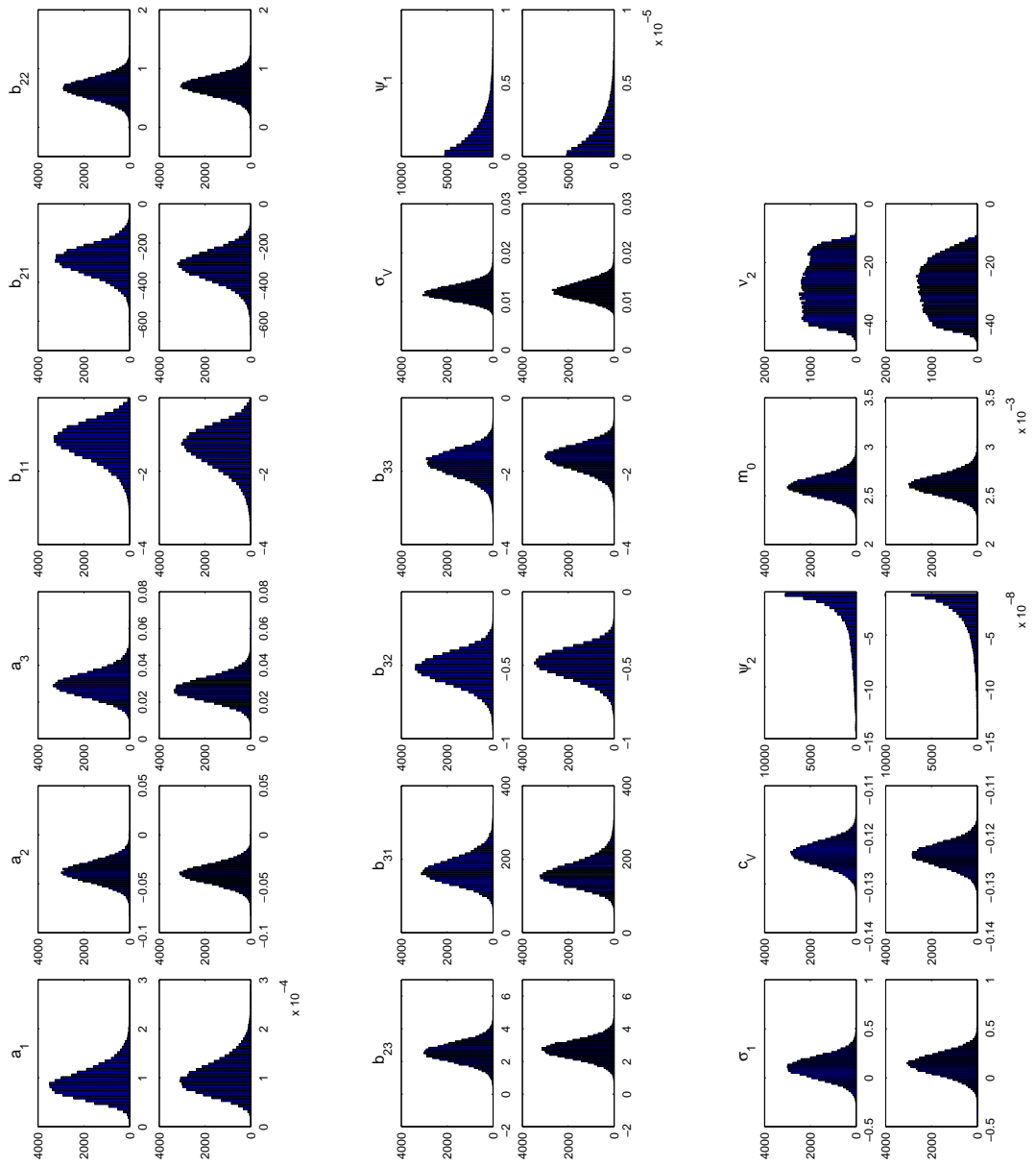


Figure B2

Smoothed state variables for $A_1(3)$ USV with and without high frequency data augmentation

