



American Finance Association

Specification Analysis of Affine Term Structure Models

Author(s): Qiang Dai and Kenneth J. Singleton

Source: *The Journal of Finance*, Vol. 55, No. 5 (Oct., 2000), pp. 1943-1978

Published by: Blackwell Publishing for the American Finance Association

Stable URL: <http://www.jstor.org/stable/222481>

Accessed: 24/11/2009 12:14

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=black>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Blackwell Publishing and American Finance Association are collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Finance*.

<http://www.jstor.org>

Specification Analysis of Affine Term Structure Models

QIANG DAI and KENNETH J. SINGLETON*

ABSTRACT

This paper explores the structural differences and relative goodness-of-fits of affine term structure models (ATSMs). Within the family of ATSMs there is a trade-off between flexibility in modeling the conditional correlations and volatilities of the risk factors. This trade-off is formalized by our classification of N -factor affine family into $N + 1$ non-nested subfamilies of models. Specializing to three-factor ATSMs, our analysis suggests, based on theoretical considerations and empirical evidence, that some subfamilies of ATSMs are better suited than others to explaining historical interest rate behavior.

IN SPECIFYING A DYNAMIC TERM STRUCTURE MODEL—one that describes the co-movement over time of short- and long-term bond yields—researchers are inevitably confronted with trade-offs between the richness of econometric representations of the state variables and the computational burdens of pricing and estimation. It is perhaps not surprising then that virtually all of the empirical implementations of multifactor term structure models that use time series data on long- and short-term bond yields simultaneously have focused on special cases of “affine” term structure models (ATSMs). An ATSM accommodates time-varying means and volatilities of the state variables through affine specifications of the risk-neutral drift and volatility coefficients. At the same time, ATSMs yield essentially closed-form expressions for zero-coupon-bond prices (Duffie and Kan (1996)), which greatly facilitates pricing and econometric implementation.

The focus on ATSMs extends back at least to the pathbreaking studies by Vasicek (1977) and Cox, Ingersoll, and Ross (1985), who presumed that the instantaneous short rate $r(t)$ was an affine function of an N -dimensional state vector $Y(t)$, $r(t) = \delta_0 + \delta'_Y Y(t)$, and that $Y(t)$ followed Gaussian and square-root diffusions, respectively. More recently, researchers have explored formulations of ATSMs that extend the one-factor Markov represen-

* Dai is from New York University. Singleton is from Stanford University. We thank Darrell Duffie for extensive discussions, Ron Gallant and Jun Liu for helpful comments, the editor and the referees for valuable inputs, and the Financial Research Initiative of the Graduate School of Business at Stanford University for financial support. We are also grateful for comments from seminar participants at the NBER Asset Pricing Group meeting, the Western Finance Association Annual Meetings, Duke University, London Business School, New York University, Northwestern University, University of Chicago, University of Michigan, University of Washington at St. Louis, Columbia University, Carnegie Mellon University, and CIRANO.

tation of the short-rate, $dr(t) = (\theta - r(t))dt + \sqrt{v}dB(t)$, by introducing a stochastic long-run mean $\theta(t)$ and a volatility $v(t)$ of $r(t)$ that are affine functions of $(r(t), \theta(t), v(t))$ (e.g., Chen (1996), Balduzzi et al. (1966)).¹ These and related ATSMs underpin extensive literatures on the pricing of bonds and interest-rate derivatives and also underlie many of the pricing systems used by the financial industry. Yet, in spite of their central importance in the term structure literature, the structural differences and relative empirical goodness-of-fits of ATSMs remain largely unexplored.

This paper characterizes, both formally and intuitively, the differences and similarities among affine specifications of term structure models and assesses their strengths and weaknesses as empirical models of interest rate behavior. We begin our specification analysis by developing a comprehensive classification of ATSMs with the following convenient features: (i) whether a specification of an affine model leads to well-defined bond prices—a property that we will refer to as *admissibility* (Section I)—is easily verified; (ii) all admissible N -factor ATSMs are uniquely classified into $N + 1$ non-nested subfamilies; and (iii) for each of the $N + 1$ subfamilies, there exists a *maximal* model that nests econometrically all other models within this subfamily. With this classification scheme in place, we answer the following questions.

[Q1] Given an ATSM (e.g., one of the popular specifications in the literature), is it “maximally flexible,” and, if not, what are the overidentifying restrictions that it imposes on yield curve dynamics?

[Q2] Are extant models, or their maximal counterparts, sufficiently flexible to describe simultaneously the historical movements in short- and long-term bond yields?

Ideally, a specification analysis of ATSMs could begin with the specification of an all-encompassing ATSM, and then all other ATSMs could be studied as special cases. However, for a specification to be admissible, constraints must be imposed on the dynamic interactions among the state variables, and these constraints turn out to preclude the existence of such an all-encompassing model. Therefore, as a preliminary step in our specification analysis, we characterize the family of admissible ATSMs.² This is accomplished by classifying ATSMs into $N + 1$ subfamilies according to the number “ m ” of the Y s (more precisely, the number of independent linear combinations of Y s) that determine the conditional variance matrix of Y , and then using this classification to provide a formal and intuitive characterization of (minimal known) sufficient conditions for admissibility. Each of the $N + 1$ subfamilies of admissible models is shown to have a *maximal* element, a feature we exploit in answering Q1 and Q2.

¹Nonlinear models, such as those studied by Chan et al. (1992), can be extended in similar fashion to multi-factor models, such as those of Andersen and Lund (1998), that fall outside the affine family.

²The problem of admissibility was not previously an issue in empirical implementations of ATSMs, because the special structure of Gaussian and CIR-style models made verification that bond prices are well defined relatively straightforward.

The usefulness of this classification scheme is illustrated by specializing to the case of $N = 3$ and describing in detail the nature of the four maximal models for the three-factor family of ATSMs. This discussion highlights an important trade-off within the family of ATSMs between the dependence of the conditional variance of each $Y_i(t)$ on $Y(t)$ and the admissible structure of the correlation matrix for Y . Gaussian models offer complete flexibility with regard to the signs and magnitudes of conditional and unconditional correlations among the Y s but at the “cost” of the apparently counterfactual assumption of constant conditional variances ($m = 0$). At the other end of the spectrum of volatility specifications lies (what we refer to as) the correlated square-root diffusion (CSR) model that has all three state variables driving conditional volatilities ($m = 3$). However, admissibility of models in this subfamily requires that the conditional correlations of the state variables be zero and that their unconditional correlations be non-negative. In between the Gaussian and CSR models lie two subfamilies of ATSMs with time-varying conditional volatilities of the state variables and unconstrained signs of (some of) their correlations.

Specializing further, we show that the Vasicek (Gaussian), BDFS, Chen, and CIR models are classified into distinct subfamilies. Moreover, comparing these models to the maximal models in their respective subfamilies, we find that, in every case except the Gaussian models, these models impose potentially strong overidentifying restrictions relative to the maximal model. Thus, we answer *Q1* by showing that there exist identified, admissible ATSMs that allow much richer interdependencies among the factors than have heretofore been studied.

One notable illustration of this point is our finding that the standard assumption of independent risk factors in CIR-style models (see, e.g., Chen and Scott (1993), Pearson and Sun (1994), and Duffie and Singleton (1997)) is not necessary either for admissibility of these models or for zero-coupon bond prices to be known (essentially) in closed form.³ At the same time we show that, when the correlations in these CSR models are nonzero, they must be positive for the model to be admissible. The data on U.S. interest rates seems to call for *negative* correlations among the risk factors (see Section II.C). Because CSR models are theoretically incapable of generating negative correlations, we conclude that they are not consistent with the historical behavior of U.S. interest rates.

Given the absence of time-varying volatility in Gaussian ($m = 0$) models and the impossibility of negatively correlated risk factors in CSR ($m = 3$) models, in answering *Q2*, we focus on the two subfamilies of $N = 3$ models in which the stochastic volatilities of the Y s are controlled by one ($m = 1$) and two ($m = 2$) state variables. The maximal model for the subfamily $m = 1$ nests the BDFS model, whereas the maximal model for the $m = 2$ subfamily nests the Chen model.

We compute simulated method of moments (*SMM*) estimates (Duffie and Singleton (1993), Gallant and Tauchen (1996)) of our maximal ATSMs. Whereas most of the empirical studies of term structure models have focused on U.S.

³ We are grateful to Jun Liu for pointing out this implication of our admissibility conditions.

Treasury yield data, we follow Duffie and Singleton (1997) and study LIBOR-based yields from the ordinary, fixed-for-variable rate swap market. The primary motivation for choosing swap instead of Treasury yields is that the former are relatively unencumbered by institutional factors that are not fully accounted for in standard arbitrage-free term structure models, including the ATSMs studied in this paper (see Section II.A). The models pass several formal goodness-of-fit tests. Moreover, the restrictions implicit in the Chen and BDFS models are strongly rejected. The substantial improvements in goodness-of-fit for the newly introduced, maximal models are traced directly to their more flexibly parameterized correlations among the state variables. *Negatively* correlated diffusions are central to the models' abilities to match the volatility structure and non-normality of changes in bond yields.

The remainder of the paper is organized as follows. Section I presents general results pertaining to the classification, admissibility, and identification of the family of N -factor affine term structure models. Section I.B specializes the classification results to the family of three-factor affine term structure models and characterizes explicitly the nature of the overidentifying restrictions in extant models relative to our more flexible, maximal models. Section II.A explains our estimation strategy and data. Section II.B discusses the econometric identification of risk premiums in affine models. Section II.C presents our empirical results. Finally, Section III concludes.

I. A Characterization of Admissible ATSMs

Absent arbitrage opportunities, the time- t price of a zero-coupon bond that matures at time T , $P(t, \tau)$, is given by

$$P(t, \tau) = E_t^Q \left[e^{-\int_t^\tau r_s ds} \right], \quad (1)$$

where E^Q denotes expectation under the risk-neutral measure. An N -factor affine term structure model is obtained under the assumptions that the instantaneous short rate $r(t)$ is an affine function of a vector of unobserved state variables $Y(t) = (Y_1(t), Y_2(t), \dots, Y_N(t))$,

$$r(t) = \delta_0 + \sum_{i=1}^N \delta_i Y_i(t) \equiv \delta_0 + \delta_y' Y(t), \quad (2)$$

and that $Y(t)$ follows an "affine diffusion,"

$$dY(t) = \tilde{K}(\tilde{\theta} - Y(t))dt + \Sigma\sqrt{S(t)}d\tilde{W}(t). \quad (3)$$

$\tilde{W}(t)$ is an N -dimensional independent standard Brownian motion under Q , $\tilde{K}$ and Σ are $N \times N$ matrices, which may be nondiagonal and asymmetric, and $S(t)$ is a diagonal matrix with the i th diagonal element given by

$$[S(t)]_{ii} = \alpha_i + \beta_i' Y(t). \quad (4)$$

Both the drifts in equation (3) and the conditional variances in equation (4) of the state variables are affine in $Y(t)$.

Provided a parameterization is admissible, we know from Duffie and Kan (1996) that

$$P(t, \tau) = e^{A(\tau) - B(\tau)' Y(t)}, \quad (5)$$

where $A(\tau)$ and $B(\tau)$ satisfy the ordinary differential equations (ODEs)

$$\frac{dA(\tau)}{d\tau} = -\tilde{\theta}' \tilde{\mathcal{K}}' B(\tau) + \frac{1}{2} \sum_{i=1}^N [\Sigma' B(\tau)]_i^2 \alpha_i - \delta_0, \quad (6)$$

$$\frac{dB(\tau)}{d\tau} = -\tilde{\mathcal{K}}' B(\tau) - \frac{1}{2} \sum_{i=1}^N [\Sigma' B(\tau)]_i^2 \beta_i + \delta_y. \quad (7)$$

These ODEs, which can be solved easily through numerical integration starting from the initial conditions $A(0) = 0$ and $B(0) = 0_{N \times 1}$, are completely determined by the specification of the risk-neutral dynamics of $r(t)$, in equations (2) through (4).

To use the (essentially) closed-form expression of equation (5) in empirical studies of ATSMs we also need to know the distributions of $Y(t)$ and $P(t, \tau)$ under the actual probability measure $\mathcal{P}$. We assume that the market prices of risk, $\Lambda(t)$, are given by

$$\Lambda(t) = \sqrt{S(t)} \lambda, \quad (8)$$

where λ is an $N \times 1$ vector of constants. Under this assumption, the process for $Y(t)$ under $\mathcal{P}$ also has the affine form,⁴

$$dY(t) = \mathcal{K}(\Theta - Y(t)) dt + \Sigma \sqrt{S(t)} dW(t), \quad (9)$$

where $W(t)$ is an N -dimensional vector of independent standard Brownian motions under $\mathcal{P}$, $\mathcal{K} = \tilde{\mathcal{K}} - \Sigma \Phi$, $\Theta = \mathcal{K}^{-1}(\tilde{\mathcal{K}} \tilde{\theta} + \Sigma \psi)$, the i th row of Φ is given by $\lambda_i \beta'_i$, and ψ is an N -vector whose i th element is given by $\lambda_i \alpha_i$.

⁴ Our formulation can be easily generalized to a nonaffine diffusion for Y under $\mathcal{P}$, while preserving the tractable pricing relations in equation (5), as long as the specification of $\Lambda(t)$ preserves the affine structure of Y under $\mathcal{Q}$. See Duffie (1998) for a generalization of our framework along this line. We do not pursue this issue here, partly because our primary focus is on the correlation structure of $Y(t)$ and partly because we find it difficult to accurately estimate the market prices of risk, even for our simple parameterization of $\Lambda(t)$ (see Section II.B).

A. A Canonical Representation of Admissible ATSMs

The general specification in equation (9) does not lend itself to a specification analysis because, for an arbitrary choice of the parameter vector $\psi \equiv (\mathcal{K}, \Theta, \Sigma, \mathcal{B}, \alpha)$, where $\mathcal{B} \equiv (\beta_1, \dots, \beta_N)$ denotes the matrix of coefficients on Y in the $[S(t)]_{ii}$, the conditional variances $[S(t)]_{ii}$ may not be positive over the range of Y . We will refer to a specification of ψ as admissible if the resulting $[S(t)]_{ii}$ are strictly positive, for all i . There is no admissibility problem if $\beta_i = 0$. However, outside of this special case, to assure admissibility we find it necessary to constrain the drift parameters ($\mathcal{K}$ and Θ) and diffusion coefficients (Σ and $\mathcal{B}$). Moreover, our requirements for admissibility become increasingly stringent as the number of state variables determining $[S(t)]_{ii}$ increases.

To formalize what we mean by the family of *admissible* N -factor ATSMs, let $m = \text{rank}(\mathcal{B})$ index the degree of dependence of the conditional variances on the number of state variables. Using this index, we classify each ATSM uniquely into one of $N + 1$ subfamilies based on its value of m . Not all N -factor ATSMs with index value m are admissible, so, for each m , we let $\mathbb{A}_m(N)$ denote those that are admissible. Next, we define the *canonical* representation of $\mathbb{A}_m(N)$ as follows.

Definition 1 [Canonical Representation of $\mathbb{A}_m(N)$]: For each m , we partition $Y(t)$ as $Y' = (Y^B, Y^D)$, where Y^B is $m \times 1$ and Y^D is $(N - m) \times 1$, and define the canonical representation of $\mathbb{A}_m(N)$ as the special case of equation (9) with

$$\mathcal{K} = \begin{bmatrix} \mathcal{K}_{m \times m}^{BB} & \mathbf{0}_{m \times (N-m)} \\ \mathcal{K}_{(N-m) \times m}^{DB} & \mathcal{K}_{(N-m) \times (N-m)}^{DD} \end{bmatrix}, \quad (10)$$

for $m > 0$, and $\mathcal{K}$ is either upper or lower triangular for $m = 0$,

$$\Theta = \begin{pmatrix} \Theta_{m \times 1}^B \\ \mathbf{0}_{(N-m) \times 1} \end{pmatrix}, \quad (11)$$

$$\Sigma = I, \quad (12)$$

$$\alpha = \begin{pmatrix} \mathbf{0}_{m \times 1} \\ \mathbf{1}_{(N-m) \times 1} \end{pmatrix}, \quad (13)$$

$$\mathcal{B} = \begin{bmatrix} I_{m \times m} & B_{m \times (N-m)}^{BD} \\ \mathbf{0}_{(N-m) \times m} & \mathbf{0}_{(N-m) \times (N-m)} \end{bmatrix}; \quad (14)$$

with the following parametric restrictions imposed:

$$\delta_i \geq 0, \quad m + 1 \leq i \leq N, \quad (15)$$

$$\mathcal{K}_i \Theta \equiv \sum_{j=1}^m \mathcal{K}_{ij} \Theta_j > 0, \quad 1 \leq i \leq m, \quad (16)$$

$$\mathcal{K}_{ij} \leq 0, \quad 1 \leq j \leq m, \quad j \neq i, \quad (17)$$

$$\Theta_i \geq 0, \quad 1 \leq i \leq m, \quad (18)$$

$$\mathcal{B}_{ij} \geq 0, \quad 1 \leq i \leq m, \quad m + 1 \leq j \leq N. \quad (19)$$

Finally, $\mathbb{A}_m(N)$ is formally defined as the set of all ATSMs that are nested special cases of our canonical model or of any equivalent model obtained by an *invariant* transformation of the canonical model. Invariant transformations, which are formally defined in Appendix A, preserve admissibility and identification and leave the short rate (and hence bond prices) unchanged.

The assumed structure of $\mathcal{B}$ assures that $\text{rank}(\mathcal{B}) = m$ for the m th canonical representation. To verify that it resides in $\mathbb{A}_m(N)$, note that the instantaneous conditional correlations among the $Y^B(t)$ are zero, whereas the instantaneous correlations among the $Y^D(t)$ are governed by the parameters $\mathcal{B}_{ij}$, because $\Sigma = I$.⁵ Because the conditional covariance matrix of Y depends only on Y^B and equation (19) holds, admissibility is established if $Y^B(t)$ is strictly positive. The positivity of Y^B is assured by zero restrictions in the upper right $m \times (N - m)$ block of $\mathcal{K}$ and the constraints in equations (17) and (18). In Appendix A we show formally that the zero restrictions in equations (10) through (14) and sign restrictions in equations (15) through (19) are sufficient conditions for admissibility. To assure that the state process is stationary, we need to impose the additional constraint that all of the eigenvalues of $\mathcal{K}$ are strictly positive. We impose this constraint in our empirical estimation.

Not only is the canonical representation admissible, but it is also “maximal” in $\mathbb{A}_m(N)$ in the sense that, given m , we have imposed the minimal known sufficient condition for admissibility and then imposed minimal normalizations for econometric identification (see Appendix A).⁶ However, our canonical representation is not the unique maximal model in $\mathbb{A}_m(N)$. Rather, there is an *equivalence class* $AM_m(N)$ of maximal models obtained by invariant transformations of the canonical representation. The representation

⁵ All of the state variables may be mutually correlated over any finite sampling interval due to feedback through the drift matrix $\mathcal{K}$.

⁶ Because the conditions for admissibility are sufficient, but are not known to be necessary, we cannot rule out the possibility that there are admissible, econometrically identified ATSMs that nest our canonical models as special cases. Importantly, all of the extant ATSMs in the literature reside within $\mathbb{A}_m(N)$, for some m and N .

in equations (10) through (19) was chosen as our canonical representation among equivalent maximal models in $AM_m(N)$, because of the relative ease with which admissibility and identification can be verified and the parametric restrictions in equations (15) through (19) can be imposed in econometric implementations.

As will be illustrated subsequently, the canonical representation of $\mathbb{A}_m(N)$ is often not as convenient as other members of $AM_m(N)$ for interpreting the state variables within a particular ATSM. In particular, the literature has often chosen to parameterize ATSMs with the riskless rate r being one of the state variables. Any such Ar ("affine in r ") representation is typically in $\mathbb{A}_m(N)$, for some m and N , and therefore has an equivalent representation in which $r(t) = \delta_0 + \delta'_y Y(t)$, with $Y(t)$ treated as an unobserved state vector (an AY or "affine in Y " representation). Similarly, any AY representation (e.g., a CIR model) typically has an equivalent Ar representation. In Section I.B we present the equivalent Ar and AY representations of several extant ATSMs and also their maximally flexible counterparts.

An implication of our classification scheme is that an exhaustive specification analysis of the family of admissible N -factor ATSMs requires the examination of $N + 1$ non-nested, maximal models.

B. Three-Factor ATSMs

In this section we explore in considerably more depth the implications of our classification scheme for the specification of ATSMs. Particular attention is given to interpreting the term structure dynamics associated with our canonical models and to the nature of the overidentifying restrictions imposed in several ATSMs in the literature. To better link up with the empirical term structure literature, we fix $N = 3$ and examine the four associated subfamilies of admissible ATSMs.

B.1. $\mathbb{A}_0(3)$

If $m = 0$, then none of the Y s affect the volatility of $Y(t)$, so the state variables are homoskedastic and $Y(t)$ follows a three-dimensional Gaussian diffusion. The elements of ψ for the canonical representation of $AM_0(3)$ are given by

$$\begin{aligned} \mathcal{K} &= \begin{bmatrix} \kappa_{11} & 0 & 0 \\ \kappa_{21} & \kappa_{22} & 0 \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix}, & \Sigma &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \\ \Theta &= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, & \alpha &= \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, & \mathcal{B} &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \end{aligned}$$

where $\kappa_{11} > 0$, $\kappa_{22} > 0$, and $\kappa_{33} > 0$.

Gaussian ATSMs were studied theoretically by Vasicek (1977) and Langer (1980), among many others. A recent empirical implementation of a two-factor Gaussian model is found in Jegadeesh and Pennacchi (1996).

B.2. $\mathbb{A}_1(3)$

The family $\mathbb{A}_1(3)$ is characterized by the assumption that one of the Y s determines the conditional volatility of all three state variables. One member of $\mathbb{A}_1(3)$ is the BDFS model:

$$\begin{aligned} du(t) &= \mu(\bar{u} - u(t))dt + \hat{\eta}\sqrt{u(t)}dB_v(t), \\ d\theta(t) &= \nu(\bar{\theta} - \theta(t))dt + \zeta dB_\theta(t), \\ dr(t) &= \kappa(\theta(t) - r(t))dt + \sqrt{u(t)}d\hat{B}_r(t), \end{aligned} \quad (20)$$

with the only nonzero diffusion correlation being $\text{cov}(dB_v(t), d\hat{B}_r(t)) = \rho_{rv}dt$. Rewriting the short rate equation in equation (20) as

$$dr(t) = \kappa(\theta(t) - r(t))dt + \sqrt{1 - \rho_{rv}^2}\sqrt{u(t)}dB_r(t) + \rho_{rv}\sqrt{u(t)}dB_v(t), \quad (21)$$

where $B_r(t)$ and $B_v(t)$ are independent, and replacing $u(t)$ by $v(t) = (1 - \rho_{rv}^2)u(t)$, $\bar{u}$ by $\bar{v} = (1 - \rho_{rv}^2)\bar{u}$, and $\hat{\eta}$ by $\eta = \sqrt{1 - \rho_{rv}^2}\hat{\eta}$, we obtain the BDFS model expressed in our notational convention for ATSMs,

$$\begin{aligned} dv(t) &= \mu(\bar{v} - v(t))dt + \eta\sqrt{v(t)}dB_v(t), \\ d\theta(t) &= \nu(\bar{\theta} - \theta(t))dt + \zeta dB_\theta(t), \\ dr(t) &= \kappa(\theta(t) - r(t))dt + \sqrt{v(t)}dB_r(t) + \sigma_{rv}\eta\sqrt{v(t)}dB_v(t), \end{aligned} \quad (22)$$

where $\sigma_{rv} = \rho_{rv}/\eta\sqrt{1 - \rho_{rv}^2}$.

The first state variable, $v(t)$, is a volatility factor, because it affects the short rate process only through the conditional volatility of r . The second state variable, $\theta(t)$, is the “central tendency” of r . The short rate mean reverts to its central tendency $\theta(t)$ at rate κ .

For interpreting the restrictions in the BDFS and related models, it is convenient to work with the following maximal model in $AM_1(3)$, presented in its Ar representation:

$$\begin{aligned} dv(t) &= \mu(\bar{v} - v(t))dt + \eta\sqrt{v(t)}dB_v(t), \\ d\theta(t) &= \nu(\bar{\theta} - \theta(t))dt + \sqrt{\zeta^2 + \boxed{\beta_\theta}v(t)}dB_\theta(t) \\ &\quad + \boxed{\sigma_{\theta v}}\eta\sqrt{v(t)}dB_v(t) + \boxed{\sigma_{\theta r}}\sqrt{\alpha_r + v(t)}dB_r(t), \\ dr(t) &= \boxed{\kappa_{rv}}(\bar{v} - v(t))dt + \kappa(\theta(t) - r(t))dt + \sqrt{\boxed{\alpha_r} + v(t)}dB_r(t) \\ &\quad + \sigma_{rv}\eta\sqrt{v(t)}dB_v(t) + \boxed{\sigma_{r\theta}}\sqrt{\zeta^2 + \beta_\theta v(t)}dB_\theta(t). \end{aligned} \quad (23)$$

The BDFS model is the special case of the expressions in equation (23) in which the parameters in square boxes are set to zero. Thus, relative to this maximal model, the BDFS model constrains the conditional correlation between r and θ to zero ($\sigma_{r\theta} = \sigma_{\theta r} = 0$). Additionally, it precludes the volatility shock v from affecting the volatility of the central tendency factor θ ($\sigma_{\theta v} = 0$). Finally, the BDFS model constrains $\kappa_{rv} = 0$ so that v cannot affect the drift of r . Freeing up these restrictions gives us a more flexible ATSM in which v is still naturally interpreted as the volatility shock but in which θ is perhaps not as naturally interpreted as the central tendency of r . The overidentifying restrictions imposed in the BDFS model are examined empirically in Section II.C.

Though the model specified by the expressions in equation (23) is convenient for interpreting the popular $\mathbb{A}_1(3)$ models, verifying that the model is maximal and, indeed, that it is admissible, is not straightforward. To check admissibility, it is much more convenient to work directly with the following equivalent AY representation (see Appendix E, Section E.1):

$$r(t) = \delta_0 + \boxed{\delta_1} Y_1(t) + Y_2(t) + Y_3(t), \quad (24)$$

$$\begin{aligned} d \begin{pmatrix} Y_1(t) \\ - \\ Y_2(t) \\ Y_3(t) \end{pmatrix} &= \left[\begin{array}{c|cc} \kappa_{11} & 0 & 0 \\ \hline 0 & \kappa_{22} & 0 \\ 0 & 0 & \kappa_{33} \end{array} \right] \left[\begin{pmatrix} \theta_1 \\ - \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} Y_1(t) \\ - \\ Y_2(t) \\ Y_3(t) \end{pmatrix} \right] dt \\ &\times \left[\begin{array}{c|cc} 1 & 0 & 0 \\ \hline \boxed{\sigma_{21}} & 1 & \boxed{\sigma_{23}} \\ \sigma_{31} & \boxed{\sigma_{32}} & 1 \end{array} \right] \sqrt{\begin{array}{c|cc} S_{11}(t) & 0 & 0 \\ \hline 0 & S_{22}(t) & 0 \\ 0 & 0 & S_{33}(t) \end{array}} dB(t), \end{aligned} \quad (25)$$

where

$$\begin{aligned} S_{11}(t) &= Y_1(t), \\ S_{22}(t) &= \alpha_2 + \boxed{[\beta_2]_1} Y_1(t), \\ S_{33}(t) &= \boxed{\alpha_3} + [\beta_3]_1 Y_1(t). \end{aligned} \quad (26)$$

That this representation is in $AM_1(3)$ follows from its equivalence to our canonical representation of $AM_1(3)$. This can be shown easily by diagonalizing Σ and by normalizing the scale of S_{22} and S_{33} , while freeing up δ_2 and δ_3 in the expression $r(t) = \delta_0 + \delta_1 Y_1(t) + \delta_2 Y_2(t) + \delta_3 Y_3(t)$. All three diffusions

may be conditionally correlated, and all three conditional variances may depend on Y_1 . However, admissibility requires that $\sigma_{12} = 0$ and $\sigma_{13} = 0$, in which case Y_1 follows a univariate square-root process that is strictly positive.

The equivalent AY representation of the BDFS model is obtained from equation (24) by setting all of the parameters in square boxes to zero, except for σ_{32} which is set to -1 . An immediate implication of this observation is that the BDFS model unnecessarily constrains the instantaneous short rate to be an affine function of only two of the three state variables ($\delta_1 = 0$). This is an implication of the assumption that the volatility factor $v(t)$ enters r only through its volatility and, therefore, it affects r only indirectly through its effects on the distribution of $(Y_2(t), Y_3(t))$. This constraint on δ_y is a feature of many of the extant models in the literature, including the model of Anderson and Lund (1998).

B.3. $\mathbb{A}_2(3)$

The family $\mathbb{A}_2(3)$ is characterized by the assumption that the volatilities of $Y(t)$ are determined by affine functions of two of the three Y 's. A member of this subfamily is the model proposed by Chen (1996):

$$\begin{aligned} dv(t) &= \mu(\bar{v} - v(t))dt + \eta\sqrt{v(t)}dW_1(t), \\ d\theta(t) &= \nu(\bar{\theta} - \theta(t))dt + \zeta\sqrt{\theta(t)}dW_2(t), \\ dr(t) &= \kappa(\theta(t) - r(t))dt + \sqrt{v(t)}dW_3(t), \end{aligned} \quad (27)$$

with the Brownian motions assumed to be mutually independent. As in the BDFS model, v and θ are interpreted as the stochastic volatility and central tendency, respectively, of r . A primary difference between the Chen and BDFS models, and the one that explains their classifications into different subfamilies, is that θ in the former follows a square-root diffusion, whereas it is Gaussian in the latter.

A convenient maximal model for interpreting the overidentifying restrictions in the Chen model is

$$\begin{aligned} dv(t) &= \mu(\bar{v} - v(t))dt + \boxed{\kappa_{v\theta}}(\bar{\theta} - \theta(t))dt + \eta\sqrt{v(t)}dW_1(t), \\ d\theta(t) &= \nu(\bar{\theta} - \theta(t))dt + \boxed{\kappa_{\theta v}}(\bar{v} - v(t))dt + \zeta\sqrt{\theta(t)}dW_2(t), \\ dr(t) &= \boxed{\kappa_{rv}}(\bar{v} - v(t))dt \\ &\quad + \boxed{\kappa_{r\theta}}(\bar{\theta} - \theta(t))dt - \kappa(\bar{\theta} - \theta(t))dt + \kappa(\boxed{\bar{r}} - r(t))dt \\ &\quad + \boxed{\sigma_{rv}}\eta\sqrt{v(t)}dW_1(t) + \boxed{\sigma_{r\theta}}\zeta\sqrt{\theta(t)}dW_2(t) \\ &\quad + \sqrt{\boxed{\alpha_r} + \boxed{\beta_\theta}\theta(t) + v(t)}dW_3(t). \end{aligned} \quad (28)$$

The Chen model is obtained as a special case with the parameters in square boxes set to zero except that $\bar{r} = \bar{\theta}$.

Clearly, within this maximal model, θ and v are no longer naturally interpreted as the central tendency and volatility factors for r . There may be feedback between θ and v through their drifts, and both of these variables may enter the drift of r . Moreover, the volatility of r may depend on both θ and v . Also, the Chen model unnecessarily constrains the correlations between $\theta(t)$ and $r(t)$ and between $v(t)$ and $r(t)$ to 0. All of these restrictions are examined empirically in Section II.C.

Again, we turn to the equivalent AY representation of the model specified in equations (28) to verify that it is admissible and maximal (see Appendix E, Section E.2):

$$r(t) = [\delta_0] + [\delta_1] Y_1(t) + Y_2(t) + Y_3(t), \quad (29)$$

$$\begin{aligned} d \begin{pmatrix} Y_1(t) \\ Y_2(t) \\ - \\ Y_3(t) \end{pmatrix} &= \left[\begin{array}{cc|c} \kappa_{11} & [\kappa_{12}] & 0 \\ [\kappa_{21}] & \kappa_{22} & 0 \\ \hline 0 & 0 & \kappa_{33} \end{array} \right] \left[\begin{pmatrix} \theta_1 \\ \theta_2 \\ - \\ 0 \end{pmatrix} - \begin{pmatrix} Y_1(t) \\ Y_2(t) \\ - \\ Y_3(t) \end{pmatrix} \right] dt \\ &+ \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hline [\sigma_{31}] & [\sigma_{32}] & 1 \end{array} \right] \sqrt{\begin{array}{cc|c} S_{11}(t) & 0 & 0 \\ 0 & S_{22}(t) & 0 \\ \hline 0 & 0 & S_{33}(t) \end{array}} dB(t), \end{aligned} \quad (30)$$

where

$$\begin{aligned} S_{11}(t) &= [\beta_1]_1 Y_1(t), \\ S_{22}(t) &= [\beta_2]_2 Y_2(t), \\ S_{33}(t) &= [\alpha_3] + Y_1(t) + [\beta_3]_2 Y_2(t). \end{aligned} \quad (31)$$

With the first two state variables driving volatility, κ_{12} and κ_{21} must be less than or equal to zero in order to assure that Y_1 and Y_2 remain strictly positive. That is, $(Y_1(t), Y_2(t))$ is a bivariate, correlated square-root diffusion. Additionally, admissibility requires that Y_1 and Y_2 be conditionally uncorrelated and Y_3 not enter the drift of these variables. Y_3 can be conditionally correlated with (Y_1, Y_2) , and its variance may be an affine function of (Y_1, Y_2) .

The corresponding restrictions on the AY representation in equations (29) through (31) implied by the Chen model are obtained by setting the square boxes in equation (30) to zero except that $\sigma_{32} = -1$ and $\delta_0 = -\delta_1\theta_1 - \theta_2 + q\theta_2 = -\theta_2\kappa_{22}/\kappa_{33}$. As in the BDFS model, in the Chen model r is constrained to be an affine function of only two of the three state variables.

B.4. $\mathbb{A}_3(3)$

The final subfamily of models has $m = 3$ so that all three Y s determine the volatility structure. The canonical representation of $AM_3(3)$ has parameters

$$\mathcal{K} = \begin{bmatrix} \kappa_{11} & \kappa_{12} & \kappa_{13} \\ \kappa_{21} & \kappa_{22} & \kappa_{23} \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$\Theta = (\theta_1, \theta_2, \theta_3)'$, $\alpha = 0$, and $\mathcal{B} = I_3$, where $\kappa_{ii} > 0$ for $1 \leq i \leq 3$, $\kappa_{ij} \leq 0$ for $1 \leq i \neq j \leq 3$, $\theta_i > 0$ for $1 \leq i \leq 3$.

With both Σ and $\mathcal{B}$ equal to identity matrices, the diffusion term of this model is identical to that in the N -factor model based on independent square-root diffusions (often referred to as the CIR model). With $\mathcal{B}$ diagonal, the requirements of admissibility preclude relaxation of the assumption that Σ is diagonal. However, admissibility does not require that $\mathcal{K}$ be diagonal, as in the classical CIR model, but rather only that the off-diagonal elements of $\mathcal{K}$ be less than or equal to zero (see equation (17)). Thus, the canonical representation is a correlated, square-root (CSR) diffusion model. It follows that the empirical implementations of multifactor CIR-style models with independent state variables by Chen and Scott (1993), Pearson and Sun (1994), and Duffie and Singleton (1997), among others, have imposed overidentifying restrictions by forcing $\mathcal{K}$ to be diagonal. In this three-factor model, a diagonal $\mathcal{K}$ implies six overidentifying restrictions.

C. Comparative Properties of Three-Factor ATSMs

In concluding this section, we highlight some of the similarities and differences among ATSMs and motivate the subsequent empirical investigation of three-factor models.

Positivity of the instantaneous short rate r :

As a general rule, for three-factor ATSMs, $3 - m$ of the state variables in $\mathbb{A}_m(3)$ models may take on negative values. That is, the Gaussian ($m = 0$) model allows all three state variables to become negative, the $\mathbb{A}_1(3)$ model allows two of the three state variables to become negative, and so on. Therefore, only in the case of models in $\mathbb{A}_3(3)$ are we assured that $r(t) > 0$, provided that we constrain δ_0 and all elements of δ_y are non-negative. Emphasis on having $r(t)$ positive has, at least partially, led some researchers to con-

sider models outside the affine class, such as the discrete-time “level-GARCH” model of Brenner, Harjes, and Kroner (1996) and Koedijk et al. (1997) or the diffusion specifications of Andersen and Lund (1997) and Gallant and Tauchen (1998). These authors assure positivity by having the short rate volatility be proportional to the product of powers of the short rate and another non-negative process.

Conditional second moments of zero-coupon bond yields:

For ATSMs in $\mathbb{A}_m(N)$, the conditional variances of zero-coupon-bond yields are determined by m common factors. However, the additional flexibility in specifying conditional volatilities that comes with increasing m is typically accompanied by *less* flexibility in specifying conditional correlations. The nature of the conditional correlations accommodated within $AM_m(3)$ can be seen most easily by normalizing $\mathcal{K}_{(N-m) \times m}^{\text{DB}}$ to zero and $\mathcal{K}_{(N-m) \times (N-m)}^{\text{DB}}$ to a diagonal matrix and concurrently freeing up Σ^{DB} and the off-diagonal elements of Σ^{DB} . This gives an equivalent model to our canonical representation of $\mathbb{A}_m(N)$ (see Appendix C). It follows that the admissibility constraints accommodate nonzero conditional correlations of unconstrained signs between each element of Y^{D} and the entire state vector $Y(t)$. For instance, with $N = 3$ and $m = 3$, the state variables are conditionally uncorrelated. On the other hand, with $m = 2$, only two state variables determine the volatility of $Y(t)$, but $Y_1(t)$ may be conditionally correlated with both $Y_2(t)$ and $Y_3(t)$.

Unconditional correlations among the state variables:

In the Gaussian model ($m = 0$), the signs of the nonzero elements of $\mathcal{K}$ are unconstrained, and, hence, unconditional correlations among the state variables may be positive or negative. On the other hand, for the CSR model with $m = 3$, the unconditional correlations among the state variables must be non-negative. This is an implication of the zero conditional correlations and the sign restrictions on the off-diagonal elements of $\mathcal{K}$ required by the admissibility conditions. The case of $m = 1$ is similar to the Gaussian model in that Σ may induce positive or negative conditional or unconditional correlations among the Y s. Finally, in the case of $m = 2$, the first state variable may be negatively correlated with the other two, but correlation between $Y_2(t)$ and $Y_3(t)$ must be non-negative.

Notice that a limitation of the affine family of term structure models is that one cannot simultaneously allow for negative correlations among the state variables and require that $r(t)$ be strictly positive.

These observations motivate the focus of our subsequent empirical analysis of three-factor ATSMs on the two branches $\mathbb{A}_1(3)$ and $\mathbb{A}_2(3)$. Models with $m = 1$ or $m = 2$ have the potential to explain the widely documented conditional heteroskedasticity and excess kurtosis in zero-coupon-bond yields, while being flexible with regard to both the magnitudes and signs of the admis-

sible correlations among the state variables. In contrast, models in the branch $\mathbb{A}_0(3)$ have the counterfactual implication that zero yields are conditionally normal with constant conditional second moments.

Though models in $\mathbb{A}_3(3)$ allow for time-varying volatilities and correlated factors, the requirement that conditional correlations be zero and unconditional factor correlations be positive turns out to be inconsistent with historical U.S. data. Evidence suggestive of the importance of negative factor correlations can be gleaned from previous empirical studies of multifactor CIR models, which presume that the factors are statistically independent. For instance, we computed the sample correlation between the implied state variables from the two-factor ($N = 2$) CIR model studied in Duffie and Singleton (1997) and found a correlation of approximately -0.5 instead of zero.⁷ Additional evidence on the importance of negative correlations within $\mathbb{A}_1(3)$ and $\mathbb{A}_2(3)$ is presented in Section II.C.

II. Empirical Analysis of ATSMs

A. Data and Estimation Method

For our empirical analysis we study yields on ordinary, fixed-for-variable rate U.S. dollar swap contracts. A primary motivation for this choice is that swap markets provide true “constant maturity” yield data, whereas in the Treasury market the maturities of “constant maturity” yields are only approximately constant or the data represent interpolated series. Additionally, the on-the-run treasuries that are often used in empirical studies are typically on “special” in the repo market, so, strictly speaking, the Treasury data should be adjusted for repo specials prior to an empirical analysis. Unfortunately, the requisite data for making these adjustments are not readily available, and, consequently, such adjustments are rarely made.

Though the institutional structures of dollar swap and U.S. Treasury markets are different, some of the basic distributional characteristics of the associated yields are similar. For instance, principal component analyses yield very similar looking “level,” “slope,” and “curvature” factors for both markets.⁸ Additionally, the shapes of the “term structures” of yield volatilities examined in Section II.C are qualitatively similar for both data sets. Nevertheless, empirical studies of ATSMs with swap and Treasury yield data may lead to different conclusions for at least two reasons. One is the difference in institutional features outlined above. Another is that the histories of yield data available in swap markets are in general shorter than those for

⁷ Duffie and Singleton (1997) assume that the two- and 10-year swap yields are priced perfectly by their two-factor model. Thus, using their pricing model evaluated at the maximum likelihood estimates of the parameters, implied state variables are computed as functions of these two swap yields.

⁸ See, for example, Litterman and Scheinkman (1997) for a principal components analysis of the U.S. Treasury data.

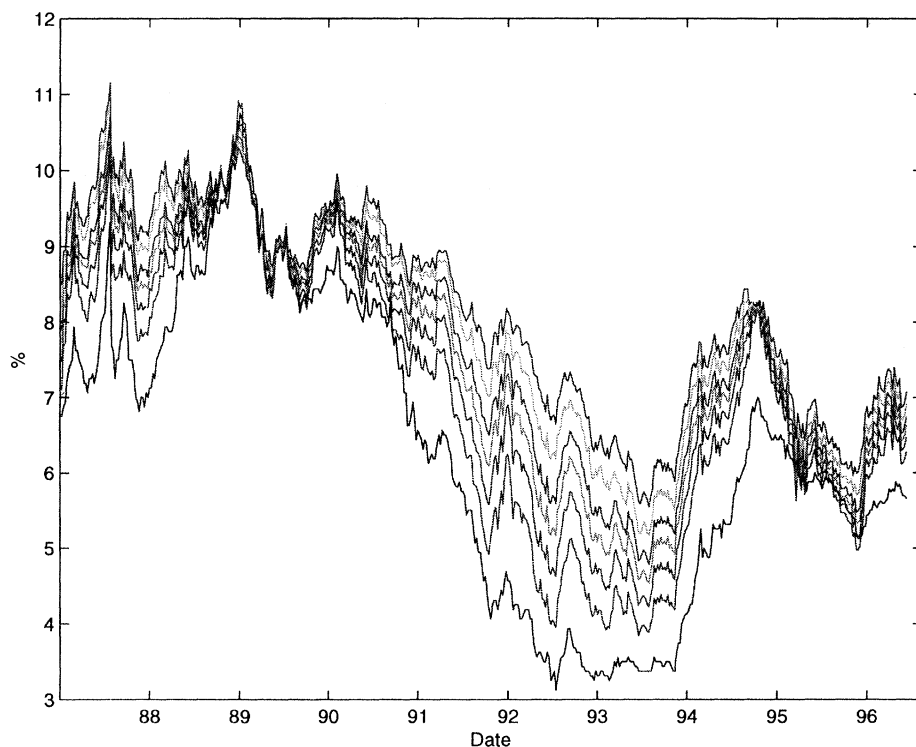


Figure 1. Time Series of Swap Yields. Our sample covers the period from April 3, 1987, to August 23, 1996, weekly. The yields plotted in this graph include, from the lowest to the highest line (with occasional cross-overs when the yield curves are inverted), six-month LIBOR, two-, three-, four-, five-, seven-, and 10-year swap rates. The maturities used for estimation are the six-month LIBOR, two-, and 10-year swap rates.

Treasury markets. In particular, available histories of swap yields do not include the periods of the oil price shocks of the early 1970s, the monetary experiment in the early 1980s, and so on.

The observed data y were chosen to be the yields on six-month LIBOR and two-year and 10-year fixed-for-variable rate swaps sampled weekly from April 3, 1987, to August 23, 1996 (see Figure 1 for a time series plot of the LIBOR and swap yields). The length of the sample period was determined in part by the unavailability of reliable swap data for years prior to 1987. The yields are ordered in y according to increasing maturity (i.e., y_1 is the six-month LIBOR rate, etc.).

The conditional likelihood function of the state vector $Y(t)$ is not known for general affine models. Therefore, we pursue the method of simulated moments (SMM) proposed by Duffie and Singleton (1993) and Gallant and Tauchen (1996). A key issue for the SMM estimation strategy is the selection of moments. Following Gallant and Tauchen (1996), we use the scores of the likelihood function from an auxiliary model that describes the time series

properties of bond yields as the moment conditions for the *SMM* estimator. More precisely, let y_t denote a vector of yields on bonds with different maturities, $x'_t = (y'_t, y'_{t-1}, \dots, y'_{t-\ell})$, and $f(y_t|x_{t-1}, \gamma)$ denote the conditional density of y associated with the auxiliary description of the yield data. We searched for the best f for our data set along numerous model expansion paths, guided by a model selection criterion, as outlined in Gallant and Tauchen (1996). The end result of this search was the auxiliary model

$$f(y_t|x_{t-1}, \gamma) = c(x_{t-1})[\epsilon_0 + [h(z_t|x_{t-1})]^2]n(z_t), \quad (32)$$

where $n(\cdot)$ is the density function of the standard normal distribution, ϵ_0 is a small positive number,⁹ $h(z|x)$ is a Hermite polynomial in z , $c(x_{t-1})$ is a normalization constant, and x_{t-1} is the conditioning set. We let z_t be the normalized version of y_t , defined by

$$z_t = R_{x,t-1}^{-1}(y_t - \mu_{x,t-1}). \quad (33)$$

In the terminology of Gallant and Tauchen (1996), the auxiliary model may be described as “Non-Gaussian, VAR(1), ARCH(2), Homogeneous-Innovation.” “VAR(1)” refers to the fact that the shift vector $\mu_{x,t-1}$ is linear with elements that are functions of $L_\mu = 1$ lags of y , in that

$$\mu_{x,t-1} = \begin{pmatrix} \psi_1 + \psi_4 y_{1,t-1} + \psi_7 y_{2,t-1} + \psi_{10} y_{3,t-1} \\ \psi_2 + \psi_5 y_{1,t-1} + \psi_8 y_{2,t-1} + \psi_{11} y_{3,t-1} \\ \psi_3 + \psi_6 y_{1,t-1} + \psi_9 y_{2,t-1} + \psi_{12} y_{3,t-1} \end{pmatrix}. \quad (34)$$

“ARCH(2)” refers to the fact that the scale transformation $R_{x,t-1}$ is taken to be of the $\text{ARCH}(L_r)$ -form, with $L_r = 2$,

$$R_{x,t-1} = \begin{bmatrix} \tau_1 + \tau_7 |\epsilon_{1,t-1}| & \tau_2 & \tau_4 \\ + \tau_{25} |\epsilon_{1,t-2}| & & \\ 0 & \tau_3 + \tau_{15} |\epsilon_{2,t-1}| & \tau_5 \\ & + \tau_{33} |\epsilon_{2,t-2}| & \\ 0 & 0 & \tau_6 + \tau_{24} |\epsilon_{3,t-1}| \\ & & + \tau_{42} |\epsilon_{3,t-2}| \end{bmatrix}, \quad (35)$$

⁹ Our implementation of SMM with an auxiliary model differs from many previous implementations by our inclusion of the constant ϵ_0 in the SNP density function. Though ϵ_0 is identified if the scale of $h(z|x)$ is fixed, Gallant and Long (1997) encountered numerical instability in estimating SNP models with ϵ_0 treated as a free parameter. Therefore, we chose to fix both ϵ_0 and the constant term of $h(z|x)$ at nonzero constants. We set $\epsilon_0 = 0.01$ and verified that the estimated auxiliary model was essentially unchanged by this setting instead of zero.

where $\epsilon_t = y_t - \mu_{x,t-1}$. Thus, the starting point for our SNP conditional density for y is a first-order vector autoregression (VAR), with innovations that are conditionally normal and follow an ARCH process of order two: $n(y|\mu_x, \Sigma_x)$, where $\Sigma_{x,t-1} = R_{x,t-1}R'_{x,t-1}$.

“Non-Gaussian” refers to the fact that the conditional density is obtained by scaling the normal density $n(z_t)$ (the “Gaussian, VAR(1), ARCH(2)” part) by the square of the Hermite polynomial $h(z_t|x_{t-1})$, where h is a polynomial of order $K_z = 4$ in z_t , that is,

$$h(z_t|x_{t-1}) = A_1 + \sum_{l=1}^4 \sum_{i=1}^3 A_{3(l-1)+1+i} z_{t,t}^l. \quad (36)$$

Finally, “Homogeneous-Innovation” refers to the fact that the coefficients in the Hermite polynomial $h(z_t|x_{t-1})$ are constants, independent of the conditioning information.¹⁰

Having selected the moment conditions, we proceed using the standard “optimal” GMM criterion function (Hansen (1982), Duffie and Singleton (1993)), a quadratic form in the sample moments

$$\frac{1}{T} \sum_{t=1}^T \frac{\partial}{\partial \gamma} \log f(y_t^\phi | x_{t-1}^\phi, \gamma_T). \quad (37)$$

Under regularity, this SMM estimator is consistent for ϕ , even if the auxiliary model does not describe the true joint distribution of y_t . Efficiency considerations, on the other hand, motivated our extensive search for among semi-nonparametric specifications of $f(y_t|x_{t-1}, \gamma)$. The analysis in Gallant and Long (1997) implies that, for our term structure model and selection strategy for an auxiliary density $f(y_t|x_{t-1}, \gamma)$, our SMM estimator is asymptotically efficient.¹¹

The auxiliary model selected shows that the time series of LIBOR and swap rates over the sample period we examined are remarkably “uneventful”: a low order ARCH-like specification was able to capture the time variation in the conditional second moments. This contrasts with the findings in the Treasury market by Andersen and Lund (1997), for example, who found, for a different sample period, a GARCH-like or high order ARCH-like specification for the conditional variance.

¹⁰ Note that the term “Homogeneous-Innovation” does not imply that the conditional variances of the yields are constants (because of the “ARCH(2)” terms).

¹¹ More precisely if, for a given order of the polynomial terms in the approximation to the density f , sample size is increased to infinity, and then the order of the polynomial is increased, the resulting SMM estimator approaches the efficiency of the maximum likelihood estimator. It follows that our SMM estimator is more efficient (asymptotically) than the quasi-maximum likelihood estimator proposed recently by Fisher and Gilles (1996).

With A_1 normalized to 1, the free parameters of the SNP model are

$$\begin{aligned} \gamma &= (A_j: 2 \leq j \leq 13; \psi_j: 1 \leq j \leq 12; \\ \tau_j: j &= 1, 2, \dots, 7, 15, 24, 25, 33, 42). \end{aligned} \quad (38)$$

B. Identification of the Market Prices of Risk

In Gaussian and square-root diffusion models of $Y(t)$, the parameters λ governing the term premiums enter the $A(\tau)$ and $B(\tau)$ in equation (5) symmetrically with other parameters, and this leads naturally to the question of under what circumstances λ is identified in ATSMs. This section argues that λ is generally identified, except for certain Gaussian models.

The “identification condition” in GMM estimation is the assumption that the expected value of the derivative of the moment equations with respect to the model parameters have full rank (Hansen (1982), Assumption 3.4). To simplify notation in our setting, we let $z'_t \equiv (y_t, x'_{t-1})$ and $f(z_t, \gamma)$ denote the auxiliary SNP conditional density function used to construct moment conditions. Using this notation, the rank condition for our SMM estimation problem is that the matrix

$$D_0 \equiv E \left[\frac{\partial^2 \log f}{\partial \gamma \partial \phi'} (z_t^{\phi_0}, \gamma_0) \right] = E \left[\frac{\partial^2 \log f}{\partial \gamma \partial z_t^{\phi'}} (z_t^{\phi_0}, \gamma_0) \frac{\partial z_t^{\phi_0}}{\partial \phi'} \right] \quad (39)$$

has full rank. The rank of D_0 is at most $\min(\dim(\gamma), \dim(\phi))$, so clearly a necessary condition for identification is that $\dim(\gamma) \geq \dim(\phi)$, where $\dim$ denotes the dimension of the vector.

The market price of risk λ (or any other parameter) is not identified if there exists another parameter, say $\delta_0 \in \phi$, such that $\partial z_t^\phi / \partial \lambda$ and $\partial z_t^\phi / \partial \delta_0$ are collinear, in which case D_0 clearly does not have full rank. For an example, it is easy to check that in the case of the one-factor Gaussian model estimated with a zero yield, the market price of risk is not identified because the partial derivatives of the zero yield with respect to both λ and δ_0 are constant and therefore proportional to each other.

In a one-factor setting, there are two sources of identification of λ . One is the use of coupon bond yields instead of zero yields to estimate the model. The nonlinear mapping between the coupon yield and the underlying (Gaussian or otherwise) state variable implies that both $\partial z_t^\phi / \partial \lambda$ and $\partial z_t^\phi / \partial \theta$ are state dependent and are not collinear. Another source of identification is the assumption that the state variable follows a non-Gaussian process. In the case of a square-root (CIR) model, for example, it is easy to check that, even though $\partial z_t^\phi / \partial \lambda$ is a constant, none of the partial derivatives of z_t^ϕ with respect to the other structural parameters is a constant. Consequently, λ is identified.

This intuition for one-factor models can easily be generalized to the case of N -factor affine models for $N > 1$. The identification results for the general case are as follows:¹² When zero yields are used to estimate a Gaussian model (in the $\mathbb{A}_0(N)$ branch), one out of N market prices of risk is not identified.¹³ When zero yields are used to estimate non-Gaussian models (in the $\mathbb{A}_m(N)$ branch with $1 \leq m \leq N$), one out of N market prices of risk is not identified unless, at least for one k with $m + 1 \leq k \leq N$, β_k is not identically zero. When coupon yields are used to estimate affine models, all of the market prices of risk are identified.

Intuitively, in cases where all market prices are identified, the common source of identification is the stochastic or time-varying nature of risk premia, which can be induced either by a non-Gaussian model or a nonlinear mapping between the observed yields and the underlying state variables. When the risk premia are not time varying, they may not be separately identified from the level of the unobserved short rate.¹⁴

C. Empirical Analysis of Swap Yield Curves

We estimated six ATSMs in $\mathbb{A}_1(3)$ and $\mathbb{A}_2(3)$ and report the overall goodness-of-fit, chi-square tests for these models in Table I. Both the Chen and BDFS models, denoted by $\mathbb{A}_1(3)_{\text{BDFS}}$ and $\mathbb{A}_2(3)_{\text{Chen}}$, respectively, have large chi-square statistics relative to their degrees of freedom. In contrast, the corresponding *maximal* models, denoted by $\mathbb{A}_m(3)_{\text{Max}}$, for $m = 1, 2$, are not rejected at conventional significance levels. However, the improved fits of $\mathbb{A}_1(3)_{\text{Max}}$ (compared to $\mathbb{A}_1(3)_{\text{BDFS}}$) and $\mathbb{A}_2(3)_{\text{Max}}$ (compared to $\mathbb{A}_2(3)_{\text{Chen}}$) were achieved with six and eight additional degrees of freedom, so we were concerned about overfitting. This concern was reinforced by the relatively large standard errors for most of the estimated parameters in the Max models, displayed in the second columns of Tables II and III. Therefore, we also present the results for two intermediate models, $\mathbb{A}_1(3)_{\text{DS}}$ and $\mathbb{A}_2(3)_{\text{DS}}$ (the *DS* indicating that these are our preferred models), that constrain some of the parameters in the Max models. Relative to the more constrained BDFS and Chen models, the $\mathbb{A}_1(3)_{\text{DS}}$ and $\mathbb{A}_2(3)_{\text{DS}}$ models allow nonzero values of the parameters $(\sigma_{r\theta}, \sigma_{\theta r})$ and $(\kappa_{\theta v}, \kappa_{rv}, \sigma_{rv})$, respectively. The *DS* models are not rejected at conventional significance levels and have fewer parameters than the Max models, and most of the estimated parameters are statistically significant at conventional levels. Therefore, we will focus primarily on the *DS* models in subsequent discussion.

¹² Details are available from the authors upon request and from the *Journal of Finance* website.

¹³ More precisely, δ_0 and the N prices of risk can not be separately identified. The fact that there is only one underidentified market price of risk in an N -factor model with $N - m$ Gaussian-like factors may appear surprising at first. This is explained by the fact that we have constrained the long run means of the Gaussian-like factors to zero so that δ_0 is the only free parameter that determines the overall level of the instantaneous short rate.

¹⁴ This intuition suggests that if the instantaneous short rate is observed, or if the model is forced to exactly match the unconditional mean of an additional yield, an otherwise unidentified market price of risk can become identified.

Table I
Overall Goodness-of-Fit

This table lists the test statistics (third column) for overidentifying restrictions obtained by setting the simulated sample moments in equation (37) to zero. The degree of freedom (fourth column) is the difference between the dimension of γ_T and that of ϕ . The fifth column gives the p -values for a chi-square distribution with the specified degrees of freedom. $\mathbb{A}_1(3)_{\text{Max}}$, specified in equation (23), is the “maximal” model in branch $\mathbb{A}_1(3)$. $\mathbb{A}_1(3)_{\text{BDFS}}$ is obtained by constraining all parameters in the square boxes in equation (23) to zero. $\mathbb{A}_1(3)_{\text{DS}}$ is our preferred model in this branch, obtained by relaxing $\sigma_{\theta r}$ and $\sigma_{r\theta}$ from $\mathbb{A}_1(3)_{\text{BDFS}}$. $\mathbb{A}_2(3)_{\text{Max}}$, specified in equation (28), is the “maximal” model in branch $\mathbb{A}_2(3)$. $\mathbb{A}_2(3)_{\text{Chen}}$ is obtained by constraining all parameters in the square boxes in equation (28) to zero, except that $\bar{r} = \bar{\theta}$. $\mathbb{A}_2(3)_{\text{DS}}$ is our preferred model in this branch, obtained by relaxing $\kappa_{\theta v}$ and σ_{rv} from $\mathbb{A}_2(3)_{\text{Chen}}$. κ_{rv} is not a free parameter but is nonzero. It has a deterministic relationship to $\kappa_{\theta v}$ through their dependence on a common degree of freedom, that is, κ_{21} (see equation (A9)).

Branch $\mathbb{A}_1(3)$		χ^2	d.f.	p -value
$\mathbb{A}_1(3)_{\text{BDFS}}$	BDFS	84.212	25	0.000%
$\mathbb{A}_1(3)_{\text{DS}}$	BDFS + $\sigma_{\theta r}$ + $\sigma_{r\theta}$	28.911	23	18.328%
$\mathbb{A}_1(3)_{\text{Max}}$	Maximal	28.901	19	6.756%
Branch $\mathbb{A}_2(3)$		χ^2	d.f.	p -value
$\mathbb{A}_2(3)_{\text{Chen}}$	Chen	129.887	26	0.000%
$\mathbb{A}_2(3)_{\text{DS}}$	Chen + $\kappa_{\theta v}$ + (κ_{rv}) + σ_{rv}	22.931	24	52.387%
$\mathbb{A}_2(3)_{\text{Max}}$	Maximal	16.398	18	56.479%

The reason that the *DS* models do a better job “explaining” the swap rates, as measured by the χ^2 statistics, than the $\mathbb{A}_1(3)_{\text{BDFS}}$ and $\mathbb{A}_2(3)_{\text{Chen}}$ models is that the former allow a more flexible correlation structure of the state variables. In the $\mathbb{A}_1(3)$ branch, the $\mathbb{A}_1(3)_{\text{BDFS}}$ model only allows a nonzero conditional correlation between the short rate and its stochastic volatility ($\sigma_{rv} \neq 0$). The $\mathbb{A}_1(3)_{\text{DS}}$ model also allows the short rate and its stochastic central tendency to be conditionally correlated ($\sigma_{r\theta} \neq 0$ and $\sigma_{\theta r} \neq 0$). (Recall, from equation (23), that relaxing these constraints affects both the $\theta(t)$ and $r(t)$ processes.) In the $\mathbb{A}_2(3)$ branch, the $\mathbb{A}_2(3)_{\text{Chen}}$ model assumes that $r(t)$, $\theta(t)$, and $v(t)$ are all pairwise, conditionally uncorrelated. In contrast, the model $\mathbb{A}_2(3)_{\text{DS}}$ allows the short rate to be conditionally correlated with its stochastic volatility ($\sigma_{rv} \neq 0$) and allows the stochastic volatility to influence the conditional mean of the short rate ($\kappa_{rv} \neq 0$) and its stochastic central tendency ($\kappa_{\theta v} \neq 0$).¹⁵

Moreover, in both branches, it is the introduction of *negative* conditional correlations among the state variables that seems to be important (see the third columns of Tables II and III). Such negative correlations are ruled out a priori in the CSR models (family $\mathbb{A}_3(3)$ in Section I). Hence, these findings support our focus on the branches $\mathbb{A}_1(3)$ and $\mathbb{A}_2(3)$ in attempting to describe the conditional distribution of swap yields.

¹⁵ Though we relax three constraints, this amounts to two additional degrees of freedom, because $\kappa_{\theta v}$ and κ_{rv} are controlled by the single parameter κ_{21} in the AY representation, given the constraint $\delta_1 = 0$. See equation (A9).

Table II
EMM Estimators for Branch $\mathbb{A}_1(3)$

The EMM estimators reported here and in Table III are obtained by minimizing the EMM criterion function based on simulated sample moments in equation (37). The parameters in the first column pertain to the *Ar* representation (equation (23)) of models in the $AYM_1(3)$ branch. Parameters indicated by “fixed” are restricted to zero. *t*-ratios are in parentheses.

Parameter	Estimate (<i>t</i> -ratio)		
<i>Ar</i>	$\mathbb{A}_1(3)_{\text{BDFS}}$	$\mathbb{A}_1(3)_{\text{DS}}$	$\mathbb{A}_1(3)_{\text{Max}}$
μ	0.602 (4.246)	0.365 (6.981)	0.366 (5.579)
ν	0.0523 (7.138)	0.226 (14.784)	0.228 (9.720)
κ_{rv}	0 (fixed)	0 (fixed)	0.0348 (0.001)
κ	2.05 (9.185)	17.4 (3.405)	18 (3.547)
$\bar{v}$	0.000156 (6.630)	0.015 (1.544)	0.0158 (1.323)
$\bar{\theta}$	0.14 (8.391)	0.0827 (16.044)	0.0827 (13.530)
$\sigma_{\theta v}$	0 (fixed)	0 (fixed)	0.0212 (0.012)
σ_{rv}	491 (3.224)	4.27 (2.645)	4.2 (1.936)
$\sigma_{r\theta}$	0 (fixed)	-3.42 (-1.754)	-3.77 (-1.668)
$\sigma_{\theta r}$	0 (fixed)	-0.0943 (-2.708)	-0.0886 (-2.470)
ζ^2	0.000113 (9.493)	0.0002 (4.069)	0.000208 (2.742)
α_r	0 (fixed)	0 (fixed)	3.26e-14 (0.000)
η^2	5.18e-05 (1.934)	0.00782 (1.382)	0.00839 (1.222)
β_θ	0 (fixed)	0 (fixed)	7.9e-10 (0.000)
λ_v	6.79e+04 (1.886)	-0.344 (-0.064)	-0.27 (-0.045)
λ_θ	31 (3.392)	31.7 (2.410)	30.2 (1.289)
λ_r	1.66e+03 (3.312)	9.32 (2.296)	9.39 (1.153)

Figure 2 displays the *t*-ratios for testing whether the fitted scores of the auxiliary model, computed from the models $\mathbb{A}_1(3)_{\text{BDFS}}$, $\mathbb{A}_1(3)_{\text{DS}}$, $\mathbb{A}_2(3)_{\text{Chen}}$, and $\mathbb{A}_2(3)_{\text{DS}}$, are zero. For a correctly specified model (and assuming that asymptotic approximations to distributions are reliable), the sample scores should be small relative to their standard errors. The graphs for models $\mathbb{A}_1(3)_{\text{BDFS}}$ and $\mathbb{A}_2(3)_{\text{Chen}}$ show that about half of the fitted scores have large-sample *t*-ratios larger than 2. In contrast, only two of the *t*-ratios are larger than 2 for model $\mathbb{A}_1(3)_{\text{DS}}$, and none are larger than 2 for model $\mathbb{A}_2(3)_{\text{DS}}$.

The first 12 scores of the auxiliary model, marked by “A” near the horizontal axis, are associated with parameters that govern the non-normality of the conditional distribution of the swap yields. The second 12 scores, marked by “ Ψ ,” are related to parameters describing the conditional first moments, and the last 12 scores, marked by “ τ ,” are related to the parameters of the conditional covariance matrix. A notable feature of the *t*-ratios for the individual scores is that they are often large for models $\mathbb{A}_1(3)_{\text{BDFS}}$ and $\mathbb{A}_2(3)_{\text{Chen}}$ for all three groups A, Ψ , and τ . Thus, the additional nonzero conditional correlations in the *DS* models help explain not only the conditional second moments of swap yields but their persistence and non-normality also.

The estimated values of the parameters for the *Ar* representations of the models are displayed in Tables II and III. Though perhaps not immediately evident from the *Ar* representations, the *DS* models maintain the constraint

Table III
EMM Estimators for Branch $\mathbb{A}_2(3)$

The parameters pertain to the Ar representation (equation(28)) of models in the $AYM_2(3)$ branch. Parameters indicated by “fixed” are restricted to zero except that $\bar{r}$ is constrained to be equal to $\bar{\theta}$, whenever it is “fixed.” t -ratios are in parentheses.

Parameter	Estimate (t -ratio)		
Ar	$\mathbb{A}_1(3)_{Chen}$	$\mathbb{A}_1(3)_{DS}$	$\mathbb{A}_1(3)_{Max}$
μ	1.24 (4.107)	0.636 (4.383)	0.291 (1.648)
$\kappa_{\theta v}$	0 (fixed)	-33.9 (-2.377)	-12.4 (-1.132)
κ_{rv}	0 (fixed)	-35.3 (-2.367)	-274 (-1.077)
$\kappa_{v\theta}$	0 (fixed)	0 (fixed)	-0.0021 (-0.411)
ν	0.0757 (4.287)	0.103 (2.078)	0.0871 (1.027)
κ	2.19 (8.618)	2.7 (7.432)	3.54 (4.457)
$\bar{v}$	0.000206 (7.456)	0.000239 (5.792)	0.000315 (2.051)
$\bar{\theta}$	0.0416 (7.909)	0.0259 (4.006)	0.0136 (0.994)
$\bar{r}$	0.0416 (fixed)	0.0259 (fixed)	0.053 (2.784)
σ_{rv}	0 (fixed)	-182 (-3.620)	-133 (-2.438)
$\sigma_{r\theta}$	0 (fixed)	0 (fixed)	-0.0953 (-0.167)
α_r	0 (fixed)	0 (fixed)	1.12e-09 (0.000)
η^2	0.000393 (2.873)	0.000119 (2.083)	7.04e-05 (1.679)
ζ^2	0.00253 (7.507)	0.00312 (3.129)	0.00237 (0.002)
β_θ	0 (fixed)	0 (fixed)	1.92e-05 (0.002)
λ_v	-1.9e+03 (-3.946)	1.3e+04 (1.761)	7.58e+03 (1.064)
λ_θ	-35.2 (-5.080)	-152 (-2.923)	-174 (-1.050)
λ_r	-121 (-4.973)	-692 (-4.066)	-349 (-1.540)

from the $\mathbb{A}_1(3)_{BDFS}$ and $\mathbb{A}_2(3)_{Chen}$ models that $\delta_1 = 0$. That is, in the AY representation, the instantaneous riskless rate is an affine function of only the second and third state variables. The test statistics in Table I suggest that this constraint is not inconsistent with the data.

The estimates of the mean reversion parameters (μ, ν, κ) of the state variables $(v(t), \theta(t), r(t))$ are (0.37, 0.23, 17.4) and (0.64, 0.10, 2.7) for models $\mathbb{A}_1(3)_{DS}$ and $\mathbb{A}_2(3)_{DS}$, respectively. As in previous empirical studies (e.g., Balduzzi et al. (1996) and Andersen and Lunk (1998)), the “central tendency” factor $\theta(t)$ shows much slower mean reversion (smaller ν) than the rate κ at which gaps between θ and r are closed in the short rate equation. Put differently, in model $\mathbb{A}_1(3)_{DS}$, $r(t)$ reverts relatively quickly to a process $\theta(t)$ that is itself reverting slowly to a constant long-run mean $\bar{\theta}$.¹⁶ In both DS models, the “central tendency” factor has the smallest mean reversion.

An important cautionary note at this juncture is that comparisons across models of mean reversion coefficients (or, more generally, coefficients of the drifts) may not be meaningful even if the models are nested. The reason is that changing the correlations among the state variables can be thought of as a “rotation” of the unobserved states $Y(t)$. Therefore, the meaning of labels such as “central tendency” or “volatility” in terms of yield curve move-

¹⁶ This interpretation does not hold exactly in model $\mathbb{A}_2(3)_{DS}$, because κ_{rv} is nonzero.

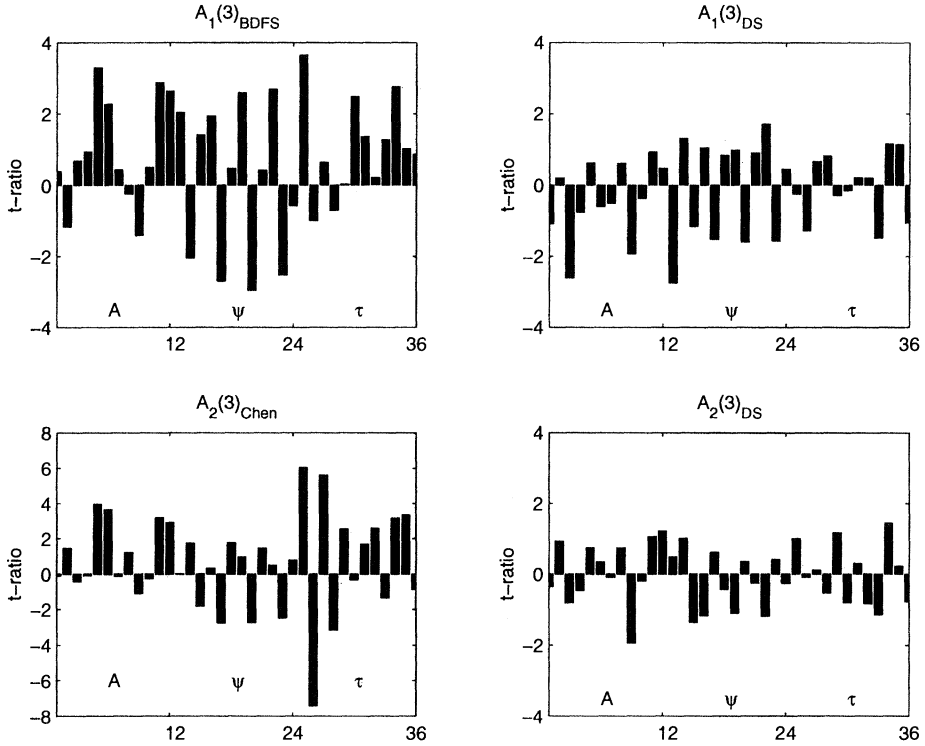


Figure 2. Fitted SNP Scores. Each subpanel plots the fitted SNP scores for one of the four models, normalized by their standard deviations. The fitted scores are the simulated sample moments given in equation (37), evaluated at the EMM estimators. Each normalized score has an asymptotic standard normal distribution. The horizontal axes are the SNP parameters γ . The first group of 12 is designated by the letter “A,” the middle group of 12 is designated by the letter “ Ψ ,” and the last group of 12 is designated by the letter “ τ .” The meanings of the parameters and the interpretations of the t -ratios are discussed in Section II.C.

ments may not be the same across models. To illustrate this point, consider the models in $A_1(3)$. In the model $A_1(3)_{BDFS}$, the correlation between changes in $\theta(t)$ and changes in the 10-year swap rate is 0.98. The close association between the long-term swap rate and central tendency is intuitive, because $r(t)$ mean reverts to $\theta(t)$. Nevertheless, this interpretation is not invariant to relaxation of the constraints $\sigma_{\theta r} = 0$ and $\sigma_{r\theta} = 0$, which gives model $A_1(3)_{DS}$. In the latter model, changes in $\theta(t)$ are most highly correlated with changes in the two-year swap rate (correlation = 0.95). Because the two- and 10-year swap rates differ in their persistence, this explains the larger value of ν (faster mean reversion of $\theta(t)$) in model $A_1(3)_{DS}$ than in model $A_1(3)_{BDFS}$.¹⁷

¹⁷ Similar observations apply to the volatility factor $v(t)$. In both models, $v(t)$ is well proxied by a butterfly position that is long 10-year swap and LIBOR contracts and short two-year contracts. However, the weights in these butterflies turn out to be quite different across the models.

Table IV
Moments of Pricing Errors (in basis points)

“Mean” (second column) is the sample mean of the pricing errors for the swap yields and models indicated in the first column. “Std” (third column) is the sample standard deviation and “ ρ ” (fourth column) is the first-order autocorrelation of the pricing errors. The columns labeled *Q-Invert* and *Q-Steep* display the sample means of the pricing errors for the days on which the slope of the yield curve was in the lowest (inverted) and highest (steep) quartile of its distribution.

Model/Swap		Mean	Std.	ρ	Q-Invert	Q-Steep
$\mathbb{A}_1(3)_{DS}$	3 yr	-11.3	9.6	0.95	-8.1	-16.6
	5 yr	16.9	16.5	0.97	-12.0	-26.6
	7 yr	-12.7	10.1	0.94	-9.1	-17.6
$\mathbb{A}_2(3)_{DS}$	3 yr	-43.1	11.6	0.97	-55.4	-27.7
	5 yr	-63.3	12.1	0.96	-75.6	-49.1
	7 yr	-47.5	8.3	0.94	-54.1	-38.1

Does the evidence recommend one of the intermediate models, $\mathbb{A}_1(3)_{DS}$ or $\mathbb{A}_2(3)_{DS}$, over the other? Ultimately, the answer to this question must depend on how the models will be used (e.g., risk management, pricing options, etc.). Even within the term structure context, these models are not nested, so formal assessments of relative fit are nontrivial. However, we offer several observations that suggest that, focusing on term structure dynamics within the affine family, model $\mathbb{A}_1(3)_{DS}$ provides a somewhat better fit. Consider first the properties of the time series of pricing errors. Table IV presents the within-sample means, standard deviations, and first-order autocorrelations of the pricing error for the yields on swaps with the three intermediate maturities three, five, and seven years, none of which were used in estimating the parameters.¹⁸ Model $\mathbb{A}_1(3)_{DS}$ has notably smaller average pricing errors than model $\mathbb{A}_2(3)_{DS}$, though both models have a tendency to imply higher yields than what we observed.

Second, the feedback effect in the drift due to $\kappa_{rv} \neq 0$ and $\kappa_{\theta v} \neq 0$ in model $\mathbb{A}_2(3)_{DS}$ is also accommodated by model $\mathbb{A}_1(3)_{DS}$. However, the results for model $\mathbb{A}_1(3)_{DS}$ suggest that nonzero values of these κ s are not essential for fitting the moments of swap yields used in estimation, once $\sigma_{\theta r}$ and $\sigma_{r\theta}$ are allowed to be nonzero. Within model $\mathbb{A}_2(3)_{DS}$, the admissibility conditions preclude relaxation of the constraint $\sigma_{\theta r} = 0$, because of its richer formulation of conditional volatility. Admissibility also requires that $\kappa_{\theta v}$ (and therefore κ_{rv}) be negative. Consequently, the stochastic central tendency and the stochastic volatility must have a positive unconditional correlation. Taken together, these findings suggest that the negative correlations among the state variables called for by the data are not easily accommodated within the $\mathbb{A}_2(3)$ branch.

¹⁸ These pricing errors were computed by inverting the models for the implied values of the state variables, using the six-month and two- and 10-year swap yields, and then computing the differences between the actual and model-implied swap rates for the intermediate maturities, with the latter evaluated at the implied state variables.

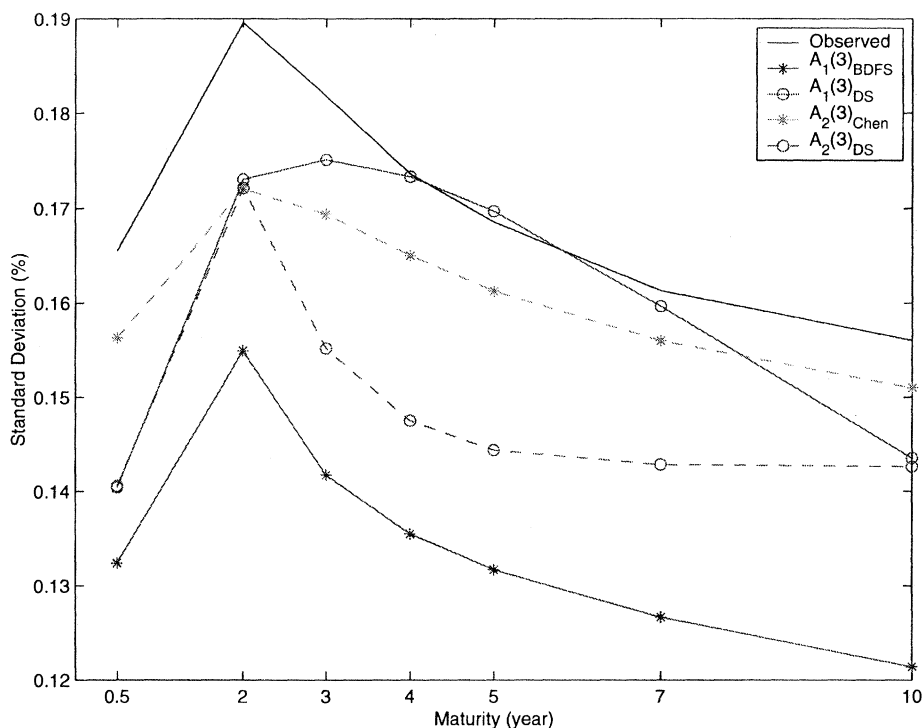


Figure 3. Term Structure of Yield Volatility. The volatility here is defined as the sample standard deviation of *weekly changes* of LIBOR and swap yields (see Figure 1). Except for the observed term structure of volatility (line), which is computed from the observed yields, the other four curves are computed from yields simulated from the four models, using EMM estimators reported in Table II and Table III.

Third, we also examined the shapes of the implied term structures of (unconditional) swap yield volatilities. The solid, uppermost line in Figure 3 displays the historical sample standard deviations of differences in yields. The other lines display the sample variances computed using long, simulated time series of swap yields from the models evaluated at their estimated parameter values.¹⁹ Notably, the term structure of historical sample volatilities is hump-shaped, with a peak around two years. Hump-shaped volatility curves can be induced in ATSMs either through negative correlation among the state variables or by hump-shaped loadings $B(\tau)$ on $Y(t)$ in equation (5). The intuition for this lies in the interplay between the negative correlations among the shocks to the risk factors and different speeds of mean reversion of the state variables. For expository ease, consider the case

¹⁹ We stress that the model-implied volatilities were computed by simulation and *not* from yields computed with the implied state variables. Thus, Figure 3 displays the population volatilities implied by the models, conditional on the estimated parameter values. We have found that using implied swap yields to compute sample moments often leads to substantially biased estimates of the population values.

of two factors where the first factor has a faster rate of mean reversion (larger κ) than the second. In affine models, κ plays a critical role in the rate at which the factor weights (the $B(\tau)$ in equation (7)) tend to zero as maturity τ is increased. At short maturities, the volatilities of both factors will typically affect overall yield volatility. As τ increases, the influence of the first factor will die out at a faster rate than that of the second factor. Thus, for long maturities, yield volatility will be driven primarily by the second factor and volatility will decline with maturity. A hump can occur, because the negative correlation contributes to a lower yield volatility at the shorter maturities. As maturity increases, the negative contribution of correlation to yield volatility declines as the importance of the first factor declines. That models with independent, mean-reverting state variables cannot induce a hump can be seen from inspection of the loadings implied by the CIR model. Models in $\mathbb{A}_1(3)$ and $\mathbb{A}_2(3)$ can exploit both of these mechanisms to match historical volatilities (whereas models in $\mathbb{A}_3(3)$ only have the latter mechanism).²⁰ All of the model-implied, volatility term structures in Figure 3 have a hump. However, model $\mathbb{A}_1(3)_{DS}$ appears to fit the volatility of swap yields much better than model $\mathbb{A}_2(3)_{DS}$.

Finally, when we computed the implied yield curves from model $\mathbb{A}_2(3)_{DS}$, we found that there were often pronounced “kinks” at the short end of the yield curve, whereas those implied by model $\mathbb{A}_1(3)_{DS}$ were generally smooth.

In light of the small goodness-of-fit statistics for the $\mathbb{A}_2(3)_{DS}$ model, we were puzzled by the frequency of kinks in yield curves, the large average pricing errors, and the underestimation of yield volatilities. The preceding discussion of the constraints on the conditional correlations implied by the admissibility conditions, together with inspection of the form of the risk-neutral drifts, leads us to the following conjecture: the market prices of risk were set, in part, to replicate the effects of a nonzero $\sigma_{\theta r}$ (which cannot be done directly) at the expense of sensible shapes of implied yield curves and smaller pricing errors. To explore the validity of this conjecture, we simply reduced the market prices of risk by 20 percent in absolute value in model $\mathbb{A}_2(3)_{DS}$ and found that the implied yield curves were essentially free of kinks and, equally importantly, seemed to line up well with the historical yield curves.²¹

There is also evidence that all of the models examined fail to capture some aspects of swap yield distributions. In particular, in Table IV, columns 5 and 6, we report the average pricing errors for dates when the slope of the swap curve was in the lowest (“Q-Invert”) and highest (“Q-Steep”) quartiles of the historically observed slopes.²² In the case of model $\mathbb{A}_1(3)_{DS}$, the average pricing errors are larger when the swap curve is steeply upward sloping than when it is inverted. The reverse is true for model $\mathbb{A}_2(3)_{DS}$. This suggests

²⁰ These observations provide further motivation for our interest in the branches $\mathbb{A}_1(3)$ and $\mathbb{A}_2(3)$.

²¹ The criterion function used in estimation does not impose a penalty for kinks in spot curves or choppy forward-rate curves. Such penalties could, of course, be introduced in practice. Nor does the criterion function force the means of the swap rates observed historically and simulated from the models to be the same.

²² Slope is the difference between the 10- and two-year swap yields.

that there may be some omitted *nonlinearity* in these affine models.²³ Also, though the standard deviations of the pricing errors are small relative to those of the swap yields themselves, the errors are highly persistent (see column 4 of Table IV). Such persistence points to some misspecification of the model for intermediate maturities.

III. Conclusion

In this paper we present a complete characterization of the admissible and identified affine term structure models, according to the most general known sufficient conditions for admissibility. For N -factor models, there are $N + 1$ non-nested classes of admissible models. For each class, we characterize the “maximally flexible” canonical model and the nature of the admissible factor correlations and conditional volatilities that these canonical models can accommodate.

Our empirical analysis of the family of three-factor affine models leads us to the following conclusions about their empirical properties. First, across a wide variety of parameterizations of ATSMs, the data consistently called for negative conditional correlations among the state variables. Such correlations are precluded in multifactor CIR models and, therefore, this finding would not have been directly evident from previous empirical studies of these models.²⁴ Non-zero conditional correlations are also precluded in the affine version of the Chen model, and only limited nonzero correlations were permitted in the BDFS model. The empirical results from the $\mathbb{A}_1(3)$ and $\mathbb{A}_2(3)$ branches suggests that the limited correlation accommodated by these models largely explains the associated large chi-square statistics. The importance of negative correlation may not have been more apparent from previous empirical studies, because many of these studies used data on the short rate alone to estimate multifactor models. We find that a key role of the factor correlations is in explaining the shape of the term structure of volatility of bond yields, and this is revealed most clearly through the simultaneous analysis of long- and short-term bond yields.

Second, within the affine family, we demonstrated an important trade-off between the structure of factor volatilities on the one hand and admissible non-zero conditional correlations of the factors on the other. The “maximally flexible” models in $\mathbb{A}_1(3)$ give the most flexibility in specifying conditional correlations, while still accommodating some time-varying volatility (in this case driven by one factor, $m = 1$). Models in $\mathbb{A}_2(3)$ offer more flexibility in specifying time-varying volatility (as it may depend on two factors, $m = 2$), but admissibility requires a relatively more restrictive correlation structure. For our data set on dollar swap yields and our sample period, we find that flexibility with regard

²³ In a one-factor setting, Ait-Sahalia (1996) finds evidence for nonlinearity in the drifts of short rates, although some recent work such as Chapman and Pearson (1999) suggests that such evidence needs to be interpreted with caution. Boudoukh et al. (1998) also provide evidence for a nonlinear relationship between slope and level in a two-factor setting.

²⁴ As noted in Section I, our finding was implicitly present in studies of CIR models, because the correlations among the implied state variables are strongly negative, in contrast to the implications of the models.

to the specification of correlations is more important than flexibility in specifying volatilities. Overall, the preferred $\mathbb{A}_1(3)_{DS}$ model in $\mathbb{A}_1(3)$ seems to fit better than the preferred model $\mathbb{A}_2(3)_{DS}$ in $\mathbb{A}_2(3)$. We conjecture that, for other markets or different sample periods, where conditional volatility is much more pronounced in the data, the relative goodness-of-fit of models in the branches $\mathbb{A}_1(3)$ and $\mathbb{A}_2(3)$ may change. Indeed, our analysis suggests that the affine family of term structure models will have the most difficulty describing interest rate behavior in settings where conditional volatility is pronounced and the factors are strongly negatively correlated.

Though extant ATSMs have evidently failed to capture key features of historical dollar interest rate behavior, many basic features of these models are supported by our empirical analysis. Specifically, both the $\mathbb{A}_1(3)_{DS}$ and $\mathbb{A}_2(3)_{DS}$ models build on the literature that posits a short-rate process with a stochastic central tendency and volatility. The (implicit) restriction in these three-factor models that the short rate can be expressed as an affine function of only two of the three factors is supported by the empirical evidence, and, hence, this restriction is imposed in the *DS* models. Furthermore, consistent with the analysis of a central tendency factor in Andersen and Lund (1998) and Balduzzi et al. (1996), we find that the short rate tends to mean-revert relatively quickly to a factor that itself has a relatively slow rate of mean reversion to its own constant long-run mean. However, in our preferred model $\mathbb{A}_1(3)_{DS}$, neither of the two factors determining the short rate, besides itself, is literally interpretable as the central tendency of r . Also, even though factors may “look” like central tendency or volatility, the meaning of these constructs in terms of their induced changes in the shapes of yield curves varies substantially across specifications of the factor correlations.

Finally, all of the models examined in this paper presume that the market prices of risk are proportional to the volatilities of the state variables. Two important, and potentially restrictive, implications of this formulation are that the state variables follow affine diffusions under both the actual and risk-neutral probabilities and that the signs of the market prices of risk do not change over time. The evidence suggests that our formulation of risk premiums may underlie the difficulty we found in matching the sample moments of swap yields, particularly with model $\mathbb{A}_2(3)_{DS}$. Nonlinear formulations of the risk premiums (including formulations with time-dependent signs) can be accommodated directly within the affine framework, as long as the state variables follow affine diffusions under the risk-neutral distribution. The empirical significance of such extensions of the affine framework explored here is an interesting topic for future research.

Appendix A. Invariant Transformations

In defining the class of admissible ATSMs, it will be necessary to undertake various transformations and rescalings of the state and parameter vectors in ways that leave the instantaneous short rate, and hence bond prices, unchanged. We refer to such transformations as “invariant transformations.” More precisely, consider an ATSM with state vector $Y(t)$, Brownian motions $W(t)$,

and parameter vector $\phi = (\delta_0, \delta_y, \mathcal{K}, \Theta, \Sigma, \{\alpha_i, \beta_i : 1 \leq i \leq N, \lambda\})$. An *invariant affine transformation* $\mathcal{T}_A$ is defined by an $N \times N$ nonsingular matrix L and an $N \times 1$ vector ϑ , such that $\mathcal{T}_A Y(t) = LY(t) + \vartheta$, $\mathcal{T}_A \phi = (\delta_0 - \delta'_y L^{-1} \vartheta, L'^{-1} \delta_y, L \mathcal{K} L^{-1}, \vartheta + L \Theta, L \Sigma, \{\alpha_i - \beta'_i L^{-1} \vartheta, L'^{-1} \beta_i : 1 \leq i \leq N, \lambda\})$ are the state vector and the parameter vector, respectively, under the transformed model. The Brownian motions are not affected. Such transformations are generally possible, because of the linear structure of ATSMs and the fact that the state variables are not observed. A *diffusion rescaling* $\mathcal{T}_D$ rescales the parameters of $[S(t)]_{ii}$ and the i th entry of λ by the same constant. That is, for any $N \times N$ nonsingular matrix D , $\mathcal{T}_D \phi = (\delta_0, \delta_y, \mathcal{K}, \Theta, \Sigma D^{-1}, \{D_{ii}^2 \alpha_i, D_{ii}^2 \beta_i : 1 \leq i \leq N, D \lambda\})$ is the parameter vector for the transformed model. The state vector and the Brownian motions are not affected. Such rescalings may be possible, because only the combinations $\Sigma S(t) \Sigma'$ and $\Sigma S(t) \lambda$ enter the pricing equations (5), (6), and (7). A *Brownian motion rotation* $\mathcal{T}_O$ takes a vector of unobserved, independent Brownian motions and rotates it into another vector of independent Brownian motions. That is, for any $N \times N$ orthogonal matrix O (i.e., $O^{-1} = O^T$) that commutes with $S(t)$, $\mathcal{T}_O W(t) = OW(t)$ and $\mathcal{T}_O \phi = (\delta_0, \delta_y, \mathcal{K}, \Theta, \Sigma O^T, \{\alpha_i, \beta_i : 1 \leq i \leq N, O \lambda\})$ are the Brownian motions and the parameter vector, respectively, for the transformed model. The state vector is not affected. Finally, a *permutation* $\mathcal{T}_P$ simply reorders the state variables, which has no observable consequences. It is easily checked that any two ATSMs linked by any combination of the above invariant transformations are equivalent in the sense that the implied bond prices (including the short rate) and their distributions are exactly the same.

B. Admissibility of the Canonical Model

For an arbitrary affine model, deriving sufficient conditions for admissibility is complicated by the fact that admissibility is a joint property of the drift ($\mathcal{K}$ and Θ) and diffusion (Σ and $\mathcal{B}$) parameters in equation (9). A key motivation for our choice of canonical representations is that we can treat the drift and diffusion coefficients separately in deriving sufficient conditions for admissibility. Therefore, verification of admissibility is typically straightforward. In this Appendix, we provide sufficient conditions for our canonical representation of $\mathbb{A}_m(N)$ to be well defined.

The canonical representation of $\mathbb{A}_m(N)$ has the conditional variances of the state variables controlled by the first m state variables:

$$S_{ii}(t) = Y_i(t), \quad 1 \leq i \leq m, \quad (\text{A1})$$

$$S_{jj}(t) = \alpha_j + \sum_{k=1}^m [\beta_j]_k Y_k(t), \quad m+1 \leq j \leq N, \quad (\text{A2})$$

where $\alpha_j \geq 0$, $[\beta_j]_i \geq 0$.²⁵ Therefore, as long as $Y^B(t) \equiv (Y_1, Y_2, \dots, Y_m)'$ is non-negative with probability one, the canonical representation of $Y(t) = (Y^B(t), Y^{D'}(t))'$, where $Y^D(t) \equiv (Y_{m+1}, Y_{m+2}, \dots, Y_N)$, will be admissible.

²⁵ Any model within $\mathbb{A}_m(N)$ can be transformed to an equivalent model with this volatility structure through an invariant transformation.

In general, Y^B follows the diffusion

$$dY^B(t) = \mathcal{K}^B(\Theta - Y(t))dt + \Sigma^B \sqrt{S(t)} dW(t). \quad (\text{A3})$$

To assure that $Y^B(t)$ is bounded at zero from below, the drift of $Y^B(t)$ must be non-negative, and its diffusion must vanish at the zero boundary. Sufficient conditions for this are

$$\begin{aligned} \text{C1: } \mathcal{K}^{\text{BD}} &= 0_{m \times (N-m)}, & \text{C2: } \Sigma^{\text{BD}} &= 0_{m \times (N-m)}, \\ \text{C3: } \Sigma_{ij} &= 0, \quad 1 \leq i \neq j \leq m, & \text{C4: } \mathcal{K}_{ij} &\leq 0, \quad 1 \leq i \neq j \leq m, \\ \text{C5: } \mathcal{K}^{\text{BB}} \Theta^B &> 0. \end{aligned}$$

Condition C1 is imposed because otherwise there would be a positive probability that the drift of Y^B at the zero boundary becomes negative (because $Y^D(t)$ is not bounded from below). Conditions C2 and C3 are imposed to prevent $Y^B(t)$ from diffusing across zero due to nonzero correlation between $Y^B(t)$ and $Y^D(t)$. Condition C4 (same as equation (17)) is imposed because otherwise, with $Y^B \geq 0$, there is a positive probability that large values of $Y_j(t)$ will induce a negative drift in $Y_i(t)$ at its zero boundary, for $1 \leq i \neq j \leq m$. Together, conditions C4 and C5 assure that the drift condition

$$\mathcal{K}_{ii} \Theta_i + \sum_{j=1; j \neq i}^m \mathcal{K}_{ij} (\Theta_j - Y_j(t)) \geq 0 \quad (\text{A4})$$

holds for all i , $1 \leq i \leq m$.

Under conditions C1–C5 the existence of an (almost surely) non-negative and nonexplosive solution to our canonical representation in equation (9) is assured because its drift and diffusion functions are continuous and satisfy a growth condition (see Ikeda and Watanabe (1981), Chap. IV, Theorem 2.4). The uniqueness of the solution is assured because the drift satisfies a Lipschitz condition and the diffusion function satisfies the Yamada condition (see Yamada and Watanabe (1971), Theorem 1.²⁶ The state space for the solution is $\mathcal{R}_+^m \otimes \mathcal{R}^{N-m}$.

Finally, condition C5 implies that the zero boundary of Y^B is at least reflecting. This is because, under conditions C1–C3, the subvector $Y^B(t)$ is an *autonomous* multivariate correlated square-root process governed by

$$dY^B(t) = \mathcal{K}^{\text{BB}}(\Theta^B - Y^B(t))dt + \sqrt{S^{\text{BB}}(t)} dW^B(t). \quad (\text{A5})$$

If the off-diagonal elements of $\mathcal{K}^{\text{BB}}$ are zero, then equation (A5) represents an m -dimensional independent square-root process. That the zero boundary is reflecting is trivial in this case. Under condition C4, the drift of the correlated square-root process dominates that of the independent square-root

²⁶ To appeal to Yamada and Watanabe (1971), we note that, without loss of generality, Σ may be normalized to the identity matrix (see Appendix C). This normalization is imposed in our canonical model.

process. By appealing to Lemma A.3 of Duffie and Kan (1996), we conclude that the zero-boundary for the correlated square-root process is at least reflecting.²⁷

C. Normalizations on the Canonical Representation

Besides these restrictions to assure admissibility, normalizations must be imposed to achieve econometric identification. The normalizations imposed on the canonical representation of branch $\mathbb{A}_m(N)$ are as follows.

Scale of the State Variables $\mathcal{B}_{ii} = 1$, $1 \leq i \leq m$, $\alpha_i = 1$, $m + 1 \leq i \leq N$, and $\Sigma_{ii} = 1$, $1 \leq i \leq N$. Fixing the scale of $Y(t)$ in this way allows δ_y to be treated as a free parameter vector.

Level of the State Vector $\alpha_i = 0$, $1 \leq i \leq m$, $\Theta_i = 0$, $m + 1 \leq i \leq N$. Fixing the level of the state vector in this way allows δ_0 and Θ^B to be treated as free parameters.

Inter-dependencies of the State Variables Three considerations arise:

- (i) The upper-diagonal blocks of $\mathcal{K}$, Σ , and $\mathcal{B}$, which control the inter-dependencies among the elements of Y^B , are not separately identified; the upper-diagonal block of $\mathcal{B}$ is normalized to be diagonal.
- (ii) The lower-diagonal blocks of $\mathcal{K}$ and Σ , which determine the inter-dependencies among the elements of Y^D , are not separately identified, the lower-diagonal block of Σ is normalized to be diagonal.
- (iii) The lower-left blocks of $\mathcal{K}$ and Σ , which determine the interdependencies between the elements of Y^B and Y^D , are not separately identified. We are free to normalize either $\mathcal{K}^{DB}$ or Σ^{DB} to zero. We choose to set $\Sigma^{DB} = 0$ in our canonical representation.²⁸

Signs The signs of δ_y and $Y(t)$ are indeterminate if $\mathcal{B}$ is free. Normalizing the diagonal elements of the upper-diagonal block of $\mathcal{B}$ to 1 has the effect of fixing the sign of Y^B , and consequently Θ_i and δ_i , $1 \leq i \leq m$. The sign of Y^D is determined once we impose the normalization that $\delta_i \geq 0$, $m + 1 \leq i \leq N$.

Brownian Motion Rotations For the case of $m = 0$, not all elements of $\mathcal{K}$ are identified. An orthogonal transformation can make $\mathcal{K}$ either upper or lower triangular. Also, even in cases with $m \neq 0$, if S_{ii} and S_{jj} are proportional for $i \neq j$, then the parameters $\mathcal{K}_{ij}$ and $\mathcal{K}_{ji}$ are not separately identified. One of them may be normalized to zero.

²⁷ Condition C5 may be replaced by the stronger condition $\mathcal{K}^{BB}\Theta^B \geq 1/2$, as in Duffie and Kan (1996). The stronger condition, under which the zero boundary for Y^B is entrance, is the multivariate generalization of the Feller condition.

²⁸ Starting from a model with nonzero Σ^{DB} , the affine transformation with

$$L = \begin{bmatrix} I_{m \times m} & 0_{m \times (N-m)} \\ -\Sigma_{(N-m) \times m}^{DB} & I_{(N-m) \times (N-m)} \end{bmatrix}$$

transforms the model to an equivalent model with $\Sigma^{DB} = 0_{(N-m) \times m}$.

D. Generating $\mathbb{A}_m(\mathbf{N})$

Starting with the canonical representation of $\mathbb{A}_m(N)$, we can generate an infinite number of equivalent “maximal” models by application of invariant transformations with coefficients that are either known constants or functions of the parameters of the canonical model. Thus, the canonical representation is the basis for an equivalence class of maximal ATSMs. And we can alternatively define $\mathbb{A}_m(N)$ as the set of admissible models that are econometrically nested within one of the maximal models in this equivalence class.

E. Alternative Representations

The following subsections derive the equivalent AY and Ar representations of the $AM_1(3)$ and $AM_2(3)$ models discussed in Section I.

E.1. Equivalent Representations of $AM_1(3)$ Models

As mentioned in Section I.B, equations (24) through (26) are an equivalent AY representation of the canonical representation.

The Ar form of this maximal model is obtained by the following steps. Starting from the model in equations (24) through (26), we apply the affine transformation $(\mathcal{T}_A:(L, \vartheta))$ with

$$L = \begin{bmatrix} [\beta_3]_1(1 + \sigma_{23})^2 & 0 & 0 \\ 0 & q & 0 \\ \delta_1 & 1 & 1 \end{bmatrix}, \quad \vartheta = \begin{pmatrix} 0 \\ \delta_0 + \delta_1 \theta_1 \\ \delta_0 \end{pmatrix}, \quad (\text{A6})$$

and the diffusion rescaling $(\mathcal{T}_D:D)$ with $\text{diag}(D) = ([\beta_3]_1(1 + \sigma_{23})^2, q, 1 + \sigma_{23})'$, where $q = (\kappa_{33} - \kappa_{22})/\kappa_{33}$. Then, relabeling the new state variables as $v(t)$, $\theta(t)$ and $r(t)$, respectively, and redefining the free parameters, we obtain the model in equation (23), where

$$\begin{aligned} (\mu, \nu, \kappa) &= (\kappa_{11}, \kappa_{22}, \kappa_{33}), & (\bar{v}, \bar{\theta}) &= ([\beta_3]_1 \theta_1 (1 + \sigma_{23})^2, \delta_0 + \delta_1 \theta_1), \\ (\eta, \zeta) &= (\sqrt{[\beta_3]_1(1 + \sigma_{23})^2}, \sqrt{q^2 \alpha_2}), & \sigma_{r\theta} &= (1 + \sigma_{32})/q, \\ \kappa_{rv} &= \delta_1(\kappa_{11} - \kappa_{33})/[\beta_3]_1/(1 + \sigma_{23})^2, & \sigma_{\theta v} &= q\sigma_{21}/[\beta_3]_1/(1 + \sigma_{23})^2, \\ \sigma_{rv} &= (\delta_1 + \sigma_{21} + \sigma_{31})/[\beta_3]_1/(1 + \sigma_{23})^2, & \sigma_{\theta r} &= q\sigma_{23}/(1 + \sigma_{23}), \\ \beta_\theta &= q^2[\beta_2]_1/[\beta_3]_1/(1 + \sigma_{23})^2, & \alpha_r &= \alpha_3(1 + \sigma_{23})^2, \end{aligned}$$

and

$$(\lambda_v, \lambda_\theta, \lambda_r) = ([\beta_3]_1(1 + \sigma_{23})^2 \lambda_1, q \lambda_2, (1 + \sigma_{23}) \lambda_3).$$

Finally, it is easily verified that the constraints on the $AM_1(3)$ canonical model that give the BDFS model are

$$\delta_1 = 0, \quad (\sigma_{21}, \sigma_{23}, \sigma_{32}) = (0, 0, 0), \quad \alpha_3 = 0, \quad \beta_{12} = 0. \quad (A7)$$

E.2. Equivalent Representations of $AM_2(3)$ Models

The *AY* representation in equations (29) through (31) can be transformed into the canonical representation by diagonalizing Σ , normalizing $[\beta_2]_2 = 1$ so that δ_2 is free, normalizing $\alpha_3 = 1$ so that δ_3 is free, and normalizing $[\beta_1]_1 = 1$ so that $[\beta_3]_1$ is free.

To transform this *AY* model into its *Ar* representation we apply the affine transformation

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 0 & q & 0 \\ \delta_1 & 1 & 1 \end{bmatrix}, \quad \vartheta = \begin{pmatrix} 0 \\ 0 \\ \delta_0 \end{pmatrix}, \quad (A8)$$

coupled with a diffusion rescaling that sets the diagonal elements of Σ to 1: $\text{diag}(D) = (1, q, 1)'$, where $q = (\kappa_{33} - \kappa_{22} - \delta_1 \kappa_{12})/\kappa_{33}$. The resulting *Ar* model is given by equation (20), where

$$\begin{aligned} (\mu, \nu, \kappa) &= (\kappa_{11}, \kappa_{22}, \kappa_{33}), & (\bar{\nu}, \bar{\theta}) &= (\theta_1, q\theta_2), \\ (\eta, \zeta) &= (\sqrt{[\beta_1]_1}, \sqrt{q[\beta_2]_2}), & \bar{r} &= \delta_0 + \delta_1 \theta_1 + \theta_2, \\ \kappa_{v\theta} &= \kappa_{12}/q, & \kappa_{\theta v} &= q\kappa_{21}, \\ \kappa_{rv} &= \kappa_{21} + \delta_1(\kappa_{11} - \kappa_{33}), & \sigma_{rv} &= \sigma_{31} + \delta_1, \\ \sigma_{r\theta} &= (1 + \sigma_{32})/q, & \alpha_r &= \alpha_3, \\ \beta_\theta &= [\beta_3]_2/q, & (\lambda_v, \lambda_\theta, \lambda_r) &= (\lambda_1, q\lambda_2, \lambda_3), \end{aligned} \quad (A9)$$

and

$$\kappa_{rv} = \kappa_{21} + \delta_1(\kappa_{11} - \kappa_{33}), \quad \sigma_{rv} = (\sigma_{31} + \delta_1).$$

To transform this *AY* model in $\mathbb{A}_2(3)$ to a sensible *Ar* model, we must require that q be positive. This is because if q is negative, then the short rate would be mean reverting to a central tendency factor that is the negative of a CIR process. This does not make sense. Suppose $\delta_1 = 0$, as in the Chen model; then $q > 0$ implies that $\kappa_{33} > \kappa_{22}$, so the central tendency has a slower mean reversion than the volatility factor, which makes sense. A model with $q < 0$ can not nest the Chen model. The requirement that a more general model nest the Chen model puts an implicit restriction on how general the nesting

model can be. This creates a possibility that the most general model estimated from the data may not nest the Chen model (i.e., the maximal model may have the property that $q < 0$).

REFERENCES

- Ait-Sahalia, Yacine, 1996, Testing continuous-time models of the spot interest rate, *Review of Financial Studies* 9, 385–426.
- Andersen, Torben G., and Jesper Lund, 1997, Estimating continuous-time stochastic volatility models of the short-term interest rate, *Journal of Econometrics* 72, 343–377.
- Andersen, Torben G., and Jesper Lund, 1998, Stochastic volatility and mean drift in the short term interest rate diffusion: Sources of steepness, level and curvature in the yield curve, Working paper, Northwestern University.
- Balduzzi, Pierluigi, Sanjiv R. Das, Silverio Foresi, and Rangarajan K. Sundaram, 1996, A simple approach to three factor affine term structure models, *Journal of Fixed Income* 6, 43–53.
- Boudoukh, Jacob, Matthew Richardson, Richard Stanton, and Robert F. Whitelaw, 1998, The stochastic behavior of interest rates: Implications from a multifactor, nonlinear continuous-time model, Working paper, New York University.
- Brenner, Robin J., Richard H. Harjes, and Kenneth F. Kroner, 1996, Another look at models of the short term interest rate, *Journal of Financial and Quantitative Analysis* 31, 85–107.
- Chan, K. C., G. Andrew Karolyi, Francis A. Longstaff, and Anthony B. Sanders, 1992, An empirical comparison of alternative models of the short-term interest rate, *Journal of Finance* 47, 1209–1227.
- Chapman, David A., and Neil D. Pearson, 2000, Is the short rate drift actually nonlinear? *Journal of Finance* 55, 355–388.
- Chen, Lin, 1996, *Stochastic Mean and Stochastic Volatility—A Three-Factor Model of the Term Structure of Interest Rates and Its Application to the Pricing of Interest Rate Derivatives* (Blackwell Publishers, Oxford, U.K.).
- Chen, Ren-RAW, and Louis Scott, 1993, Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates, *Journal of Fixed Income* 3, 14–31.
- Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53, 385–408.
- Duffee, Greg R., 1998, Forecasting future interest rates: Are affine models failures? Working paper, Federal Reserve Board.
- Duffie, J. Darrell, and Rui Kan, 1996, A yield-factor model of interest rates, *Mathematical Finance* 6, 379–406.
- Duffie, J. Darrell, and Kenneth J. Singleton, 1993, Simulated moments estimation of Markov models of asset prices, *Econometrica* 61, 929–952.
- Duffie, J. Darrell, and Kenneth J. Singleton, 1997, An econometric model of the term structure of interest rate swap yields, *Journal of Finance* 52, 1287–1321.
- Fisher, Mark, and Christian Gilles, 1996, Estimating exponential affine models of the term structure, Working Paper, Federal Reserve Board, Washington, D.C.
- Gallant, A. Ronald, and Jonathan R. Long, 1997, Estimating stochastic differential equations efficiently by minimum chi-square, *Biometrika* 84, 125–141.
- Gallant, A. Ronald, and George Tauchen, 1996, Which moments to match? *Econometric Theory* 12, 657–681.
- Gallant, A. Ronald, and George Tauchen, 1998, Reprojecting partially observed systems with application to interest rate diffusions, *Journal of the American Statistical Association* 93, 10–24.
- Hansen, Lars, 1982, Large sample properties of generalized method of moment estimators, *Econometrica* 50, 1029–1056.
- Ikedu, Nobuyuki, and Shinzo Watanabe, 1981, *Stochastic Differential Equations and Diffusion Processes* (North Holland, Amsterdam).
- Jegadeesh, Narasimhan, and George G. Pennacchi, 1996, The behavior of interest rates implied by the term structure of eurodollar futures, *Journal of Money, Credit, and Banking* 28, 426–446.

- Koedijk, Kees G., Francois G. J. A. Nissen, Peter C. Schotman, and Christian C. P. Wolff, 1997, The dynamics of short term interest rate volatility reconsidered, *European Finance Review* 1, 105–130.
- Langetieg, Terence C., 1980, A multivariate model of the term structure, *Journal of Finance* 35, 71–97.
- Litterman, Robert, and Jose Scheinkman, 1991, Common factors affecting bond returns, *Journal of Fixed Income* 1, 54–61.
- Pearson, Neil D., and Tong-Sheng Sun, 1994, Exploiting the conditional density in estimating the term structure: An application to the Cox, Ingersoll, and Ross model, *Journal of Finance* 49, 1278–1304.
- Vasicek, Oldrich A., 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* 5, 177–188.
- Yamada, Toshio, and Shinzo Watanabe, 1971, On the uniqueness of solutions of stochastic differential equations, *Journal of Mathematics of Kyoto University* 11, 155–167.