NEW YORK UNIVERSITY
NYU STERN
LEONARD N. STERN SCHOOL OF BUSINESS

Department of Finance
Working Paper Series 1998

FIN-98-083

Specification Analysis of Affine Term Structure Models

Qiang Dai, Kenneth J. Singleton

October 27, 1998

This Working Paper Series has been generously supported by a grant from

CDC Asset Management
AMERICAS

Specification Analysis of Affine Term Structure Models [1]

Qiang Dai
New York University
and
Kenneth J. Singleton
Stanford University and NBER

First Version: November 15, 1996
Current Version: October 27, 1998

[1]We would like to thank Darrell Duffie for extensive discussions, Ron Gallant and Jun Liu for helpful comments, 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, and other places where the paper is presented, and comments from the editor and two referees. Authors' email addresses: qdai@stern.nyu.edu for Dai and ken@future.stanford.edu for Singleton.

Abstract

In this paper, we explore the features of affine term structure models that are empirically important for explaining the joint distribution of yields on short- and long-term interest rate swaps. We begin by showing that the family of N-factor affine models can be classified into $N+1$ non-nested sub-families of models. For each sub-family, we derive a maximal model with the property that every admissible member of this family is equivalent to or a nested special case of our maximal model. Second, using our classification scheme and maximal models, we show that many of the three-factor models in the literature impose potentially strong over-identifying restrictions on the joint distribution of short- and long-term rates. Third, we compute simulated-method-of-moments estimates for several members of one of the four branches of three-factor models, and test the over-identifying restrictions implied by these models. We conclude that many of the extant affine models in the literature fail to describe important features of the distribution of long- and short- term rates. The source of the model misspecification is shown to be overly strong restrictions on the correlations among the state variables. Relaxing these restrictions leads to a model that passes several goodness-of-fit tests over our sample period.

I Introduction

Recently, considerable attention has been focused on the "affine" class of term structure models (*ATSMs*) in which the drifts and volatility coefficients of the state-variable processes are affine functions of the underlying state vector (e.g., Duffie and Kan [13]). *ATSMs* accommodate potentially rich term-structure dynamics because, in multi-factor models, the conditional variance of each factor can be a positive affine function of all of the factors, the shocks driving the factors may be correlated, and there may be affine dependencies among the factors through their drifts. However, both theoretical and empirical studies of affine models have focused exclusively on seemingly *very* special cases. For instance, Chen and Scott [10], Pearson and Sun [25], and Duffie and Singleton [15] assume that the short rate is a sum of a vector of independent, univariate square-root diffusions. Alternatively, the models of Chen [9] and Balduzzi, Das, Foresi and Sundaram [7], in which the short rate itself is a state variable, assume zero correlations among some of the shocks, and impose strong restrictions on the dependencies among the factors through their drifts and conditional volatilities. Therefore, we are led to inquire:

Q1 Are these special cases restrictive, or are they the most flexible specifications of *ATSMs* that yield well-defined bond prices?

Q2 If they are restrictive, what are the over-identifying restrictions they impose on yield curve dynamics?

Q3 In their least restrictive forms, are *ATSMs* sufficiently flexible to describe simultaneously the historical movements in short- and long-term bond yields?

In this paper we show that the answer to *Q1* is indeed yes: extant affine term structure models have implicitly imposed potentially strong over-identifying restrictions on the joint distributions of long- and short-term bond yields. To show this, we provide a complete characterization of the admissible, identified multi-factor *ATSMs*. Using the classification scheme of Dai, Liu, and Singleton [11] for general affine diffusions, we propose a convenient classification of N-factor *ATSMs* into $N+1$ non-nested sub-families of models. For each of these $N+1$ sub-families, we derive a *maximal* model with the property that every other well-defined *ATSM* within this sub-family is equivalent to, or a nested special case of, the maximal model. Furthermore, all of the extant *ATSMs* cited above are shown to be restricted special cases of our maximal models. As such, we answer *Q2* by providing a full characterization of the over-identifying restrictions previously imposed.

The reason there is not an all-encompassing *ATSM* that nests all extant models as special cases is that the parameter space must be constrained so that bond prices are well-defined. This *admissibility problem* arises because the volatility of the i^{th} factor, $Y_i(t)$, is given by $\sqrt{\alpha_i + \beta_i' Y(t)}$, and, therefore, $\alpha_i + \beta_i' Y(t)$ must be positive over the range of $Y(t)$ for bond prices to be well defined (Duffie and Kan [13], Dai, Liu, and Singleton [11], Appendix A). Whether or not a parameterization is admissible depends jointly on the characteristics of the drift and diffusion coefficients of the state vector $Y(t)$ and one must typically trade off flexibility in specifying the drift against the richness of the conditional volatilities and correlations of Y. We proceed by classifying the family of N-factor *ATSMs* into $N+1$ sub-families in such a way that sufficient conditions for admissibility are easily verified and the over-identifying restrictions in extant *ATSMs* are easily interpreted. The classification scheme is based on the number of the factors (m) that determine the volatilities of all N factors.

Having classified the admissible N-factor *ATSMs*, we next specialize to the case of $N = 3$ and describe in detail the nature of the 4 maximal models for the 3-factor family of *ATSMs*. From this discussion we see that potentially strong over-identifying restrictions were imposed in most *ATSMs* in the literature, the primary exception being some Gaussian (Vasicek [26]) models. Moreover, this analysis reveals several new insights into the nature of these restrictions. Specifically, *ATSMs* allow for more interdependencies among the factors through their drifts, without jeopardizing admissibility or identification, than has heretofore been recognized. For instance, we can allow for feedback through the drifts of the stochastic central tendency and volatility factors in the Chen [9] and Balduzzi, Das, Foresi and Sundaram [7] (hereafter *BDFS*) models. Similarly, there is no need to constrain the drifts of the square-root diffusion in *CIR*-style models to be independent across factors – correlated square-root diffusions are not inconsistent with admissibility or our ability to obtain (essentially) closed-form expressions for zero-coupon bond prices. Furthermore, in the cases of the *Chen* and *BDFS* models, several of the zero restrictions on the correlations among the diffusions can be relaxed. These observations lead to new, and as yet unexplored, *ATSMs*.

Our discussion of admissibility highlights an important trade-off between the generality of the dependence of the conditional variance of each $Y_i(t)$

on $Y(t)$ (within the affine framework) and the admissible structure of the correlations among these state variables. *Gaussian* models offer complete flexibility with regard to the signs and magnitudes of conditional and unconditional correlations among the Ys, 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 the correlated square-root diffusion (*CSR*) model that has all three state variables driving conditional volatilities ($m = 3$). However, admissibility of this model requires that the conditional correlations of the state variables be zero and that their unconditional correlations be non-negative. We conclude that *CSR* models are theoretically incapable of generating the negative correlations among the state variables that the historical data seems to call for. In between the *Gaussian* and *CSR* models lie $N - 1$ families of *ATSM*s with time-varying conditional volatilities of the state variables and unconstrained signs of (some of) their correlations.

In light of these observations, to address *Q3*, we consider the case of $N = 3$ and focus on the two sub-families of *ATSM*s 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 sub-family $m = 1$ nests the *BDFS* model, while the maximal model for the $m = 2$ sub-family nests the *Chen* model. In both of these special cases, the instantaneous short rate has a stochastic central tendency and stochastic volatility. We show that our maximal models differ from these extant specifications by accommodating much richer interdependencies among the state variables through both their drift and diffusion coefficients.

We compute simulated method of moments (*SMM*) estimates (Duffie and Singleton [14] and Gallant and Tauchen [18]) of our maximal *ATSM*s using short-, intermediate-, and long-term swap yields simultaneously. The models pass several formal goodness-of-fit tests. Moreover, the restrictions implicit in the *Chen* model and *BDFS* are strongly rejected. The substantial improvement in goodness-of-fit for the canonical model is traced directly to its richer parameterization of correlations among the state variables. The *negatively* correlated diffusions are central to the model's ability to match the volatility structure of swap rates. Finally, we analyze additional properties of the maximal models and their nested special cases, and conclude that (a) the maximal models are over-parameterized; (b) within the affine class of models, the best description of the swap data is given by an intermediate model in the branch $m = 1$, which extends the *BDFS* model by allowing conditional correlations between the short rate and its stochastic central tendency.

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

II The Affine Bond Pricing Model

Consider a frictionless economy with riskless borrowing and lending opportunities. Fix a standard Brownian motion $W = (W_1, W_2, \ldots, W_N)$ in $\mathbb{R}^N$ restricted to some time interval $[0, T]$ on a given probability space $(\Omega, \mathcal{F}, P)$. We also fix the standard filtration $\mathbb{F} = \{\mathcal{F}_t : t \in [0, T]\}$ of W, and let $\mathcal{F} = \mathcal{F}_T$. Assume that: (a) the prices of M bonds follow the Ito process $X = (X_1, X_2, \ldots, X_M)$ in $\mathbb{R}^M$,

$$dX(t) = \mu_X(t)dt + \sigma_X(t)dW(t), \tag{1}$$

where $\sigma_X(t)$ is an $M \times N$ matrix; (b) the instantaneous short rate process $r(t)$ is measurable with respect to $\mathcal{F}_t$; and (c) there are no arbitrage opportunities. Then, under further technical conditions (see Duffie [12] and Hansen and Richard [20]), there exists a state price deflator $\pi(t)$, such that $\pi(t)X(t)$ is a martingale under P; i.e., for any time t and $s > t$,

$$X(t) = E_t\left[\frac{\pi(s)}{\pi(t)}X(s)\right]. \tag{2}$$

The ratio $\frac{\pi(s)}{\pi(t)}$ is the stochastic discount factor or pricing kernel for pricing the M securities in the absence of arbitrage. By Ito's lemma, it can be shown that the pricing kernel satisfies

$$\frac{d\pi(t)}{\pi(t)} = -r(t)dt - \Lambda(t)'dW(t), \tag{3}$$

where $\sigma_X(t)\Lambda(t) = \mu_X(t) - r(t)X(t)$.

The preceding characterization of the pricing kernel process $\pi(t)$ for pricing bonds requires little more than the absence of arbitrage opportunities. The general affine term structure model is obtained by imposing the additional assumptions that

$$r(t) = \delta_0 + \sum_{i=1}^{N} \delta_i Y_i(t) \equiv \delta_0 + \delta_y' Y(t) \tag{4}$$

and

$$\Lambda(t) = \sqrt{S(t)}\lambda, \tag{5}$$

where, $\delta_y = (\delta_1, \ldots, \delta_N)'$, and $\lambda = (\lambda_1, \ldots, \lambda_N)'$ are N-vectors of constants. The state variables $Y_i(t)$, $i = 1, 2, \ldots, N$, are assumed to follow the N-dimensional stochastic process

$$dY(t) = \mathcal{K}\left(\Theta - Y(t)\right) dt + \Sigma\sqrt{S(t)}dW(t), \tag{6}$$

where $Y(t) = (Y_1(t), Y_2(t), \cdots, Y_N(t))'$, $\mathcal{K}$ and Σ are $N \times N$ matrices, which may be non-diagonal and asymmetric. $S(t)$ in (5) and (6) is a diagonal matrix with the i^{th} diagonal element given by

$$[S(t)]_{ii} = \alpha_i + \beta_i' Y(t). \tag{7}$$

This characterization of the affine term structure model is the continuous-time, affine counterpart to the formulations of the pricing kernels in Backus and Zin [4] and Backus, Foresi, and Telmer [3]. Our formulation generalizes the continuous-time pricing kernels assumed by Bakshi and Chen [5] and Nielsen and Saá-Requejo [24], and is equivalent to that of Fisher and Gilles [16]. Thus, the subsequent analysis of the affine term structure models applies to all of these frameworks. Of course, it also applies to equilibrium term structure models that lead to pricing kernels with this affine structure such as the CIR model.

The time t price $P(t,\tau)$ for a zero-coupon bond with maturity τ is given by setting $X(t+\tau) = 1$ in (2):

$$P(t,\tau) = E_t\left[\frac{\pi(t+\tau)}{\pi(t)}\right], \tag{8}$$

which, by the Girsanov theorem, is equivalent to

$$P(t,\tau) = E_t^Q\left[e^{-\int_t^{t+\tau} r(u)du}\right], \tag{9}$$

where $E_t^Q[\cdot] = E^Q[\cdot|\mathcal{F}_t]$ is the expectation with respect to the "risk-neutral" measure Q conditional on the filtration at time t. The dynamics of the state variables under Q, which is needed in order to evaluate bond prices using (9), is given by

$$dY(t) = \tilde{\mathcal{K}}\left(\tilde{\theta} - Y(t)\right) dt + \Sigma\sqrt{S(t)}d\tilde{W}(t), \tag{10}$$

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

The risk-neutral drift $\mu(t)$ and diffusion $\sigma(t)$ of $Y(t)$ have the feature that both $\mu(t)$ and $\sigma(t)\sigma(t)'$ are affine functions of $Y(t)$. This assures that the zero coupon bond prices are log linear in the state vector $Y(t)$.[1] Specifically, it can be shown [see Duffie and Kan [13]] that the zero-coupon bond prices are given by

$$P(t,\tau) = e^{A(\tau)-B(\tau)'Y(t)}, \tag{11}$$

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

$$\frac{dA(\tau)}{d\tau} = -\tilde{\theta}'\tilde{\mathcal{K}}'B(\tau) + \frac{1}{2}\sum_{i=1}^{N}\left[\Sigma' B(\tau)\right]_i^2 \alpha_i - \delta_0, \tag{12}$$

$$\frac{dB(\tau)}{d\tau} = -\tilde{\mathcal{K}}'B(\tau) - \frac{1}{2}\sum_{i=1}^{N}\left[\Sigma' B(\tau)\right]_i^2 \beta_i + \delta_y. \tag{13}$$

These ODEs can be solved easily through numerical integration, starting from the initial conditions: $A(0) = 0$, $B(0) = 0_{N\times 1}$. Consequently, estimation of models that simultaneously price long- and short-term rates is computationally feasible.

Equations (4) - (9) characterize what we will refer to as the general AY representation of a multi-factor, affine bond pricing model.[2]

[1] Our specification of the state variable dynamics under the actual measure is also affine [see (6)]. This is not necessary for the log linearity of zero coupon bond prices, which only requires that the risk-neutral dynamics of the state variables be affine.

[2] There is a different formulation of affine models in the literature that starts with the diffusion model for $r(t)$ and adds state variables by allowing the drift and diffusion coefficients of r to depend on unobserved state variables (e.g., Balduzzi, Das and Foresi [6] and Chen [9]). See Section III.B for a proof that this alternative approach produces a model that is analytically equivalent to a member of the class of affine models examined here.

III A Characterization of Admissible *ATSM*s

Ideally, a specification analysis could be conducted with the general affine term structure specification (6). This is not possible, however, because, for an arbitrary choice of the parameter vector $\psi \equiv (\mathcal{K}, \Theta, \Sigma, \mathcal{B}, \alpha)$, the conditional variances $[S(t)]_{ii}$ may not be positive over the range of Y. To assure positive variances– what we will refer to as admissibility– it appears necessary to trade off flexibility in specifying the drift parameters ($\mathcal{K}$ and Θ) against generality in the diffusion coefficients (Σ and $\mathcal{B}$). Accordingly, using results in Dai, Liu, and Singleton [11], we proceed by classifying N-factor *ATSM*s into $N+1$ classes for which there are intuitive and easily verified sufficient conditions for admissibility. Then we specialize to the case of $N = 3$ and use our classification scheme to interpret the over-identifying restrictions imposed in extant *ATSM*s. This section focuses on the practical implications of our characterization of admissible *ATSM*s. Formalities are presented in Appendix A.

III.A A Canonical Representation of Admissible *ATSM*s

The admissibility problem is intimately related to the presence of stochastic volatility in the state process: there is no admissibility problem if $\beta_i = 0$ and the requirements for admissibility become increasingly stringent as the number of state variables determining $[S(t)]_{ii}$ increases. More formally, we let $\mathcal{B} \equiv (\beta_1, \ldots, \beta_N)$ denote the matrix of coefficients on Y in the $[S(t)]_{ii}$ and introduce the index $m = \text{rank}(\mathcal{B})$ of the degree of dependence of the conditional variances on the number of state variables. We then classify all *ATSM*s with the same m into the same subfamily. This classification of *ATSM*s leads to $N+1$ non-nested and mutually exclusive families with intuitive and easily verifiable sufficient conditions for admissibility for each subfamily or "branch". The branch of admissible *ATSM*s with $\text{rank}(\mathcal{B}) = m$ is denoted by $\mathbb{A}_m(N)$.

To characterize the *ATSM*s in $\mathbb{A}_m(N)$ we introduce the following *canonical* representation of admissible N-factor models with $\text{rank}(\mathcal{B}) = m$:

Definition III.1 (Canonical Representation of $\mathbb{A}_m(N)$) *Partitioning Y as $Y(t)' = (Y^{\mathrm{B}\prime}, Y^{\mathrm{D}\prime})$, where Y^{B} is $m \times 1$ and Y^{D} is $(N-m) \times 1$, the* *canonical representation of* $\mathbb{A}_m(N)$ *is the special case of (6) with*

$$\mathcal{K} = \begin{bmatrix} \mathcal{K}^{\mathrm{BB}}_{m\times m} & 0_{(N-m)\times m} \\ \mathcal{K}^{\mathrm{DB}}_{(N-m)\times m} & \mathcal{K}^{\mathrm{DD}}_{(N-m)\times(N-m)} \end{bmatrix}, \tag{14}$$

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

$$\Theta = \begin{bmatrix} \Theta^{\mathrm{B}}_{m\times 1} \\ 0_{(N-m)\times 1} \end{bmatrix}, \tag{15}$$

$$\Sigma = I, \tag{16}$$

$$\alpha = \begin{bmatrix} 0_{m\times 1} \\ 1_{(N-m)\times 1} \end{bmatrix}, \tag{17}$$

$$\mathcal{B} = \begin{bmatrix} I_{m\times m} & B^{\mathrm{DB}}_{(N-m)\times m} \\ 0_{(N-m)\times m} & 0_{(N-m)\times(N-m)} \end{bmatrix}; \tag{18}$$

with the following parametric restrictions imposed:

$$\delta_i \geq 0,\ \ m+1 \leq i \leq N, \tag{19}$$

$$\mathcal{K}_i\Theta \equiv \sum_{j=1}^{m} \mathcal{K}_{ij}\Theta_j > 0,\ \ 1 \leq i \leq m, \tag{20}$$

$$\mathcal{K}_{ij} \leq 0,\ \ 1 \leq j \leq m,\ \ j \neq i, \tag{21}$$

$$\Theta_i \geq 0,\ \ 1 \leq i \leq m, \tag{22}$$

$$\mathcal{B}_{ij} \geq 0,\ \ 1 \leq i \leq m,\ \ m+1 \leq j \leq N. \tag{23}$$

This representation is canonical, because it serves as a basis for the *ATSM*s that are known to be admissible under our sufficient conditions. More precisely, we show in Appendix A that the model (14)–(23) is *maximally flexible* in the sense that it imposes the known sufficient conditions for admissibility (Dai, Liu, and Singleton [11]) along with minimal normalizations for econometric identification. Starting from this canonical representation, we let $AM_m(N)$ denote the equivalence class of maximally flexible *ATSM*s generated through *invariant* transformations of the state vector. Invariant transformations, which are formally defined in Appendix A, preserve admissibility and identification and leave the short rate (and hence bond prices) unchanged. The set $\mathbb{A}_m(N)$ can now be characterized as the set of all *ATSMs* that are econometrically nested special cases of an *ATSM* in $AM_m(N)$.[3] The representation (14) – (23) 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 (19) – (23) can be imposed in econometric implementations.

To interpret our canonical representation, note that the conditional variance of the i^{th} component of $Y^{\mathrm{B}}(t)$, $Y_i^B(t)$, is $Y_i^B(t)$. This feature, combined with the assumption that the last $N-m$ rows of $\mathcal{B}$ are zero (see (18)), imply that $\text{rank}(\mathcal{B}) = m$. The instantaneous conditional correlations among the $Y^{\mathrm{B}}(t)$ are zero, while the instantaneous correlations among the $Y^{\mathrm{D}}(t)$ are governed by the parameters $\mathcal{B}_{ij}$, because $\Sigma = I$. All of the state variables may be mutually correlated over any finite sampling interval due to feedback through the drift matrix $\mathcal{K}$. Since the conditional covariance matrix of Y depends only on Y^{B} and (23) holds, admissibility is established if $Y^{\mathrm{B}}(t)$ is strictly positive. The positivity of Y^{B} is assured by zero restrictions in the upper right $(N-m) \times m$ block of $\mathcal{K}$ and the constraints (20) and (21).[4] In order to assure that the state process is stationary, we need to impose an additional constraint, namely, all of the eigen-values of $\mathcal{K}$ are strictly positive. We impose this condition in our empirical estimation.

For the purposes of interpreting the state variables and the restrictions on their dynamic properties with a particular *ATSMs*, it is sometimes more convenient to work with an equivalent model in $AM_m(N)$. Indeed, 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). In Section III.B we present the equivalent Ar and AY representations of several extant *ATSMs*, as well as

[3]Since 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.

[4]As discussed more formally in Appendix A, the zero-restrictions in (14) – (18) and sign restrictions in (19) – (23) represent a complete implementation of the generalization of the existence condition of Duffie and Kan [13] presented in Dai, Liu, and Singleton [11].

their maximally flexible counterparts.

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

III.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 the nature of the over-identifying 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 4 associated sub-families of admissible *ATSMs*.

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

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

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

$$\Theta = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \alpha = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathcal{B} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

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

Gaussian *ATSMs* were studied theoretically by Vasicek [26] and Langetieg [23], among many others. A recent empirical implementation of a 2-factor Gaussian model is Jegadeesh and Pennacchi [22].

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

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

$$\begin{aligned} dv(t) &= \mu(\bar{v} - v(t))dt + \eta\sqrt{v(t)}dB_v(t), \\ d\theta(t) &= \nu(\bar{\theta} - \theta(t))dt + \zeta dB_\theta(t), \\ dr(t) &= \kappa(\theta(t) - r(t))dt + \sqrt{v(t)}d\hat{B}_r(t), \end{aligned} \tag{24}$$

with the only non-zero diffusion correlation being $\text{cov}(dB_v(t), d\hat{B}_r(t)) = \rho_{rv}dt$. Rewriting (25) as

$$dr(t) = \kappa(\theta(t) - r(t))dt + \sqrt{v(t)}dB_r(t) + \sigma_{rv}\sqrt{v(t)}dB_v(t), \tag{25}$$

where $\sigma_{rv} = \rho_{rv}/\eta$, and $B_r(t)$ and $B_v(t)$ are independent, gives the $BDFS$ model in the standard notation for *ATSMs*. 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:

$$\begin{aligned} dv(t) &= \mu(\bar{v} - v(t))dt + \eta\sqrt{v(t)}dB_v(t), \\ d\theta(t) &= \nu(\bar{\theta} - \theta(t))dt + \sqrt{\zeta^2 + \boxed{\beta_\theta}v(t)}dB_\theta(t) \\ &\quad + \boxed{\sigma_{\theta v}}\eta\sqrt{v(t)}dB_v(t) + \boxed{\sigma_{\theta r}}\sqrt{\alpha_r + v(t)}dB_r(t), \\ dr(t) &= \boxed{\kappa_{rv}}(\bar{v} - v(t))dt + \kappa(\theta(t) - r(t))dt + \sqrt{\boxed{\alpha_r} + v(t)}dB_r(t), \\ &\quad + \sigma_{rv}\eta\sqrt{v(t)}dB_v(t) + \boxed{\sigma_{r\theta}}\sqrt{\zeta^2 + \beta_\theta v(t)}dB_\theta(t). \end{aligned} \tag{26}$$

The state variable v is naturally interpreted as a volatility factor, because it determines the conditional variances of all three state variables and, in particular, of r. The state variable θ, on the other hand, affects the drift of the short rate, but not its volatility. The conditional correlation between θ and r (and between v and r) may be nonzero, however.

The $BDFS$ model is the special case of (26) in which the parameters in square boxes are set to zero. Thus, relative to this maximal model, the $BDFS$ model constrains the conditional correlations between r and θ to zero. Additionally, it precludes the volatility shock v from affecting the volatility of the central tendency factor θ. 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*, and one in which θ is perhaps not as naturally interpreted as the central tendency of r. The over-identifying restrictions imposed in the *BDFS* model are examined empirically in Section V.

Though (26) is convenient for interpreting the popular $\mathbb{A}_1(3)$ models, verifying that (26) 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:[5]

$$r(t) = \delta_0 + \boxed{\delta_1} Y_1(t) + Y_2(t) + Y_3(t), \tag{27}$$

$$d\begin{bmatrix} Y_1(t) \\ \hline Y_2(t) \\ Y_3(t) \end{bmatrix} = \left[\begin{array}{c|cc} \kappa_{11} & 0 & 0 \\ \hline 0 & \kappa_{22} & 0 \\ 0 & 0 & \kappa_{33} \end{array}\right] \left[\begin{bmatrix} \theta_1 \\ \hline 0 \\ 0 \end{bmatrix} - \begin{bmatrix} Y_1(t) \\ \hline Y_2(t) \\ Y_3(t) \end{bmatrix} \right] dt$$
$$+ \left[\begin{array}{c|cc} 1 & 0 & 0 \\ \hline \boxed{\sigma_{21}} & 1 & \boxed{\sigma_{23}} \\ \sigma_{31} & \boxed{\sigma_{32}} & 1 \end{array}\right] \sqrt{\left[\begin{array}{c|cc} S_{11}(t) & 0 & 0 \\ \hline 0 & S_{22}(t) & 0 \\ 0 & 0 & S_{33}(t) \end{array}\right]} \, dB(t) \tag{28}$$

where

$$\begin{aligned} S_{11}(t) &= Y_1(t), \\ S_{22}(t) &= \alpha_2 + \boxed{[\beta_2]_1} Y_1(t), \\ S_{33}(t) &= \boxed{\alpha_3} + [\beta_3]_1 Y_1(t). \end{aligned} \tag{29}$$

That this representation is in $AM_1(3)$ follows from its equivalence to our canonical representation of $AM_1(3)$, which is easily shown 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.

[5]The equivalence of the Ar and AY representations is shown in Appendix A.5.1 where the invariant transformations that take us from one representation to the other are given explicitly.

The equivalent AY representation of the $BDFS$ model is obtained from (27) by setting all of the parameters in square boxes to zero in (27), 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 Andersen and Lund [2].

III.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 Ys. A member of this sub-family is the model proposed by Chen [9],

$$\begin{aligned} dv(t) &= \mu(\bar{v} - v(t))dt + \eta\sqrt{v(t)}dW_1(t), \\ d\theta(t) &= \nu(\bar{\theta} - \theta(t))dt + \zeta\sqrt{\theta(t)}dW_2(t), \\ dr(t) &= \kappa(\theta(t) - r(t))dt + \sqrt{v(t)}dW_3(t), \end{aligned} \tag{30}$$

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, while it is Gaussian in the latter.

A convenient maximal model for interpreting the over-identifying restrictions in the *Chen* model is

$$\begin{aligned} dv(t) &= \mu(\bar{v} - v(t))dt + \boxed{\kappa_{v\theta}}(\bar{\theta} - \theta(t))dt + \eta\sqrt{v(t)}\,dW_1(t), \\ d\theta(t) &= \nu(\bar{\theta} - \theta(t))dt + \boxed{\kappa_{\theta v}}(\bar{v} - v(t))dt + \zeta\sqrt{\theta(t)}\,dW_2(t), \\ dr(t) &= \boxed{\kappa_{rv}}(\bar{v} - v(t))dt - \kappa(\bar{\theta} - \theta(t))dt + \kappa(\boxed{\bar{r}} - r(t))dt \\ &\quad + \boxed{\sigma_{rv}}\eta\sqrt{v(t)}\,dW_1(t) + \boxed{\sigma_{r\theta}}\zeta\sqrt{\theta(t)}\,dW_2(t) \\ &\quad + \sqrt{\boxed{\alpha_r} + \boxed{\beta_\theta}\theta(t) + v(t)}\,dW_3(t). \end{aligned} \tag{31}$$

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 V.

Again, we turn to the equivalent AY representation of (31) in order to verify that it is admissible and maximal:

$$r(t) = \boxed{\delta_0} + \boxed{\delta_1} Y_1(t) + Y_2(t) + Y_3(t), \tag{32}$$

$$d\begin{bmatrix} Y_1(t) \\ Y_2(t) \\ \hline Y_3(t) \end{bmatrix} = \left[\begin{array}{cc|c} \kappa_{11} & \boxed{\kappa_{12}} & 0 \\ \boxed{\kappa_{21}} & \kappa_{22} & 0 \\ \hline 0 & 0 & \kappa_{33} \end{array}\right] \left[\begin{bmatrix} \theta_1 \\ \theta_2 \\ \hline 0 \end{bmatrix} - \begin{bmatrix} Y_1(t) \\ Y_2(t) \\ \hline Y_3(t) \end{bmatrix} \right] dt$$

$$+ \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hline \boxed{\sigma_{31}} & \boxed{\sigma_{32}} & 1 \end{array}\right] \sqrt{\left[\begin{array}{cc|c} S_{11}(t) & 0 & 0 \\ 0 & S_{22}(t) & 0 \\ \hline 0 & 0 & S_{33}(t) \end{array}\right]}\, dB(t), \tag{33}$$

where

$$\begin{aligned} S_{11}(t) &= [\beta_1]_1 Y_1(t), \\ S_{22}(t) &= [\beta_2]_2 Y_2(t), \\ S_{33}(t) &= \boxed{\alpha_3} + Y_1(t) + \boxed{[\beta_3]_2} Y_2(t). \end{aligned} \tag{34}$$

In Appendix A.5.2, we show that how this Ar representation is linked to the AY representation, and how the AY representation is linked to the canonical representation for $AM_2(3)$.

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 (32) - (34) implied by the Chen model are obtained by setting the square boxes in (33) to zero except that $\sigma_{32} = -1$ and $\delta_0 = -\delta_1\theta_1 - \theta_2 + q\theta_2 = -\theta_2\kappa_{22}/\kappa_{33}$. Like the $BDFS$ model, in the Chen model r is constrained to be an affine function of only two of the three state variables.

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

The final sub-family of models has $m = 3$ so that all three Ys determine the volatility structure. The canonical representation of $AM_3(3)$ has parameters

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

$$\Theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}, \quad \alpha = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathcal{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

where $\kappa_{ii} > 0$ for $1 \le i \le 3$, $\kappa_{ij} \le 0$ for $1 \le i \ne j \le 3$, $\theta_i > 0$ for $1 \le i \le 3$.

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 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 (21)). Thus, the canonical representation is a correlated, square-root (*CSR*) diffusion model. It follows that the empirical implementations of multi-factor CIR-style models with independent state variables by Chen and Scott [10], Pearson and Sun [25], and Duffie and Singleton [15], among others, have imposed potentially strong over-identifying restrictions by forcing $\mathcal{K}$ to be diagonal. In this three-factor model, a diagonal $\mathcal{K}$ implies six over-identifying restrictions.

III.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 3-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 $\mathbb{A}_1(3)$ models allows two of the three state variables to become negative, etc. 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 be non-negative.

Conditional second moments of zero-coupon bond yields:

The conditional variances of zero-coupon bond yields are affine functions of the state vector $Y(t)$. It follows that bond yields may be conditionally heteroskedastic and that the conditional variances will be determined by m common factors. The Gaussian model ($m=0$) implies that zero yields are homoskedastic, while the *CSR* model ($m=3$) allows all three state variables to induce conditional heteroskedasticity.

However, because parameter restrictions must be imposed to assure admissibility, models with $m=3$ do not necessarily offer more flexibility in parameterizing conditional correlations than models with $m<3$. The nature of the conditional correlations accommodated within $AM_m(3)$ can be seen most easily by normalizing $\mathcal{K}^{\mathrm{DB}}_{(N-m)\times m}$ to zero and $\mathcal{K}^{\mathrm{DD}}_{(N-m)\times(N-m)}$ to a diagonal matrix, and concurrently freeing up Σ^{DB} and the off-diagonal elements of Σ^{DD}. This gives an equivalent model to our canonical representation of $\mathbb{A}_m(N)$ (see Appendix A.3). It follows that the admissibility constraints accommodate non-zero conditional correlations of unconstrained signs between each element of Y^{D} and the entire state vector $Y(t)$.

For instance, with $m=3$, 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 $Y_3(t)$. Thus, in moving from $m=0$ to $m=3$, there is a trade-off between flexibility in specifying the factor structure of the conditional variances and in allowing for non-zero conditional correlation among the factors.

Unconditional correlations among the state variables:

In the Gaussian model ($m=0$), the signs of the non-zero 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 Ys. 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 3-factor *ATSMs* on the two branches $\mathbb{A}_1(3)$ and $\mathbb{A}_2(3)$. Models with $m = 1$ or $m = 2$ allow for 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 admissible correlations among the state variables. In contrast, models in the branch $\mathbb{A}_0(3)$ imply that zero yields are conditionally normal with constant conditional second moments. Though the signs of the unconditional correlations are unrestricted in Gaussian models, conditional correlations are also constants. Finally, models in $\mathbb{A}_3(3)$ do not accommodate negative correlations among the state variables. We will see subsequently that such flexibility is helpful for fitting the historical behavior of swap yields.

IV Simulated Method of Moments Estimation of *ATSMs*

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 [14] and Gallant and Tauchen [18]. Our estimation strategy can be outlined as follows:

(i) select N (LIBOR and swap) yields and a set of moments of these yields to be used in estimation, and choose an initial value for the parameter vector ϕ;

(ii) simulate a long time series of observations on the state vector $Y(t)$ using the chosen value of ϕ; compute the associated time series of model-implied zero-coupon bond prices by solving the Ricatti equations (12) and (13) and substituting these weights into (11); then use the simulated zero-coupon prices to compute the N bond yields;

(iii) compute sample versions of the selected moments using both the actual historical yields and simulated yields, and compute a measure of the distance between them;

(iv) finally, adjust ϕ and then repeat these steps until the historical and simulated moments are made as close to each other as possible.

IV.A The *SMM* Objective Function

A key issue for the SMM estimation strategy is the selection of moments in Step (i). Following Gallant and Tauchen [18], 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, $x_t' = (y_t', y_{t-1}', \ldots, y_{t-\ell}')$, and $f(y_t|x_{t-1}, \gamma)$ denote the conditional density of y associated with the auxiliary description of the yield data. The score of the log-likelihood function evaluated at the maximum likelihood (ML) estimator γ_T with sample size T satisfies

$$\frac{1}{T}\sum_{t=1}^{T} \frac{\partial}{\partial \gamma} \log f(y_t|x_{t-1}, \gamma_T) = 0. \tag{35}$$

Under suitable regularity conditions (see Duffie and Singleton [14] and Gallant and Tauchen [18]), as sample size gets large the sample mean in (35) converges to $E[\partial \log f(y_t|x_{t-1}, \gamma_0)/\partial\gamma] = 0$, where γ_0 is the probability limit of γ_T. It follows that, if the asset pricing model is correctly specified, then the score computed with $\log f(y_t|x_{t-1}, \gamma_T)$ evaluated at $y_t^{\phi_0}$s simulated from the asset pricing model using the true model parameters ϕ_0,

$$\frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} \frac{\partial}{\partial \gamma} \log f(y_t^{\phi_0}|x_{t-1}^{\phi_0}, \gamma_T), \tag{36}$$

where $\mathcal{T}$ is the simulation size, should also be approximately zero. The particular $f(y_t|x_{t-1}, \gamma)$ used is described subsequently.

Having selected the moment conditions, we proceed using the standard *GMM* criterion function (Hansen [19], Duffie and Singleton [14]), a quadratic form in the sample moments (36). The *distance* matrix in this criterion function is chosen optimally to give the most efficient *GMM* estimators based on the moment conditions (36). The requirements for this *SMM* estimator to be consistent for ϕ, beyond the requirement that the auxiliary model have at least as many unknown parameters as the dimension of ϕ, will be met by many descriptive time series models of bond yields. In particular, consistency of the *SMM* estimator does not require that the auxiliary model describe the true joint distribution of the discretely sampled bond yields. Efficiency considerations, on the other hand, lead us to choose an $f(y_t|x_{t-1}, \gamma)$ that might reasonably be assumed to nest the true conditional density of the bond yields y. Specifically, the auxiliary model is constructed using the Semi-Non-Parametric (*SNP*) framework proposed by Gallant and Tauchen [18]. The analysis in Gallant and Long [17] 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.[6] That is, we achieve the efficiency of the maximum likelihood estimators for the true conditional distributions of (discretely sampled) bond yields implied by the *ATSM*s.

Gallant and Tauchen [18] showed that the simulated *SNP* scores, evaluated at the *ML* estimates of γ and the *SMM* estimates ϕ, are asymptotically

[6] More precisely if, for a given order of the polynomial terms in the *SNP* approximation to the density f described subsequently, 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 [16].

normally distributed with zero mean. Thus, individual scores can be tested by forming t–statistics that have a standard normal asymptotic distribution. The minimized value of the GMM criterion function serves as an overall goodness-of-fit statistic with an asymptotic χ^2 distribution and degrees of freedom equal to the difference between the number of SNP parameters and the number of structural parameters.

For our empirical analysis, the observed data y was chosen to be the yields on six-month LIBOR and two-year and ten-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.).

We use the "explicit order-two weak scheme" (see Gallant and Long [17]) to simulate the state variables from the stochastic differential equation governing the state dynamics. The simulated bond prices or yields are then computed from the simulated state variables using (11). We use five subintervals for each week, and take every fifth simulated observation to construct a simulated data set of size 50000.

IV.B The SNP Auxiliary Model

In selecting an SNP approximation to the conditional density of swap yields, we started with a conditional normal distribution for the three bond yields with a linear conditional mean and ARCH specifications of the conditional variances. Then we scaled this conditional normal distribution by polynomial functions of the yields in order to accommodate non-normality of the conditional distribution. We obtained the best SNP model for our dataset through an extensive search along numerous model expansion paths, guided by a model selection criterion, as outlined in Gallant and Tauchen [18]. The resulting SNP model has the following conditional density:

$$f(y_t|x_{t-1},\gamma) = c(x_{t-1}) \; \left[\epsilon_0 + [h(z_t|x_{t-1})]^2\right] \; n(z_t), \tag{37}$$

where $n(\cdot)$ is the density function of the standard normal distribution, ϵ_0 is a small positive number,[7] $h(z|x)$ is a Hermite polynomial in z, $c(x_{t-1})$ is a

[7] Our implementation of SMM with an SNP auxiliary model differs from many previous implementations by our inclusion of the constant ϵ_0 in the SNP density function.

normalization constant, and x_{t-1} is the conditioning set. We let z_t be the normalized version of y_t, defined by

$$z_t = R_{x,t-1}^{-1}(y_t - \mu_{x,t-1}). \tag{38}$$

In the terminology of Gallant and Tauchen [18], the *SNP* 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

$$\mu_{x,t-1} = \begin{pmatrix} \psi_1 + \psi_4\, y_{1,t-1} + \psi_7\, y_{2,t-1} + \psi_{10}\, y_{3,t-1} \\ \psi_2 + \psi_5\, y_{1,t-1} + \psi_8\, y_{2,t-1} + \psi_{11}\, y_{3,t-1} \\ \psi_3 + \psi_6\, y_{1,t-1} + \psi_9\, y_{2,t-1} + \psi_{12}\, y_{3,t-1} \end{pmatrix}. \tag{39}$$

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

$$R_{x,t-1} = \begin{pmatrix} \tau_1 + \tau_7\,|\epsilon_{1,t-1}| + \tau_{25}\,|\epsilon_{1,t-2}| & \tau_2 & \tau_4 \\ 0 & \tau_3 + \tau_{15}\,|\epsilon_{2,t-1}| + \tau_{33}\,|\epsilon_{2,t-2}| & \tau_5 \\ 0 & 0 & \tau_6 + \tau_{24}\,|\epsilon_{3,t-1}| + \tau_{42}\,|\epsilon_{3,t-2}|, \end{pmatrix} \tag{40}$$

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, i.e.,

$$h(z_t|x_{t-1}) = A_1 + \sum_{l=1}^{4}\sum_{i=1}^{3} A_{3(l-1)+1+i}\, z_{i,t}^{l}. \tag{41}$$

Though ϵ_0 is identified if the scale of $h(z|x)$ is fixed, Gallant and Long [17] 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 non-zero constants. With $\epsilon_0 = 0$, we often found that some of the simulated observations were close to the zeros in the density function. In such cases, the *SNP* scores were nearly singular and this, in turn, caused spurious random spikes in the *SMM* objective function. This problem was eliminated by setting $\epsilon_0 = .01$ in our empirical analysis. The estimated parameters of our auxiliary model were essentially unchanged by setting $\epsilon = .01$ instead of at zero.

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.[8]

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

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

IV.C 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 (11) symmetrically with other parameters, and this leads naturally to the question of under what circumstances λ is identified in *ATSM*s. This section argues that λ is generally identified, outside of certain Gaussian models, and clarifies the source of this identification.

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 [19]'s 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

$$D_0 \equiv E\left[\frac{\partial^2 \log f}{\partial\gamma\partial\phi'}(z_t^{\phi_0}, \gamma_0)\right] = E\left[\frac{\partial^2 \log f}{\partial\gamma\partial z_t^{\phi'}}(z_t^{\phi_0}, \gamma_0)\frac{\partial z_t^{\phi_0}}{\partial\phi'}\right] \tag{43}$$

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.

Consider first the case where y_t consists of a single yield on a τ-year zero-coupon bond and $N = 1$. In this case,

$$y_t^{\phi} = a(\tau) + b(\tau)Y_t^{\phi}, \tag{44}$$

where $a(\tau) = -A(\tau)/\tau$ and $b(\tau) = B(\tau)/\tau$. The simulated value of the state vector Y^{ϕ} does not depend on λ, because the simulation is done under the

[8]Note that the term "Homogeneous-Innovation" does not imply that the conditional variances of the yields are constants (because of the "ARCH(2)" terms).

actual probability measure for Y. Thus, for all $ATSMs$, $\partial Y_t^{\phi}/\partial\lambda = 0$. It follows that identification of λ using zero coupon bond yields hinges on the dependence of the $a(\tau)$ and $b(\tau)$ on λ.

In the Gaussian model ($m = 0$), $b(\tau)$ depends only on the scalar $\mathcal{K}$, so that $\partial b(\tau)/\partial\lambda = 0$ and $\partial y_t^{\phi}/\partial\lambda = \partial a(\tau)/\partial\lambda$. Furthermore, $\partial y_t^{\phi}/\partial\delta_0 = \partial a(\tau)/\partial\delta_0$. Both of these derivatives are state-independent. Hence, two of the columns of $\partial z_t^{\phi_0}/\partial\phi'$ are colinear and D_0 has less than full rank. We conclude that the market prices of risk are not identified in Gaussian models estimated using zero-coupon bond yields.

Next, consider the case where $N = 1$ and Y_t follows a square-root diffusion (a one-factor CIR model). Though $\mathcal{K}$ and λ enter $b(\tau)$ symmetrically as $\kappa+\lambda$, these parameters are separately identified. This follows from the observations that $\partial y_t^{\phi}/\partial\lambda$ is state-dependent; and $\partial a/\partial\lambda \neq \partial a/\partial\mathcal{K}$ and $\partial b/\partial\lambda = \partial b/\partial\mathcal{K}$. Furthermore, $\partial Y_t^{\phi}/\partial\mathcal{K}$ is a stochastic process that does not asymptote to a constant, so $\partial z_t^{\phi}/\partial\mathcal{K}$ and $\partial z_t^{\phi}/\partial\lambda$ are not colinear.

Turning to the general case of estimation with N factors and zero-coupon bond yields, consider a model in $AM_m(N)$. If $m < N$, then there are $N-m > 0$ "Gaussian" factors in that, conditional on the first m state variables, the last $N-m$ state variables are Gaussian. Following the logic of the one-factor cases, the λ's associated with the first m factors will generally be identified. Whether the λ's associated with the $N-m$ "Gaussian" factors are identified depends on the correlations among the state variables. If δ_0 is free, then the $N-m$ θ's associated with the "Gaussian" factors are normalized to zero in the canonical representation so, with $N > 1$ and $N-m > 0$, the issue is whether δ_0 and the last $N-m$ λ's are identified. All of these parameters can indeed be indentified. For example, if the volatility of N^{th} state variable depends on the first m state variables, then the $b_i(\tau), 1 \leq i \leq m$ depend on λ_N. Consequently, $\partial y_t^{\phi}/\partial\lambda_N$ is state-dependent. On the other hand, $\partial y_t^{\phi}/\partial\delta_0$ is state-independent, because δ_0 only affects the $a_i(\tau)$. Thus, the derivatives of z_t^{ϕ} with respect to δ_0 and λ_N are not colinear. However, if the β_k are zero, for $m+1 \leq k \leq N$, then the first-order conditions with respect to δ_0 and λ_k will be colinear. Consequently δ_0 and λ_k are not separately identified.

If estimation is based on coupon-bond yields, then an additional source of nonlinearity associated with the nonlinear state-variable to yield mapping is introduced. For the case of swap yields, Duffie and Singleton [15] showed that this nonlinear mapping is given by

$$y_t^{\tau\phi} = \frac{1 - P(t,\tau)}{\sum_{k=1}^{2\tau} P(t, k/2)}, \tag{45}$$

where the $P(t,\tau) = e^{A(\tau) - B(\tau)\cdot Y_t^{\phi}}$ are the (credit-risk adjusted) zero-coupon prices implicit in the swap market. In this case, the partial derivatives $\partial y_t^{\tau\phi}/\partial\lambda$ and $\partial y_t^{\tau\phi}/\partial\delta_0$ are stochastic, and are not proportional to each other. Therefore, for all *ATSM*s, λ and δ_0 are separately identified.

V 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 $\mathbb{A}_1(3)_{BDFS}$ and $\mathbb{A}_2(3)_{Chen}$ respectively, have large chi-square statistics relative to their degrees of freedom. In contrast, the corresponding *maximal* models, denoted by $\mathbb{A}_m(3)_{Max}$, for $m = 1, 2$, are not rejected at conventional significance levels. However, the improved fits of of $\mathbb{A}_1(3)_{Max}$ (compared to $\mathbb{A}_1(3)_{BDFS}$) and $\mathbb{A}_2(3)_{Max}$ (compared to $\mathbb{A}_2(3)_{Chen}$) 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 the two intermediate models, $\mathbb{A}_1(3)_{DS}$ and $\mathbb{A}_2(3)_{DS}$ (the *DS* indicating that these are our preferred models). The *DS* models are not rejected at conventional significance levels, 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.

The key reason that the *DS* models do a better job "explaining" the swap dynamics, in terms of χ^2 statistics, than the $\mathbb{A}_1(3)_{BDFS}$ and $\mathbb{A}_2(3)_{Chen}$ models is that the former allow a more flexible correlation structure of the state variables. In the $\mathbb{A}_1(3)$ branch, the $\mathbb{A}_1(3)_{BDFS}$ model only allows a nonzero conditional correlation between the short rate and its stochastic volatility ($\sigma_{rv} \neq 0$). The $\mathbb{A}_1(3)_{DS}$ 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 (26), that relaxing these constraints affects the diffusion for both $\theta(t)$ and $r(t)$.) In the $\mathbb{A}_2(3)$ branch, the $\mathbb{A}_2(3)_{Chen}$ model assumes that $r(t)$, $\theta(t)$, and $v(t)$ are all pairwise, conditionally uncorrelated. In contrast, the model $\mathbb{A}_2(3)_{DS}$ 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 ($\kappa_{\theta v} \neq 0$).[9]

Moreover, in both branches, it is the introduction of *negative* conditional correlations among the state variables that seems to be important (see the

[9]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 (73) – (74).

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 III). 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.

Figure 2 displays the t-ratios for testing whether the fitted SNP scores of the auxiliary model (see Section IV), computed from the models $\mathbb{A}_1(3)_{BDFS}$, $\mathbb{A}_1(3)_{DS}$, $\mathbb{A}_2(3)_{Chen}$, and $\mathbb{A}_2(3)_{DS}$, 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 $\mathbb{A}_1(3)_{BDFS}$ and $\mathbb{A}_2(3)_{Chen}$ 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 $\mathbb{A}_1(3)_{DS}$, and none are larger than 2 for model $\mathbb{A}_2(3)_{DS}$.

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 $\mathbb{A}_1(3)_{BDFS}$ and $\mathbb{A}_2(3)_{Chen}$ 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 also their persistence and non-normality as well.

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 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 $(.37, .23, 17.4)$ and $(.64, .10, 2.7)$ for models $\mathbb{A}_1(3)_{DS}$ and $\mathbb{A}_2(3)_{DS}$, respectively. As in previous empirical studies (e.g., Balduzzi, Das, Foresi and Sundaram [7] and Andersen and Lund [2]), the "central tendency" factor $\theta(t)$ shows much slower mean reversion (smaller ν) than the rate at which gaps between θ and r are closed in the short rate equation (κ). 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 $\bar{\theta}$.[10] In both DS models, the "volatility" factor $v(t)$ has the fastest rate of mean reversion (μ).

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 like "central tendency" or "volatility" in terms of yield curve movements may not be the same across models. To illustrate this point, consider the models in $\mathbb{A}_1(3)$. In the model $\mathbb{A}_1(3)_{BDFS}$, the correlation between changes in $\theta(t)$ and changes in the ten-year swap rate is .98. The close association between the long-term swap rate and central tendency is intuitive, since $r(t)$ 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 $\mathbb{A}_1(3)_{DS}$. In the latter model, changes in $\theta(t)$ are most highly correlated with changes in the *two-year* swap rate (correlation = .95). This explains the larger value of ν (faster mean reversion of $\theta(t)$) in model $\mathbb{A}_1(3)_{DS}$ than in model $\mathbb{A}_1(3)_{BDFS}$. The conditional distributions of the two- and ten-year swap rates are not the same.[11]

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 non-nested so formal assessments of relative fit are non-trivial. 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 3, 5, and 7 years, none of which were used in estimating the parameters.[12] Model $\mathbb{A}_1(3)_{DS}$ has notably smaller average pricing errors than

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

[11]Similar observations apply to the volatility factor $v(t)$. In both models, $v(t)$ is well proxied by a butterfly position that is long ten-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.

[12]These pricing errors were computed by inverting the models for the implied values

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.

Related to this point, within model $\mathbb{A}_2(3)_{DS}$, the admissibility condition precludes relaxation of the constraint $\sigma_{\theta r} = 0$, because of the 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. Yet, the results from Duffie and Singleton [15] suggest that, in order to explain the two-year and ten-year swap yields, the stochastic central tendency and the stochastic volatility factors should be negatively correlated. More precisely, Duffie and Singleton [15] estimated an *ATSM* of swap yields with two ($N = 2$) independent square-root diffusions as state variables. The implied state variables should be approximately uncorrelated if their model is correctly specified.[13] In fact, the sample correlation between the implied $Y_1(t)$ and $Y_2(t)$ is approximately -.5.

Fourth, 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 (log) yields. The other lines display the sample variances computed by simulating long time series of swap yields, using the estimated parameter values, and then computing sample standard deviations with the simulated data.[14] Notably, the term structure of historical sample volatilities is hump-shaped,

of the state variables, using the six-month and two- and ten-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.

[13] Duffie and Singleton [15] assumed that the two- and ten-year swap yields were priced perfectly by their two-factor model. Thus, using their pricing model evaluated at the maximum likelihood estimates of the parameters, implied state variables were computed as functions of these two swap yields.

[14] 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 the population values.

with a peak around two years. Hump-shaped volatility curves can be induced in *ATSMs* either through negative correlation among the state variables,[15] or by hump-shaped loadings $B(\tau)$ on $Y(t)$ in (11). 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[16]). 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.

We were puzzled by the frequency of kinks in yield curves, the large average pricing errors, and the underestimation of yield volatilities implied by model $\mathbb{A}_2(3)_{DS}$, especially given its small goodness-of-fit statistics. 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, lead us to the following conjecture: The market prices of risk were set, in part, to replicate the effects of a non-zero $\sigma_{\theta r}$ (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% 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.[17]

[15]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 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 (13)) 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.

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

[17]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

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.[18] 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 that there may be some omitted *nonlinearity* in the these affine models.[19] 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.

---

practice. Nor does the criterion function force the means of the swap rates observed historically and simulated from the models to be the same.

[18]Slope is the difference between the ten- and two-year swap yields.

[19]In a one-factor setting, Ait-Sahalia [1] found evidence for non-linearity in the drifts of short rates. Boudoukh, Richardson, Stanton, and Whitelaw [8] provide evidence for a nonlinear relationship between slope and level in a two-factor setting.

VI Conclusion

In this paper we presented 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 characterized the "maximally flexible" canonical model and the nature of the admissible factor correlations and conditional volatilities that these canonical models can accommodate. We then applied this classification scheme to the family of three-factor affine term structure models in order to characterize the over-identifying restrictions implicit in several of the more popular affine term structure models in the literature.

A thorough empirical investigation of two of the four branches of the three-factor family of affine models was carried out to evaluate the goodness-of-fit of models with the central tendency and volatility of the short rate following independent affine diffusions. We found that correlation restrictions implicit in these models were strongly rejected by the data. One reason this may not have been apparent from previous studies is that empirical studies of affine models of the short rate have typically used data on the short rate alone to estimate multi-factor models. In contrast, we fit our models using data on bonds with three different maturities.

Finally, the empirical evidence suggests that the combinations of correlation and volatility specifications allowable within the affine family of models, given the requirements of admissibility, do not fully describe the conditional distributions of swap rates. In particular, further exploration of the parameterization of the risk premiums seems warranted along two dimensions. First, within the family $\mathbb{A}_2(3)$, the evidence suggests that many features of the distributions of yields would be better fit with smaller market prices of risk. Future research will explore what is lost in terms of fit by reducing λ. And within both of the families $\mathbb{A}_1(3)$ and $\mathbb{A}_2(3)$, it would be informative to explore nonlinear specifications of the risk premiums. So long as the state variables follow affine diffusions under the risk-neutral distribution, nonlinear risk premiums can be handled directly within the valuation and estimation frameworks exploited in this paper.

A Appendix to Section III

The main purpose of this appendix is to provide some technical details to the discussion in Section III. The emphasis is on intuitive understanding of the issues rather than on formal proofs. Toward this end, the appendix does two things. First, it explains why our canonical model satisfies the known (minimal) sufficient conditions for admissibility of an affine diffusion and describes the normalizations imposed to achieve a just identified model. More formal treatments of these issues for affine asset pricing models are presented in Duffie and Kan [13] and Dai, Liu, and Singleton [11]. Second, for the three-factor models in $\mathbb{A}_1(3)$ and $\mathbb{A}_2(3)$ discussed in Section III, we establish the equivalence of the canonical, AY, and Ar representations.

A.1 Invariant Transformations

In addressing both of these issues we will have frequent need to transform and rescale 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 $\phi = (\delta_0, \delta_y, \mathcal{K}, \Theta, \Sigma, \{\alpha_i, \beta_i : 1 \leq i \leq N\}, \lambda)$. A transformation of the model is represented by an operator $\mathcal{T}$ such that $\mathcal{T}Y(t)$, $\mathcal{T}W(t)$ and $\mathcal{T}\phi$ are the state vector, the vector of Brownian motions, and the parameter vector, respectively, for the transformed model. Invariant transformations are defined as follows:

Definition A.1 (Invariant Transformation) *An invariant transformation $\mathcal{T}$ of an N-factor* ATSM *is an arbitrary combination of* affine transformations $\mathcal{T}_A$, diffusion rescalings $\mathcal{T}_D$, Brownian motion rotations $\mathcal{T}_O$, *and* permutations $\mathcal{T}_P$, *such that,*

- *If $\mathcal{T} = \mathcal{T}_A$, then*

$$\mathcal{T}Y(t) = LY(t) + \vartheta, \ \mathcal{T}W(t) = W(t),$$
$$\mathcal{T}\phi = (\delta_0 - \delta_y' L^{-1}\vartheta, L'^{-1}\delta_y, L\mathcal{K}L^{-1}, \vartheta + L\Theta, L\Sigma,$$
$$\{\alpha_i - \beta_i' L^{-1}\vartheta, L'^{-1}\beta_i : 1 \leq i \leq N\}, \lambda),$$

where L is an $N \times N$ non-singular matrix, and ϑ is an $N \times 1$ vector.

- *If $\mathcal{T} = \mathcal{T}_D$, then*

$$\mathcal{T}Y(t) = Y(t), \ \mathcal{T}W(t) = W(t),$$
$$\mathcal{T}\phi = (\delta_0, \delta_y, \mathcal{K}, \Theta, \Sigma D^{-1}, \{D_{ii}^2 \alpha_i, D_{ii}^2 \beta_i : 1 \leq i \leq N\}, D\lambda),$$

where D is an $N \times N$ non-singular diagonal matrix.

- *If $\mathcal{T} = \mathcal{T}_O$, then*

$$\mathcal{T}Y(t) = Y(t), \ \mathcal{T}W(t) = OW(t),$$
$$\mathcal{T}\phi = (\delta_0, \delta_y, \mathcal{K}, \Theta, \Sigma O^{\mathrm{T}}, \{\alpha_i, \beta_i : 1 \leq i \leq N\}, O\lambda),$$

where O is an $N \times N$ orthogonal matrix (i.e., $O^{-1} = O^{\mathrm{T}}$) that commutes with $S(t)$.

- *If $\mathcal{T} = \mathcal{T}_P$, then*

$$\mathcal{T}Y(t) = PY(t), \ \mathcal{T}W(t) = PW(t),$$
$$\mathcal{T}\phi = (P\delta_0, P\delta_y, P\mathcal{K}P^{\mathrm{T}}, P\Theta, P\Sigma P^{\mathrm{T}}, P\alpha, P\mathcal{B}P^{\mathrm{T}}, P\lambda),$$

where P is an $N \times N$ permutation matrix.

Invariant affine transformations $\mathcal{T}_A$ are generally possible, because of the linear structure of *ATSM*s and the fact that the state variables are not observed. For instance, if $r(t) = \delta_0 + \delta_y' Y(t)$, then an example of a $\mathcal{T}_A$ is a transformation of $Y(t) \rightarrow LY(t)$ and $\delta_y \rightarrow L^{-1'}\delta_y$ with L a non-singular matrix. A diffusion rescaling $\mathcal{T}_D$ rescales the parameters of $[S(t)]_{ii}$ and the i^{th} entry of λ by the same constant. Such rescalings may be possible, because only the combinations $\Sigma S(t) \Sigma'$ and $\Sigma S(t) \lambda$ enter the pricing equations (11), (12), and (13). 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. Finally, a permutation $\mathcal{T}_P$ simply reorders the state variables which has no observable consequences. It is easily checked that any two *ATSM*s linked by an invariant transformation are equivalent in the sense that the implied bond prices (including the short rate) and their distributions are exactly the same.

A.2 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 (6). 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 represention 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:

$$S_{ii}(t) = Y_i(t), \quad 1 \leq i \leq m, \tag{46}$$

$$S_{jj}(t) = \alpha_j + \sum_{k=1}^{m} [\beta_j]_k Y_k(t), \quad m+1 \leq j \leq N, \tag{47}$$

where $\alpha_j \geq 0$, $[\beta_j]_i \geq 0$.[20] Therefore, as long as $Y^{\mathrm{B}}(t) \equiv (Y_1, Y_2, \ldots, Y_m)'$ is non-negative with probability one, the canonical representation of $Y(t) = (Y^{\mathrm{B}'}(t), Y^{\mathrm{D}'}(t))'$, where $Y^{\mathrm{D}}(t) \equiv (Y_{m+1}, Y_{m+2}, \ldots, Y_N)$, will be admissible.

In general, Y^{B} follows the diffusion

$$dY^{\mathrm{B}}(t) = \mathcal{K}^{\mathrm{B}}(\Theta - Y(t))dt + \Sigma^{\mathrm{B}}\sqrt{S(t)}dW(t). \tag{48}$$

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

Condition 1 $\mathcal{K}^{\mathrm{BD}} = 0_{m\times(N-m)}$,

Condition 2 $\Sigma^{\mathrm{BD}} = 0_{m\times(N-m)}$,

Condition 3 $\Sigma_{ij} = 0, \; 1 \leq i \neq j \leq m$,

Condition 4 $\mathcal{K}_{ij} \leq 0, \; 1 \leq i \neq j \leq m$,

Condition 5 $\mathcal{K}^{\mathrm{BB}}\Theta^{\mathrm{B}} > 0$.

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

Condition 1 is imposed because otherwise there would be a positive probability that the drift of Y^{B} at the zero boundary becomes negative (since $Y^{\mathrm{D}}(t)$ is not bounded from below). Conditions 2 and 3 are imposed to prevent $Y^{\mathrm{B}}(t)$ from diffusing across zero due to non-zero correlation between $Y^{\mathrm{B}}(t)$ and $Y^{\mathrm{D}}(t)$. Condition 4 (same as (21)) is imposed because otherwise, with $Y^{\mathrm{B}} \geq 0$, 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 4 and 5 assure that the drift condition

$$\mathcal{K}_{ii}\Theta_i + \sum_{j=1;j\neq i}^{m} \mathcal{K}_{ij}(\Theta_j - Y_j(t)) \geq 0 \tag{49}$$

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

Under Conditions 1 – 5, the existence of an (almost surely) non-negative and non-explosive solution to our canonical representation of (6) is assured because its drift and diffusion functions are continuous and satisfy a growth condition (see Ikeda and Watanabe [21], Chapter 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 Theorem 1 of Yamada and Watanabe [27]).[21] The state space for the solution is $\mathbb{R}_+^m \otimes \mathbb{R}^{N-m}$.

Finally, Condition 5 implies that the zero-boundary of Y^{B} is at least reflecting. This is because, under Conditions 1 – 3, the sub-vector $Y^{\mathrm{B}}(t)$ is an *autonomous* multi-variate correlated square-root process governed by

$$dY^{\mathrm{B}}(t) = \mathcal{K}^{\mathrm{BB}}(\Theta^{\mathrm{B}} - Y^{\mathrm{B}}(t))dt + \sqrt{S^{\mathrm{BB}}(t)}dW^{\mathrm{B}}(t). \tag{50}$$

If the off-diagonal elements of $\mathcal{K}^{\mathrm{BB}}$ are zero, then (50) is an m-dimensional independent square-root process. That the zero boundary is reflecting is trivial in this case. Under Condition 4, the drift of the correlated square root process dominates that of the independent square-root process. By appealing to Lemma A.3 of Duffie and Kan [13], we conclude that the zero-boundary for the correlated square-root process is at least reflecting.[22]

[21] To appeal to Yamada and Watanabe [27], we note that, without loss of generality, Σ may be normalized to the identity matrix (see Section A.3). This normalization is imposed in our canonical model.

[22] Condition 5 may be replaced by the stronger condition $\mathcal{K}^{\mathrm{BB}}\Theta^{\mathrm{B}} \geq 1/2$, as in Duffie and Kan [13]. The stronger condition, under which the zero boundary for Y^{B} is entrance, is the multi-variate generalization of the Feller condition.

A.3 Normalizations on The Canonical Representation

The preceding restrictions assure admissibility, but the resulting model is econometrically underidentified. The canonical representation in Definition III.1 results from one particular set of normalizations to achieve a just-identified model. 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:

- 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. This indeterminacy is eliminated by normalizing the upper-diagonal block of $\mathcal{B}$ to be diagonal.
- The lower-diagonal blocks of $\mathcal{K}$ and Σ, which determine the inter-dependencies among the elements of Y^{D}, are not separately identified. This indeterminacy is eliminated by normalizing the lower-diagonal block of Σ to be diagonal.
- The lower-left blocks of $\mathcal{K}$ and Σ, which determine the inter-dependencies between the elements of Y^{B} and Y^{D}, are not separately identified. We are free to normalize either $\mathcal{K}^{\mathrm{DB}}$ or Σ^{DB} to zero. We choose to set $\Sigma^{\mathrm{DB}} = 0$ in our canonical representation.[23]

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, $1 \leq i \leq m$.

[23]Starting from a model with non-zero Σ^{DB}, the affine transformation with

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

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

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 The possibility that a Brownian motion rotation $\mathcal{T}_O$ can be applied to obtain an equivalent model gives rise to a more subtle form of under-identification. 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. Second, 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.

A.4 Generating $\mathbb{A}_m(N)$ from the Canonical Representation

These normalizations ensure that the only invariant transformation that takes a canonical model to another canonical model (with the above restrictions and normalizations preserved) is the identity transformation. However, 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 *ATSM*s. 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. As shown in Section III, all of the extant *ATSM*s examined in the literature reside in one of these $\mathbb{A}_m(N)$. Therefore, our classification scheme allows us to derive the most flexible, admissible generalizations of extant models.

A.5 Alternative Representations

In the canonical representation, the admissibility conditions are intuitive and easily verified. In practical applications, however, it is often convenient to work with alternative but equivalent representations. The following subsections derive the equivalent *AY* and *Ar* representations of the $AM_1(3)$ and $AM_2(3)$ models discussed in Section III.

A.5.1 Equivalent Representations of $AM_1(3)$ Models

As mentioned in Section III.B, (27 – 29) is an equivalent AY representation of the canonical representation.

The Ar form of this maximal model is obtained by the following steps. Starting from (27 – 29), we apply the affine transformation $(\mathcal{T}_A : (L, \vartheta))$ with

$$L = \begin{pmatrix} [\beta_3]_1(1+\sigma_{23})^2 & 0 & 0 \\ 0 & q & 0 \\ \delta_1 & 1 & 1 \end{pmatrix}, \ \vartheta = \begin{pmatrix} 0 \\ \delta_0 + \delta_1\theta_1 \\ \delta_0 \end{pmatrix}, \tag{51}$$

and the diffusion rescaling $(\mathcal{T}_D : D)$ with

$$D = \begin{pmatrix} [\beta_3]_1(1+\sigma_{23})^2 & 0 & 0 \\ 0 & q & 0 \\ 0 & 0 & 1+\sigma_{23} \end{pmatrix}, \tag{52}$$

where $q = (\kappa_{33} - \kappa_{22})/\kappa_{33}$. Then, re-labeling the new state variables as $v(t)$, $\theta(t)$ and $r(t)$ respectively, and re-defining the free parameters, we obtain (26), where

$$\begin{aligned}
(\mu, \nu, \kappa) &= (\kappa_{11}, \kappa_{22}, \kappa_{33}), && (53)\\
(\bar{v}, \bar{\theta}) &= ([\beta_3]_1\theta_1(1+\sigma_{23})^2, \delta_0 + \delta_1\theta_1), && (54)\\
(\eta, \zeta) &= (\sqrt{[\beta_3]_1(1+\sigma_{23})^2}, \sqrt{q^2\,\alpha_2}), && (55)\\
\sigma_{rv} &= (\delta_1 + \sigma_{21} + \sigma_{31})/[\beta_3]_1/(1+\sigma_{23})^2, && (56)\\
\boxed{\kappa_{rv}} &= \delta_1(\kappa_{11} - \kappa_{33})/[\beta_3]_1/(1+\sigma_{23})^2, && (57)\\
\boxed{\sigma_{\theta v}} &= q\,\sigma_{21}/[\beta_3]_1/(1+\sigma_{23})^2, && (58)\\
\boxed{\sigma_{\theta r}} &= q\,\sigma_{23}/(1+\sigma_{23}), && (59)\\
\boxed{\sigma_{r\theta}} &= (1+\sigma_{32})/q, && (60)\\
\boxed{\beta_\theta} &= q^2\,[\beta_2]_1/[\beta_3]_1/(1+\sigma_{23})^2, && (61)\\
\boxed{\alpha_r} &= \alpha_3(1+\sigma_{23})^2, && (62)\\
(\lambda_v, \lambda_\theta, \lambda_r) &= ([\beta_3]_1(1+\sigma_{23})^2\lambda_1, q\lambda_2, (1+\sigma_{23})\lambda_3). && (63)
\end{aligned}$$

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

$$\delta_1 = 0,\ (\sigma_{21}, \sigma_{23}, \sigma_{32}) = (0,0,0),\ \alpha_3 = 0,\ \beta_{12} = 0. \tag{64}$$

A.5.2 Equivalent Representations of $AM_2(3)$ Models

The AY representation (32) – (34) 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

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 0 & q & 0 \\ \delta_1 & 1 & 1 \end{pmatrix}, \; \vartheta = \begin{pmatrix} 0 \\ 0 \\ \delta_0 \end{pmatrix}, \tag{65}$$

coupled with a diffusion rescaling that sets the diagonal elements of Σ to 1:

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & q & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{66}$$

where $q = (\kappa_{33} - \kappa_{22} - \delta_1 \kappa_{12})/\kappa_{33}$. The resulting Ar model is (31), where

$$(\mu, \nu, \kappa) = (\kappa_{11}, \kappa_{22}, \kappa_{33}), \tag{67}$$
$$(\bar{v}, \bar{\theta}) = (\theta_1, q\,\theta_2), \tag{68}$$
$$(\eta, \zeta) = (\sqrt{[\beta_1]_1}, \sqrt{q\,[\beta_2]_2}), \tag{69}$$
$$\bar{r} = \delta_0 + \delta_1 \theta_1 + \theta_2, \tag{70}$$
$$\kappa_{v\theta} = \kappa_{12}/q, \tag{71}$$
$$\kappa_{\theta v} = q\,\kappa_{21}, \tag{72}$$
$$\kappa_{rv} = \kappa_{21} + \delta_1(\kappa_{11} - \kappa_{33}), \tag{73}$$
$$\sigma_{rv} = (\sigma_{31} + \delta_1), \tag{74}$$
$$\sigma_{r\theta} = (1 + \sigma_{32})/q, \tag{75}$$
$$\alpha_r = \alpha_3, \tag{76}$$
$$\beta_\theta = [\beta_3]_2/q, \tag{77}$$
$$(\lambda_v, \lambda_\theta, \lambda_r) = (\lambda_1, q\lambda_2, \lambda_3). \tag{78}$$

In order 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 which 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 $q < 0$).

Table I: Overall Goodness-of-Fit

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

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

Parameter	Estimate (t-ratio)		
Ar	$\mathbb{A}_1(3)_{BDFS}$	$\mathbb{A}_1(3)_{DS}$	$\mathbb{A}_1(3)_{Max}$
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)

The parameters pertain to the Ar-representation (26) of models in the $AYM_1(3)$ branch. Parameters indicated by "fixed" are restricted to zero.

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

Parameter	Estimate (t-ratio)		
Ar	$\mathbb{A}_2(3)_{Chen}$	$\mathbb{A}_2(3)_{DS}$	$\mathbb{A}_2(3)_{Max}$
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 (1.181)
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)

The parameters pertain to the Ar-representation (31) 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".

Table IV: Moments of Pricing Errors (in basis points)

Model/Swap		Mean	Std.	ρ	Q-Invert	Q-Steep
	3yr	-11.3	9.6	.95	-8.1	-16.6
$\mathbb{A}_1(3)_{DS}$	5yr	16.9	16.5	.97	-12.0	-26.6
	7yr	-12.7	10.1	.94	-9.1	-17.6
	3yr	-43.1	11.6	.97	-55.4	-27.7
$\mathbb{A}_2(3)_{DS}$	5yr	-63.3	12.1	.96	-75.6	-49.1
	7yr	-47.5	8.3	.94	-54.1	-38.1

Mean is the sample mean of the pricing errors for the swap yields and models indicated in column 1. *Std.* is the sample standard deviation and ρ 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.

Figure 1: Time Series of Swap Yields, Weekly, 4/03/87 – 8/23/96

6-month LIBOR, 2-, 3-, 4-, 5-, 7-, and 10-year swap

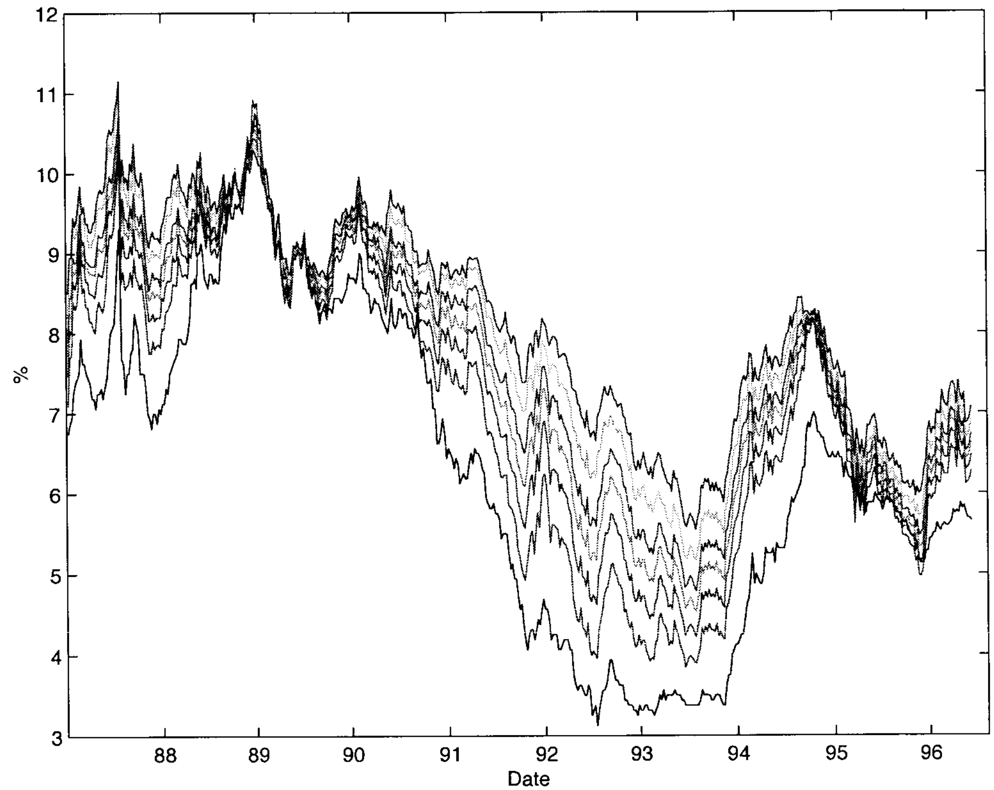


Figure 2: Fitted SNP Scores

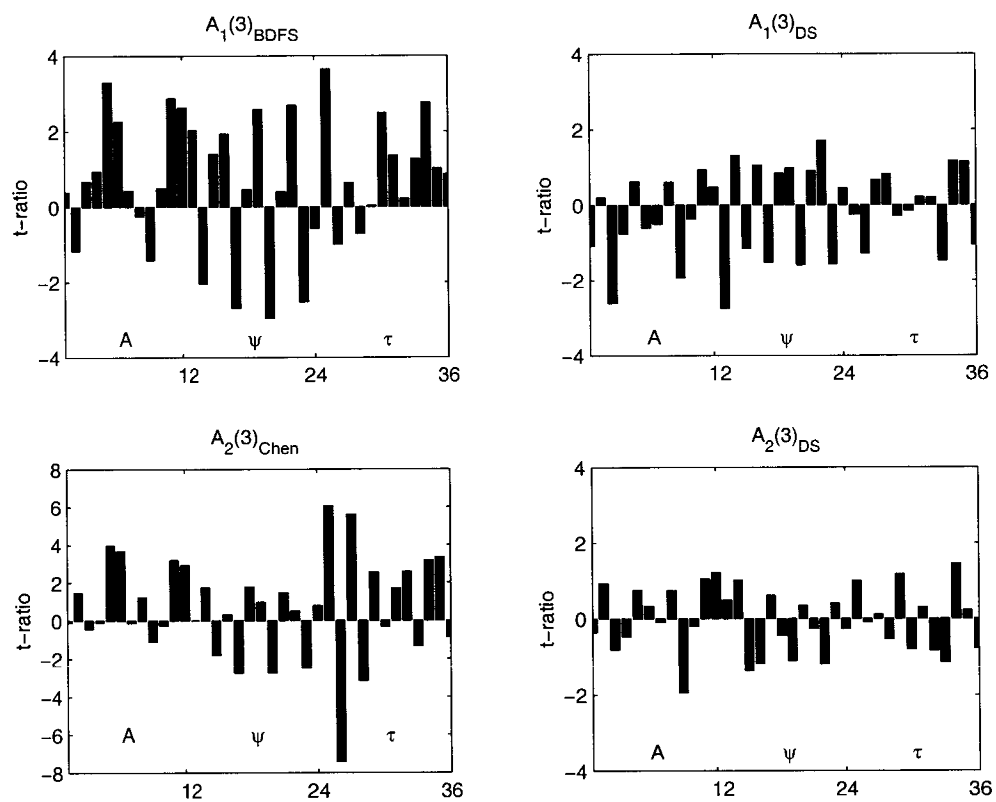
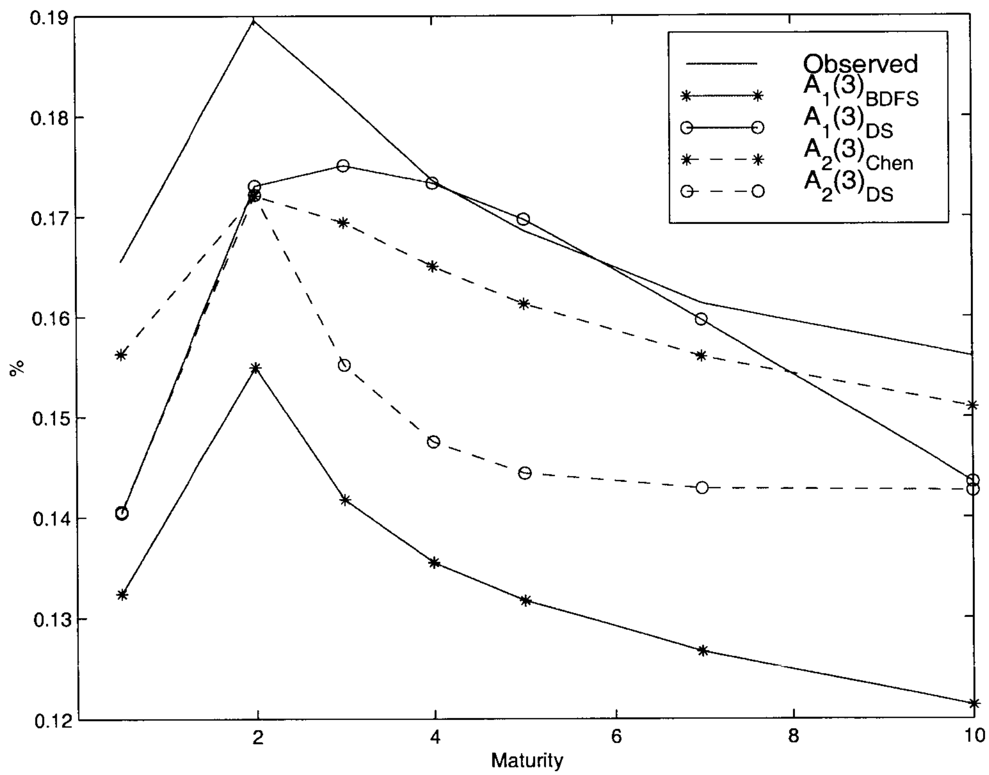


Figure 3: Term Structure of Volatility of Yield Changes



References

[1] Y. Ait-Sahalia. Testing Continuous-Time Models of the Spot Interest Rate. *Review of Financial Studies*, 9(2):385 – 426, 1996.

[2] T. G. Andersen and J. Lund. Stochastic Volatility and Mean Drift In the Short Term Interest Rate Diffusion: Sources of Steepness, Level and Curvature in the Yield Curve. Working paper, February 1996.

[3] D. Backus, S. Foresi, and C. Telmer. Affine Models of Currency Prices. Working paper, New York University, 1996.

[4] D. K. Backus and S. E. Zin. Reverse Engineering the Yield Curve. NBER Working Paper #4676, 1994.

[5] G. S. Bakshi and Z. Chen. Asset Pricing without Consumption or Market Portfolio Data. Working paper, 1997.

[6] P. Balduzzi, S. R. Das, and S. Foresi. The Central Tendency: A Second Factor in Bond Yields. Working paper, 1995.

[7] P. Balduzzi, S. R. Das, S. Foresi, and R. Sundaram. A Simple Approach to Three Factor Affine Term Structure Models. *Journal of Fixed Income*, 6:43–53, December 1996.

[8] J. Boudoukh, M. Richardson, R. Stanton, and R. F. Whitelaw. The stochastic behavior of interest rates: Implications from a multifactor, nonlinear continuous-time model. New York University Working Paper, 1998.

[9] L. Chen. *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, 1996.

[10] R. Chen and L. Scott. Maximum Likelihood Estimation For a Multifactor Equilibrium Model of the Term Structure of Interest Rates. *Journal of Fixed Income*, 3:14–31, December 1993.

[11] Q. Dai, J. Liu, and K. Singleton. Admissibility and Identification of Affine Asset Pricing Models. Working paper, Stanford University, 1997.

[12] D. Duffie. *Dynamic Asset Pricing Theory.* Princeton University Press, 2nd edition, 1996.

[13] D. Duffie and R. Kan. A Yield-Factor Model of Interest Rates. *Mathematical Finance*, 6(4):379–406, October 1996.

[14] D. Duffie and K. Singleton. Simulated Moments Estimation of Markov Models of Asset Prices. *Econometrica*, 61:929–952, 1993.

[15] D. Duffie and K. Singleton. An Econometric Model of the Term Structure of Interest Rate Swap Yields. Journal of Finance, forthcoming, 1996.

[16] M. Fisher and C. Gilles. Estimating Exponential Affine Models of the Term Structure. Working paper, 1996.

[17] A. R. Gallant and J. R. Long. Estimation Stochastic Differential Equatiosn Efficiently by Minimum Chi-Square. *Biometrika*, forthcoming, 1997.

[18] A. R. Gallant and G. Tauchen. Which Moments to Match? *Econometric Theory*, 12:657–681, 1996.

[19] L. Hansen. Large sample properties of generalized method of moment estimators. *Econometrica*, 50:1029–1056, 1982.

[20] L. P. Hansen and S. F. Richard. The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models. *Econometrica*, 55(3):587–613, May 1987.

[21] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes.* Amsterdam: North-Holland, 1981.

[22] Narasimhan Jegadeesh and George G. Pennacchi. The Behavior of Interest Rates Implied by the Term Structure of Eurodollar Futures. *Journal of Money, Credit, and Banking*, 28(3), August 1996.

[23] T. Langetieg. A Multivariate Model of the Term Structure. *Journal of Finance*, 35:71–97, 1980.

[24] L. T. Nielsen and J. Saá-Requejo. Exchange Rate and Terms Structure Dynamics and the Pricing of Derivative Securities. INSEAD working paper, 1993.

[25] N. D. Pearson and T. Sun. Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross model. *Journal of Finance*, XLIX(4):1279–1304, September 1994.

[26] O. Vasicek. An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics*, 5:177–188, 1977.

[27] T. Yamada and S. Watanabe. On the Uniqueness of Solutions of Stochastic Differential Equations. *J. Math. Kyoto Univ.*, 11(1):155–167, 1971.