NBER WORKING PAPER SERIES

SPECIFICATION ANALYSIS OF AFFINE TERM STRUCTURE MODELS

Qiang Dai
Kenneth J. Singleton

Working Paper 6128

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 1997

We would like to thank Darrell Duffie for extensive discussions, Ron Gallant 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 Northwestern University, the NBER Asset Pricing Group meeting, and the Western Finance Association Annual Meetings. This paper is part of NBER's research program in Asset Pricing. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

Specification Analysis of Affine Term
Structure Models
Qiang Dai and Kenneth J. Singleton
NBER Working Paper No. 6128
August 1997
Asset Pricing

ABSTRACT

This paper characterizes, interprets, and tests the over-identifying restrictions imposed in affine models of the term structure. Letting $r(t) = \delta' Y(t)$, where Y is an unobserved vector affine process, our analysis proceeds in three steps. First, we show that affine models can be categorized according to the different over-identifying restrictions they impose on (i) δ, and (ii) the parameters of the diffusion matrices. Second, this formulation is shown to be equivalent to a model in which there is a terraced drift structure with one of the state variables being the stochastic long-run mean of r. This equivalence allows direct comparisons of the substantive restrictions on the dynamics of interest rates imposed in CIR-style models and models in which the state variables are the stochastic long-run mean and volatility of r. Third, we compute simulated method of moments estimates of a three-factor affine term structure model, and test the over-identifying restrictions on the *joint distribution* of long- and short-term interest rates implied by extant affine models of r. We find allowing for correlated factors is key to simultaneously describing the short and long ends of the yield curve. This finding is interpreted in terms of the properties of the risk factors underlying term structure movements.

Qiang Dai
Graduate School of Business
Stanford University
Stanford, CA 94305
dq@leland.stanford.edu

Kenneth J. Singleton
Graduate School of Business
Stanford University
Stanford, CA 94305
and NBER
ken@future.stanford.edu

Introduction

RECENTLY CONSIDERABLE attention has been focused on the "affine" class of asset pricing models in which the drifts and volatility coefficients of the state-variable processes are affine functions (e.g., Duffie and Kan (1996)). Two approaches to developing affine models have been pursued in the term structure literature. One assumes that the instantaneous short rate is a linear combination of an unobserved state vector Y, $r(t) = \delta' Y(t)$, and that Y follows an affine diffusion model (hereafter AY models). This approach is rooted in the risk management literature, which has found it convenient to decompose term structure movements into changes in "level", "slope," and "curvature" factors. The unobserved Y's are the dynamic counterparts to these constructs. Examples of AY models include the multi-factor, square-root diffusion models of Y used by Chen and Scott (1993), Pearson and Sun (1994), and Duffie and Singleton (1996) to study term structures of Treasury and swap rates. The second approach posits a model for the instantaneous short rate (r) in terms of its own lag and other state variables (hereafter Ar models). See, for example, Chen (1996), Balduzzi, Das and Foresi (1995), and Backus, Foresi, and Telmer (1996). This approach has evolved from the literature on one-factor models of the instantaneous short rate r, with the additional state variables representing the stochastic long-run mean and volatility of r.

Though both approaches seek models that describe the temporal behavior of bond yields, there has been little comparative analysis of the nature of the restrictions imposed on the distributions of bond yields in extant AY and Ar models. Whether or not these restrictions are *normalizations* or *over-identifying* restrictions depends on the information about the joint, conditional distribution of bond yields used to identify the term structure model. Ar models have typically focused on the conditional distribution of a short-term rate to the exclusion of information about long-term yields. Consequently many (though not all) of the restrictions imposed on the diffusion coefficients in these models (e.g., independent diffusions) are normalizations that are necessary to identify the parameters of the state vector from knowledge of the distribution of a single short rate. However, suppose there are N state variables and the parameters are to be identified from information about the joint conditional distribution of $M(\geq N)$ bond yields. Then many of the restrictions imposed in previous Ar and AY models represent potentially strong over-identifying restrictions on the joint distribution of long-and short-term bond yields.

Starting from the premise that the goal of both Ar and AY models is to explain the term structure of interest rates – i.e., the joint distribution of long- and short-term rates – this paper characterizes, interprets, and tests the *over-identifying* restrictions imposed in affine term structure models. Our analysis proceeds in three steps. First, we show that, within the general AY model with $M = N$, not all of the parameters are identified from knowledge of the conditional distribution of bond prices. A convenient normalization that contributes to eliminating this under-identification is to assume that there is no feedback between the state variables through their drifts. Imposing this normalization, we show that affine models can be distinguished by the different over-identifying restrictions they impose on (i) δ, and (ii) the parameters of the diffusion matrices.

Second, we show that every AY model is analytically equivalent to an Ar model and *vice versa*. Moreover, the diagonal drift of AY models maps to a "terraced" drift structure in multi-factor Ar models of r. This equivalence allows direct comparisons of the substantive restrictions on the dynamics of interest rates imposed in AY and Ar models.

Specifically, we show that the models of Chen (1996), Balduzzi, Das and Foresi (1996), Andersen and Lund (1996), among others, implicitly restrict one of the δ's to zero. The dimension of the state vector Y (N) is 3, but the number of Y's that directly determine r through the relation $r(t) = \delta' Y(t)$ (n) is 2; r is a linear combination of a strict subset of the Ys. In contrast, the multi-factor Cox, Ingersoll and Ross (1985)-style (CIR) models assume that $N = n$.

Furthermore, the two-factor CIR model ($N = 2$), for example, is shown to be equivalent to an Ar model in which one of the state variables is r and the other is the stochastic long-run mean of r. Thus, CIR models have implicitly always incorporated the central tendency factor that Balduzzi, Das and Foresi (1996) and Andersen and Lund (1996) have recently argued is an essential feature of Ar models for explaining the time series properties of short-term rates. Moreover, given an identified AY model with N factors, there are N equivalent, identified Ar models with the long-run mean of r being Y_i in the i^{th} model, $i = 1, \ldots, N$. What distinguishes these Ar models, besides the different drift specifications of r, are the different specifications of the diffusion coefficients. In other words, it is the assumed form of the diffusion coefficients in Ar models that dictates which Y_i from the AY representation is the long-run mean (or stochastic volatility) or r.

Both Ar and AY models have imposed potentially strong restrictions on the conditional correlations and variances of the state variables. Many of these restrictions are unnecessary for the purpose of identifying or estimating affine term structure models using yields on bonds with multiple maturities. The state variables in the general affine model may be correlated and their conditional variances and correlations may depend on multiple state variables. At the same time, one does not have complete freedom in specifying the diffusion coefficients, because of the need to rule out certain nontrivial rotations and to assure existence of solutions to the stochastic differential equations describing the state variables. We illustrate the interplay between identification and existence conditions with several models in which volatility is stochastic and the diffusions are correlated.

The third step of our analysis is an empirical analysis of a three-factor affine term structure model that nests many previous specifications as special cases. Affine specifications of the state variables lead to closed- or nearly closed-form solutions for the prices of zero-coupon bonds, so the the simultaneous computation of "arbitrage-free" coupon bond prices and estimation of the unknown parameters using multiple bond yields is computationally feasible. We compute simulated method-of-moments estimators (Duffie and Singleton (1996) and Gallant and Tauchen (1996b)) of a three-factor affine model using data on six-month, two-year, and ten-year swap rates simultaneously. Various specification tests suggest that a three-factor AY model with correlated factors adequately describes the dynamics of swap rates.

Then we test the restrictions imposed by Chen (1996)[1] and Andersen and Lund (1996) that $N = 3$ and $n = 2$ (one of the δ's is zero), and that the conditional correlations among all of the state variables are zero. Within our affine framework, the over-identifying restriction of independent diffusions – a normalization when estimating an Ar model using data on a short rate alone – is strongly rejected. Thus, affine models with independent diffusions fail to describe the joint distribution of long- and short-term swap rates.

Moreover, the equivalence between Ar and AYD models provides insights into the reasons why zero conditional correlations are inconsistent with the data. With the first factor in the Ar representation being r, we show that the second factor is well proxied by a long-term rate. In addition, the third

[1]The "benchmark" model referred to by Chen (1996) is a three-factor model obtained by assuming the long run mean and conditional variance of r follow independent square-root processes.

factor, typically interpreted as a volatility factor in Ar models, is in fact well proxied by the slope of the swap curve! These findings are implications of the assumed structure of the diffusion coefficients. It follows that assuming independent diffusions amounts to the implicit assumption that r is conditionally uncorrelated with the level and slope of the yield curve which, not surprisingly, is counter-factual.

The common assumption that the second and third factors are the central tendency and stochastic volatility of r, respectively, is also challenged by our findings. Though most of the improvement in fit comes from relaxation of the correlation restrictions, we also reject at conventional significance levels the restriction that one of the δ's is zero. In the context of Ar representations of affine models, this finding suggests that the drift of r depends on both the second and third factors, and not simply a central tendency or stochastic long-run mean of r.

Though our focus is on the term structure, the subsequent observations are directly applicable to many affine currency pricing models, because of the close link between bond prices and forward exchange rates. Under covered interest parity, the forward premium is the difference of the domestic and foreign interest rates for the horizon of the forward contract. Thus, affine models of default-free, zero-coupon bond prices in each country lead to an affine model for the forward premium. See, for example, Nielsen and Saá-Requejo (1993) and Backus, Foresi, and Telmer (1996).

The remainder of the paper is organized as follows. Section I sets up the multi-factor affine term structure model by directly specifying the pricing kernel in an arbitrage-free economy. Section II presents a general discussion on the econometric identification of affine term structure models. The discussion leads to the presentation of an affine model that nest extant affine specifications. The over-identifying restrictions implicit in the Chen (1996) model are tested and presented in Section III. Section IV concludes.

I The Affine Bond Pricing Model

Consider a frictionless economy with riskless borrowing and lending opportunities. Let us 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 technical conditions (see Duffie (1996) and Hansen and Richard (1987)), there exists a state price deflater $\pi(t)$, 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 bond requires little more than the absence of arbitrage opportunities. The general affine term structure model is obtained by imposing the additional assumptions that

$$r(t) = \sum_{i=1}^{N} \delta_i Y_i(t) \tag{4}$$

and

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

where, $\delta = (\delta_1, \ldots, \delta_N)'$, 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

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

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

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

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 (1994) and Backus, Foresi, and Telmer (1996). Our formulation generalizes the continuous time, pricing kernels assumed by Bakshi and Chen (1997) and Nielsen and Saá-Requejo (1993), and is equivalent to that of Fisher and Gilles (1996). Thus, the subsequent analysis of the specifications of 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

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

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 $Y(t)$.[2] Specifically, it can be shown [see Duffie and Kan (1996)] 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)

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

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

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 term structure model.

[2]Our specification of the state variable dynamics under the real 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.

II Identification of Affine Models

The mapping in multi-factor models between the state variables and bond prices is typically non-linear. As such, the identification of the parameters from information about the conditional distribution of bond prices is typically not transparent. In this section we address the identification problem for the general affine term structure model defined in Section I. The identification problem for affine models is simplified considerably by the fact that zero-coupon bond are linear functions of the state vector. Even in this affine framework, there are subtle ways in which models may be under-identified due to the presence of the arbitrage pricing model the lies between the specification of the state process and the bond prices. Moreover, the interplay between existence and identification conditions is shown to restrict the flexibility one has in parameterizing the state process.[3]

We assume that identification is achieved by using information about the distributions of zero-coupon bond prices. Extensions of this discussion to the case of coupon bonds is straightforward. Furthermore, assume that the number of observed yields used in estimation (M) is equal to the number of factors (N). If $M < N$, as in recent multi-factor models of r, then additional normalizations will generally be required to achieve identification. These normalizations for the case $M < N$ become over-identifying restrictions when $M = N$. Since our focus is on modeling the term structure, we focus on the case $M = N$.[4]

We begin his section by showing that the general N-factor AY model is under-identified based on knowledge of the conditional distribution of N bond yields. A convenient normalization that contributes to eliminating this under-identification is to set $\mathcal{K}$ to a diagonal matrix. Then we establish a general equivalence result between the AY models with diagonal drifts and Ar models with "terraced" drift structures.

[3]We expect that the identification problem for other nonlinear models will be at least as challenging as for affine models. The following discussion will hopefully be informative about potential sources of under-identification or non-existence in other environments.

[4]If $M > N$, then common practice has been to introduce $M - N$ "pricing errors" as additional state variables, and to treat N yields as being measured without error (e.g., Chen and Scott (1993) and Duffie and Singleton (1996)). Since the state variables are inferred from the latter yields, the identification problem is essentially identical to the case of $M = N$.

II.A Under-identification of General Affine Models

To show that the AY model, defined by (4) – (9), is in general under-identified, let us suppose initially that the parameters of the affine model, including δ in (4), are unconstrained parameters. Then any transformation of Y to $\hat{Y}(t) = XY(t)$ and δ to $\hat{\delta}' = \delta' X^{-1}$, where X is an $N \times N$, non-singular matrix, preserves the affine structure of the model and leaves r unchanged. We shall refer to such transformations as non-singular rotations (NSR) of the state vector. The next proposition shows that zero-coupon bond prices are also invariant to NSR's. Proofs are given in the Appendix.

Proposition II.1 *Zero-coupon bond prices are invariant to NSR's of the state vector.*

This proposition follows from the observations that (i) under the real probability measure, $\hat{Y}(t) = X \times Y(t)$ (where X is a NSR) follows the diffusion

$$d\hat{Y}(t) = \hat{\mathcal{K}}\left(\hat{\Theta} - \hat{Y}(t)\right) dt + \hat{\Sigma}\,\hat{S}(t)\,dW(t), \tag{14}$$

where $\hat{\mathcal{K}} = X\mathcal{K}X^{-1}$, $\hat{\Sigma} = X\Sigma$, $\hat{\Theta} = X\Theta$, $[\hat{S}(t)]_{ii}$ has $\hat{Y}$ in place of Y and $\hat{\beta}_i = X'^{-1}\beta_i$ in place of β_i, and (ii) A and B in the ODEs (12) and (13) transform in such a way that the bond prices are invariant.

Moreover, this invariance extends to the conditional distribution of the bond prices:

Proposition II.2 *The joint conditional density of zero coupon bond prices at date t conditioned on prices at date s, $s < t$, is invariant under a NSR of the state vector.*

An implication of Proposition II.2 is that a given affine term structure model will in general serve as the "basis" of a family of observationally equivalent models generated by rotations of the initial model.

Eliminating the under-identification associated with rotations requires the imposition of normalizations/restrictions on the co-dependence of the state variables Y. There are three potential "channels" through which the state variables may be interdependent in an affine model: (i) through the feedback in the conditional mean (non-diagonal $\mathcal{K}$), (ii) through non-zero correlations of the Brownian motions (non-diagonal Σ), and (iii) as a result of dependence of the conditional variance on the state variables (non-zero elements in β_i other than in the i^{th} component). Proposition II.2 implies that the contributions of all of these channels cannot be separately identified in affine term structure models satisfying (4). Normalizations on $\mathcal{K}$, Σ, or $S(t)$ are necessary to eliminate this indeterminacy.

A common feature of the factor structures in AY models is that $\mathcal{K}$ is diagonal. Under mild regularity conditions, a diagonal $\mathcal{K}$ turns out to be a normalization that goes a long way toward eliminating under-identification associated with NSRs. For the purpose of interpreting term structure dynamics, this normalization has the attractive feature that mean reversion of Y_i is governed only by κ_i. Co-dependence among the state variables, and over-identifying restrictions on this dependence, are captured through the specifications of the matrices Σ and $\mathcal{B} \equiv [\beta_1, \beta_2, \ldots, \beta_N]$.

More precisely, we assume that

Assumption II.1 *The eigenvalues of $\mathcal{K}$ are positive real numbers, and there exists an $N \times N$ nonsingular matrix X with the property that*

$$X\mathcal{K}X^{-1} = \hat{\mathcal{K}} = \begin{bmatrix} \kappa_1 & 0 & \cdots & 0 \\ 0 & \kappa_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \kappa_N \end{bmatrix}. \tag{15}$$

The assumption that the eigenvalues are real rules out some potentially interesting dynamics associated with complex eigenvalues. However, as we show subsequently, all of the affine models that have been studied in the literature we are aware of presume that the eigenvalues of $\mathcal{K}$ are real.[5] This assumption can be relaxed if one has a specific parameterization of $\mathcal{K}$ in mind with complex eigenvalues and one is willing to restrict Σ *a priori* to assure identification. The assumption that the real parts of the eigenvalues are positive is necessary for stationarity of the distribution of Y. A sufficient condition for there to be a nonsingular matrix X that diagonalizes $\mathcal{K}$ is that the eigenvalues of $\mathcal{K}$ are distinct.

Assumption (II.1) and Proposition (II.2) imply:

[5]Beaglehole and Tierney (1991) discuss examples of Gaussian affine models with complex eigenvalues, but these or similar models have not been pursued in the empirical term structure or currency pricing literatures.

Proposition II.3 *Under Assumption (II.1), normalizing the feedback matrix $\mathcal{K}$ to be diagonal in the general affine model (4) – (7) leads to an observationally equivalent model for r and zero-coupon bond prices.*[6]

With $\mathcal{K}$ diagonal, the only admissible NSR's are diagonal matrices. In Section II.C we discuss additional normalizations that fully identify $\mathcal{K}$ and δ.

Notice that normalization to a diagonal $\mathcal{K}$ does not restrict the dependence of r on Y. In particular, r may be a linear combination of a strict subset of the state variables which, without loss of generality, we can take to be the first n Y's: $r(t) = \sum_{i=1}^{n} \delta_i Y_i(t)$. If $N - n$ of the δ's are zero, then the state variables $Y_i(t)$, $n < i \leq N$ are *auxiliary* factors in that they do not affect the short rate directly, but only indirectly by influencing the distribution of the *primary* factors ($Y_i(t)$, $1 \leq i \leq n$). To distinguish between models with one or more δ's being zero, we let $AYD(N,n)$ denote affine, diagonal drift models with a total of N factors (and yields) and n primary factors.[7]

There is another source of under-identification of the drift function in affine models. Specifically, when the α_i's are unconstrained, the N components of Θ are not uniquely determined. To see this, suppose that Y is shifted to $\hat{Y} = Y + \vartheta$. The long-run mean of the diffusion for $\hat{Y}$ is $\Theta + \vartheta$ and the conditional variances are of the form $\sqrt{\alpha_i - \beta_i'\vartheta + \beta_i'\hat{Y}(t)}$. The shift vector ϑ must satisfy the linear restriction $\delta'\vartheta = 0$ to preserve the relation (4) ($\delta'\hat{Y} = \delta'Y$). Thus, this shift-indeterminacy is eliminated by imposing $N-1$ normalizations on the vector Θ and the constants α_i in $S(t)$. For instance, one could normalize $N-1$ of the Θ's to be zero.

II.B An Equivalent Affine Model

Instead of parameterizing r as a linear combination of an unobserved state vector Y, many have parameterized the diffusion for r directly. Typical

[6]The freedom to diagonalize $\mathcal{K}$ presumes, of course, that there are not *a priori* restrictions on Σ and $S(t)$ in a model with non-diagonal $\mathcal{K}$. As noted previously, AY models typically impose a diagonal $\mathcal{K}$ and restrict Σ and $S(t)$. Our point is that we can interpret the restriction on $\mathcal{K}$ as a normalization and focus on the diffusion coefficients in characterizing the over-identifying restrictions. We could instead normalize Σ to be diagonal.

[7]The bond prices in the $AYD(N,n)$ model may be solved by writing the $AYD(N,n)$ as an $AYD(N,N)$ model, with the δ_i's associated with the auxiliary factors set to zero. Then (11) gives the zero coupon bond prices, and (12) and (13) give the factor loadings. The zero restrictions on δ_i's for the auxiliary factors appear only in (13).

parameterizations of such Ar (affine-r) models are special cases of the three-factor model[8]

$$dZ(t) = \begin{pmatrix} dr(t) \\ d\theta(t) \\ dv(t) \end{pmatrix} = \begin{pmatrix} \kappa\,(\theta(t) - r(t)) \\ \nu\,(\bar{\theta} - \theta(t)) \\ \mu\,(\bar{v} - v(t)) \end{pmatrix} dt + \Sigma_z\, S_z(t)\, dW(t), \qquad (16)$$

where $\theta(t)$ is interpreted as the long-run mean and $v(t)$ is the volatility of r. In this section we show that the general AYD model has an Ar representation similar to (16) and use this equivalence to interpret the restrictions on bond prices implicit in (16).

Consider an $AYD(N, n)$ model, where n may be strictly less than N. Any transformation of Y by a nonsingular matrix L with elements that are known functions of the parameters of the $AYD(N, n)$ model produces an equivalent model. And any transformation in which the first row of L is δ' gives an equivalent Ar model. A particularly revealing Ar model is obtained by the following transformation of Y:

$$Z(t) = \theta_z + LY(t), \qquad (17)$$

where L is a block-diagonal matrix, partitioned conformably with a partition of Y into n primary and $N - n$ auxiliary state variables, given by

$$L = \begin{pmatrix} L^{(n)} & 0_{n\times(N-n)} \\ 0_{(N-n)\times n} & I_{(N-n)\times(N-n)} \end{pmatrix}, \qquad (18)$$

with $L^{(n)}_{ji} = \gamma^{(j)}_i$, where $\gamma^{(j)}_i \equiv 0$ for $i < j$ and $1 \leq j \leq n$, and the $\gamma^{(j)}_i$ are constants for $i \geq j$ and $1 \leq j \leq n$. The shift vector is $\theta_z = (\theta_{z1}, \theta_{z2}, \ldots, \theta_{zn}, 0, \ldots, 0)'$. The diffusion representation of $Z(t)$ has drift function $\mu_z(t)$ and diffusion matrix $\Sigma_z S_z(t)$, where $\Sigma_z = L\Sigma$,

$$Z(t) = \begin{pmatrix} r(t) \\ Z_2(t) \\ \vdots \\ Z_{n-1}(t) \\ Z_n(t) \\ Z_{n+1}(t) \\ \vdots \\ Z_N(t) \end{pmatrix}, \quad \mu_z(t) = \begin{pmatrix} \kappa_1\,(Z_2(t) - Z_1(t)) \\ \kappa_2\,(Z_3(t) - Z_2(t)) \\ \vdots \\ \kappa_{n-1}\,(Z_{n-2}(t) - Z_{n-1}(t)) \\ \kappa_n\,(\bar{\theta} - Z_n(t)) \\ \kappa_{n+1}\,(\theta_{n+1} - Z_{n+1}(t)) \\ \vdots \\ \kappa_N\,(\theta_N - Z_N(t)) \end{pmatrix}, \qquad (19)$$

[8] See, for example, Chen (1996) and Balduzzi, Das and Foresi (1996). Andersen and Lund (1996) study a model with the same drift as (16), but their parameterization of volatility does not fit within the affine class.

and $[S_z]_{ii} = \sqrt{\alpha_{zi} + \beta_{zi}' Z(t)}$. The κ_i's, $\bar{\theta}$, θ_i's, α_{zi} and β_{zi} are functions of the parameters in the $AYD(N, n)$ model and known constants. (Details are given in Appendix B.)

Inspection of the form of this Ar representation leads to several observations about the factor structures in affine term structure models. The drift in (19) is such that, for the first $n - 1$ state variables, $Z_{i+1}(t)$ serves as the stochastic long-run mean of $Z_i(t)$. The drifts of the last $N - n$ Z's are of exactly the same form as the drifts of the $N - n$ auxiliary Y's. The fact that the last $N - n$ state variables affect the drift of r only through the diffusion coefficients is the Ar counterpart to the dependence of r on only the first n Y's in the AYD representation. With this correspondence in mind, we refer to Ar models with drifts of the form (19) as $Ar(N, n)$ models.

This equivalence result implies that the potential sources of testable over-identifying restrictions in all $AYD(N, n)$ and $Ar(N, n)$ models can be classified into two categories: restrictions on δ in (4) and restrictions on Σ and $S(t)$. We next discuss each of these in turn.

II.C Restrictions on the Number of Primary Factors

In $AYD(N, n)$ models, the normalization of $\mathcal{K}$ to be diagonal does not fully rule out non-trivial NSRs. When $n = N$, i.e., when all δ_i are known to be nonzero, then any transformation of the state vector by a non-singular diagonal matrix X will lead to an observationally equivalent model, since $X\mathcal{K}X^{-1}$ will also be diagonal. This indeterminacy is eliminated in $AYD(N, N)$ models by normalizing the δ_i to 1, for all $i = 1, \ldots, N$. With δ fixed, the requirement $\hat{\delta} = (X^{-1})'\delta = \delta$ implies that the diagonal elements of X must be unity. This was the normalization imposed by Chen and Scott (1993), Pearson and Sun (1994), and Duffie and Singleton (1996) in specifying their multi-factor CIR models. These are $AYD(N, N)$ models, because $\mathcal{K}$ was assumed to be diagonal and r depended on all N state variables.

An immediate implication of the preceding discussion is that one $Ar(N, N)$ representation of the N-factor CIR model is given by (19). In the case of $N = 2$, we see that the two-factor CIR model has an equivalent representation in which one of the state variables is r and the other is a stochastic "central tendency" factor to which r mean reverts. Recall that previous implementations of two-factor AY models (e.g., Duffie and Singleton (1996)) found that the two factors are highly correlated with the "level" (a long-term rate) and "slope" (long minus short rate) of the yield curve. Thus, the central tendency of r can be interpreted as the "level" or "slope" of the yield curve, depending on how one chooses to order the factors in the AY representation. These equivalent representations of the drift of r are not observationally equivalent, in general, because the choice of a particular Y as the central tendency factor has implications for the specification of the diffusion coefficients. The interpretation of the factors in the Ar models we estimate is explored more extensively in Section III.D.

Setting $N = 3$ and $n = 2$ in (19) leads to a representation of $\mu_z(t)$ that is identical to the drift in (16). It follows that extant three-factor affine models of the short rate have implicitly assumed that r depends directly on two of the three state variables in the model: $r(t) = Y_1(t) + Y_2(t)$.[9] This is a testable, over-identifying assumption (see below).

More generally, whether or not a primary risk factor has a stochastic long-run mean in an Ar representation depends on the value of n relative to N. In the case of $AYD(N, N)$ models, the equivalent Ar representation has each factor mean reverting to its own stochastic long-run mean in a terraced fashion. On the other hand, an $Ar(N, n)$ model implies an $AYD(N, n)$ representation in which $N - n$ of the δ's are restricted to be zero. These restrictions show up in the $Ar(N, n)$ model as $N - n$ state variables having constant long-run means.

When one is interested in testing the null hypothesis that $n = N_0$ (for $N_0 < N$) against the alternative $n = N$, δ cannot be normalized to the unit vector, because this null hypothesis is equivalent to $\delta_i = 0$ for $N_0 < i \leq N$. If estimation proceeds under the alternative (with the δ_i being free parameters for $N_0 < i \leq N$) then nontrivial, diagonal NSR's are again admissible. In this case, a δ_i can be treated as a free parameter in estimation by normalizing the i^{th} element of a β_j (for any j for which β_j has a non-zero i^{th} element) to a non-zero constant.

A third case arises under the null hypothesis when one (or more) of the δ_i is restricted to be zero. With $\delta_i = 0$, a normalization is still needed to preclude a nontrivial rescaling of Y_i. As above, the i^{th} element of one of the β_j could be normalized to unity, for example.

For example, Chen (1996) and Balduzzi, Das and Foresi (1996), in their $Ar(3, 2)$ models, assumed that $\Sigma_z = I$ and the volatility of r was $\sqrt{\beta_{z1}' Z(t)}$,

[9] This remark applies as well to the non-affine model in Andersen and Lund (1996), since the equivalence of the representations of the AYD and Ar drifts does not depend on an affine volatility model.

with $\beta'_{z1} = (0, 0, 1)$. Thus, they normalized the third element of β_{z1} to unity. It is easily verified that this is equivalent to normalizing the third element of β_1 to unity in the equivalent $AYD(3,2)$ specification of the volatility of Y_1. From the preceding remarks, it follows that this normalization identifies δ_3 in the $AYD(3,3)$ model with $r(t) = Y_1(t) + Y_2(t) + \delta_3 Y_3(t)$. δ_3 can be treated as a free parameter in estimation and the restriction $\delta_3 = 0$ that they imposed is therefore testable.

II.D Restrictions on Σ and $S(t)$

Consider again the general $AYD(N, n)$ model. The normalizations on $\mathcal{K}$ and δ do not assure identification of the parameters governing the diffusion coefficient of Y. Additional normalizations can be imposed to achieve identification of "maximally flexible" $AYD(N, n)$ representations. However, such normalizations will not in general guarantee that the terms $\alpha_i + \beta_i Y(t)$ in $S(t)$ are positive with probability one, a condition that is required in affine models for the existence of a strong solution to (6). Existence of a solution to the model typically requires the imposition of restrictions that are not required by the standard conditions for identification. Consequently, there is not a single equivalence class of just-identified affine models, but rather multiple branches of non-nested, identified models that satisfy the existence conditions in different ways. We begin this section with some general remarks about identification of the parameters in $S(t)$ and Σ, and then illustrate the practical implications of these observations for the three-factor model that will be examined empirically.

The diffusion is parameterized in terms of the product $\Sigma S(t)$ so normalizations must also be imposed to rule out rescalings by non-singular diagonal matrices. We choose to normalize the diagonal elements of Σ to be 1. All subsequent discussion of the identification of Σ and $S(t)$ presumes this normalization. One could instead fix the scales of β_i, $i = 1, 2, \ldots, N$.

There is a more subtle scale-invariance that arises in some special cases of affine models. Notice that the PDE's that determine bond prices and the conditional density of the state variables depend on the diffusion only through the combination $(\Sigma S)(\Sigma S)'$, and $(\Sigma S)(S\lambda)$. It follows that a new model with ΣS replaced by $\Sigma S U$, and $S\lambda$ replaced by $U' S\lambda$, for an arbitrary unitary matrix U, will produce the same observable predictions. (A unitary matrix U has the property that $U' = U^{-1}$.) If U commutes with S (i.e., $US = SU$), then a new affine model defined by replacing Σ by ΣU and λ by

$U'\lambda$ is observationally equivalent to the original affine model. We call such a transformation a unitary rotation (UR). The requirement $U'U = I_{N\times N}$ imposes $N(N+1)/2$ restrictions on a UR, so U has at most $N(N-1)/2$ free parameters. The requirement that U commute with S further restricts the free parameters in U. If S is completely unrestricted, then the only unitary matrix that commutes with S is the identity matrix. In this case, no further normalizations on Σ are necessary. On the other hand, if S is the identity matrix, as would be the case in Gaussian diffusions, then any unitary matrix commutes with S. In this case, we need to impose $N(N-1)/2$ restrictions on Σ. A convenient normalization for the Gaussian model is to restrict Σ to be upper (or lower) triangular. An intermediate case where a normalization is necessary to preclude nontrivial UR's is when $[S(t)]_{ii}$ and $[S(t)]_{jj}$ are proportional for some $i \neq j$.

In addition to ruling out rotations and rescalings, a parameterization of $S(t)$ and Σ must be such that the affine model is well defined. Duffie and Kan (1996) derived what are essentially necessary conditions for the existence of a strong solution to the stochastic differential equations in (6). Condition A in Duffie and Kan (1996) is reproduced here as Condition 1 (the Existence Condition):

Condition 1 (Existence) *For all i with $\beta_i \neq 0$*
(a) For all Y such that $[S(t)]_{ii} = 0$, $\beta_i{}'\mathcal{K}(\Theta - Y) > \beta_i{}'\Sigma\Sigma'\beta_i/2$
(b) For all j, if $(\beta_i{}'\Sigma)_j \neq 0$, then $[S(t)]_{ii}$ and $[S(t)]_{jj}$ are proportional.

The requirements of the Existence Condition may significantly influence the extent to which $S(t)$ and Σ can be freely parameterized. Different assumptions about $\mathcal{B} = (\beta_1, \beta_2, \beta_3)$ and Σ may interact with the Existence Condition to lead down different, non-nested branches within the class of admissible affine models. Following are three illustrative examples:

Example 1

If all the elements of β_i are free, then Condition 1(b) requires that $\mathcal{B}'\Sigma$ be diagonal. Since the diagonal elements of Σ are normalized to 1, this requirement fixes all of the off-diagonal elements of Σ in terms of the free parameters in $\mathcal{B}$.

Example 2

If $\mathcal{B}$ is diagonal, as in multi-factor CIR models, then Σ must be diagonal. Thus, the only admissible extension of the square root model is to allow the *volatility of each factor to have the form $\sqrt{\alpha_i + \beta_{ii} Y_i(t)}$. Correlation among the factors violates the existence conditions.*

Example 3

If the volatilities of two or more of the state variables are proportional, then some of the off-diagonal elements of Σ can be unconstrained parameters.

The third example is of particular interest, because a typical specification of Ar models has v being the volatility of r: $[S_z(t)]_{11} = \sqrt{v}$ and $[S_z(t)]_{33} = \eta\sqrt{v}$. To illustrate the flexibility in parameterizing Σ provided by this proportionality restriction in $S_z(t)$, consider the following $AYD(3,2)$ model with $r(t) = Y_1(t) + Y_2(t)$:

$$dY(t) = \mu_Y(t)dt + \Sigma S(t)dW(t), \tag{20}$$

where

$$\mu_Y(t) = \begin{pmatrix} \kappa(0 - Y_1(t)) \\ \nu(\bar{\theta} - Y_2(t)) \\ \mu(\bar{v} - Y_3(t)) \end{pmatrix}, \tag{21}$$

$S_{ii} = \sqrt{\alpha_i + \beta_i' Y(t)}$, $\alpha = (0, \alpha_2, 0)'$, and

$$\Sigma = \begin{pmatrix} 1 & \sigma_{12} & \sigma_{13} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} \beta_1' \\ \beta_2' \\ \beta_3' \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & b_{22} & 0 \\ 0 & 0 & b_{33} \end{pmatrix}. \tag{22}$$

The normalizations in this system were chosen as follows: The zero long-run mean of the first state variable in (21) and $\alpha_3 = 0$ are normalizations that preclude non-trivial shifts in Y. (With these normalizations, $\bar{\theta}$ and $\bar{v}$ are free parameters in (21).) α_1 is zero, because of the proportionality between the first the third diagonal elements of $S(t)$.

In the matrix Σ, $\sigma_{32} = \sigma_{21} = \sigma_{23} = 0$ is a requirement of the existence conditions $[\mathcal{B}'\Sigma]_{12} = [\mathcal{B}'\Sigma]_{21} = [\mathcal{B}'\Sigma]_{23} = 0$, because $[S(t)]_{11}$ and $[S(t)]_{33}$ are not proportional to $[S(t)]_{22}$. $\sigma_{31} = 0$ is a normalization that precludes non-trivial UR's. It follows that further relaxation of the fixed parameters in $\mathcal{B}$ and Σ_z, while preserving identification and satisfying Condition 1(b), is not possible.

An equivalent $Ar(3,2)$ model is obtained by applying the the transformation

$$Z(t) = LY(t) + L\vartheta, \tag{23}$$

where

$$L = \begin{pmatrix} 1 & 1 & 0 \\ 0 & \frac{\kappa-\nu}{\kappa} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \vartheta = \begin{pmatrix} -\frac{\nu}{\kappa-\nu}\bar{\theta} \\ \frac{\nu}{\kappa-\nu}\bar{\theta} \\ 0 \end{pmatrix}. \tag{24}$$

By construction, $Z_1(t) = r(t)$ and the drift of $Z(t)$ is the same as the drift in (16). Also, co-dependence among the Z's through the diffusion coefficients is governed by

$$\Sigma_z = \begin{pmatrix} 1 & \sigma_{r\theta} & \sigma_{rv} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{25}$$

$$\begin{pmatrix} \alpha_{z1} \\ \alpha_{z2} \\ \alpha_{z3} \end{pmatrix} = \begin{pmatrix} 0 \\ \hat{a} \\ 0 \end{pmatrix}, \begin{pmatrix} \beta'_{z1} \\ \beta'_{z2} \\ \beta'_{z3} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & \zeta^2 & 0 \\ 0 & 0 & \eta^2 \end{pmatrix}, \tag{26}$$

where $\sigma_{r\theta} = \frac{\kappa}{\kappa-\nu}(\sigma_{12}+1)$, $\sigma_{rv} = \sigma_{13}$, $\hat{a} = (\frac{\kappa-\nu}{\kappa})^2(\alpha_2 - \frac{\nu}{\kappa-\nu}\bar{\theta}b_{22})$, $\zeta^2 = \frac{\kappa-\nu}{\kappa}b_{22}$, and $\eta^2 = b_{33}$. Two special cases of this model are:[10]

- *Chen's "benchmark" model.* $\hat{a} = 0$, $\sigma_{r\theta} = 0$ and $\sigma_{rv} = 0$.
- *Balduzzi, Das, and Foresi model.* $\sigma_{r\theta} = 0$ and $\zeta = 0$.

All of these $AYD(3,2)$ (equivalently, $Ar(3,2)$) models can be nested in a less restrictive $AYD(3,3)$. As noted previously, the normalization $[\beta_1]_3 = 1$ allows δ_3 to be treated as a free parameter. Relaxing the constraint $\delta_3 = 0$ leads to a three-factor affine term structure model with $r(t) = Y_1(t) + Y_2(t) + \delta_3 Y_3(t)$. The $AYD(3,3)$ model described by (21) and (22) is the basis of our empirical analysis of affine models in the next section.

[10] In Chen's "benchmark" model, $\theta(t)$ – the "central tendency" factor – is assumed to follow a CIR process (by restricting $\hat{a}$ to zero), while in Balduzzi, Das and Foresi (1996), it is assumed to follow a Vasicek process (by restricting ζ in (26) to zero). Balduzzi, Das, and Foresi also allow for correlation between r and v ($\sigma_{r\theta} \neq 0$).

III Specification Tests of Affine Models

Empirical studies of Ar models of r have typically focused on fitting the short rate alone. This section describes an approach to evaluating the goodness of fit of any affine model, and presents the results from a specification analysis of the $Ar(3,2)$ model given by (25) and (26) and its $AYD(3,3)$ extension. Particular attention is given to the consequences of relaxing the assumptions on Σ_z and δ_3 in extant Ar models for describing the joint distribution of long- and short-term bond yields.

III.A Simulated Method of Moments Estimation

Let ψ_0 denote the parameter vector describing the AYD model of interest. Following Duffie and Singleton (1993) and Gallant and Tauchen (1996b), we compute a simulated method-of-moments (SMM) estimator of ψ_0. As Gallant and Tauchen (1996b) have recently shown, the scores of the likelihood function from an auxiliary model that describes the time series properties of bond yields or currency prices can serve 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 $f(y_t|x_{t-1}, \phi_0)$ 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 of ϕ_0 with sample size T (ϕ_T) satisfies

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\phi}\log f(y_t|x_{t-1}, \phi_T) = 0. \tag{27}$$

Under suitable regularity conditions (see Duffie and Singleton (1993) and Gallant and Tauchen (1996b)), as sample size gets large the sample mean in (27) converges to $E[\partial \log f(y_t|x_{t-1}, \phi_0)/\partial\phi]$. It follows that, if the asset pricing model is correctly specified, then the sample mean of the score evaluated at y's simulated from the asset pricing model ($\hat{y}_\tau$),

$$\frac{1}{\mathcal{T}}\sum_{\tau=1}^{\mathcal{T}}\frac{\partial}{\partial\phi}\log f(\hat{y}_\tau|\hat{x}_{\tau-1}, \phi_T), \tag{28}$$

where $\mathcal{T}$ is the simulation size, should also be approximately zero. Thus, by choosing the estimates of the term structure model to make the sample mean in (28) as close to zero as possible, we obtain estimates of the affine term structure model.

The requirements for the SMM estimator to be consistent for ψ_0, beyond the requirement that the auxiliary model have at least as many unknown parameters as the dimension of ψ_0, 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. To select an auxiliary model, we used the Semi-Non-Parametric (SNP) framework proposed by Gallant and Tauchen (1996b). Under plausible regularity conditions, an SNP auxiliary model can approximate arbitrarily well the joint conditional distribution of discretely sampled bond yields. Gallant and Long (1997) show that, for our term structure model and selection strategy for an auxiliary density $f(y_t|x_{t-1}, \phi_0)$, the SMM estimator is asymptotically efficient.[11] That is, we achieve the efficiency of the maximum likelihood estimator for the true conditional distribution of (discretely sampled) bond yields implied by the $AYD(3,3)$ model. It follows that our SMM estimator is more efficient (asymptotically) than the quasi-maximum likelihood estimator proposed recently by Fisher and Gilles (1996).

For our illustrations, y was chosen to be the yields on six-month LIBOR and two-year and ten-year fixed-for-variable rate swaps over the sample period April 3, 1987 to August 23, 1996. 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.). Duffie and Singleton (1996) found, for a somewhat shorter sample period, that a two-factor CIR model did not simultaneously describe all three of these yields. One outcome of the subsequent empirical analysis is an assessment of the adequacy of the $Ar(3,2)$ model (25)-(26) or its $AYD(3,3)$ extension as a description of the swap term structure.

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

[11]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.

polynomial functions of the yields in order to accommodate non-normality of the conditional distribution. After examining the properties of several auxiliary models [see Section III.B], we selected the auxiliary model with the following conditional density for our empirical analysis:

$$f(y_t|x_{t-1}, \phi_0) = c(x_{t-1})\ \left[\epsilon_0 + [h(z_t|x_{t-1})]^2\right]\ n(z_t), \tag{29}$$

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

$$z_t = R_{x,t-1}^{-1}(y_t - \mu_{x,t-1}). \tag{30}$$

The shift vector $\mu_{x,t-1}$ is assumed to be linear with elements that are functions of $L_\mu = 1$ lags of y,

$$\mu_{x,t-1} = \begin{pmatrix} \psi_1 + \psi_4\, y_{1,t-1} + \psi_7\, y_{2,t-1} + \psi_{10}\, y_{3,t-1} \\ \psi_2 + \psi_5\, y_{1,t-1} + \psi_8\, y_{2,t-1} + \psi_{11}\, y_{3,t-1} \\ \psi_3 + \psi_6\, y_{1,t-1} + \psi_9\, y_{2,t-1} + \psi_{12}\, y_{3,t-1} \end{pmatrix}. \tag{31}$$

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

$$R_{x,t-1} = \begin{pmatrix} \tau_1 + \tau_7\,|\epsilon_{1,t-1}| & \tau_2 & \tau_4 \\ +\tau_{25}\,|\epsilon_{1,t-2}| & & \\ 0 & \tau_3 + \tau_{15}\,|\epsilon_{2,t-1}| & \tau_5 \\ & +\tau_{33}\,|\epsilon_{2,t-2}| & \\ 0 & 0 & \tau_6 + \tau_{24}\,|\epsilon_{3,t-1}| \\ & & +\tau_{42}\,|\epsilon_{3,t-2}| \end{pmatrix} \tag{32}$$

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 $\Sigma_{x,t-1} = R_{x,t-1}\, R'_{x,t-1}$.

More complex conditional densities are accommodated by scaling $n(z_t)$ by the square of the Hermite polynomial $h(z_t|x_{t-1})$. In general, h is a polynomial of order K_z in z_t, with coefficients that are polynomials of order K_x in x_{t-1} and the conditioning information x_{t-1} consists of L_p lags of y_t. We set $L_p = 1$, so that the conditioning information is $x_{t-1} = y_{t-1}$. Additionally, we set

$K_z = 4$, with all of the interaction terms suppressed, and $K_x = 0$. With these choices, our h depends only on z and can be represented as:

$$h(z_t|x_{t-1}) = A_1 + \sum_{l=1}^{4}\sum_{i=1}^{3} A_{3(l-1)+1+i}\, z_{i,t}^{l} \tag{33}$$

The normalizing constant $c(x_{t-1})$ is the inverse of the integral over y_t of the product of $[h(z_t)]^2$ and $n(z_t)$.

The state variables are simulated using the Euler approximation of the stochastic differential equation governing the state dynamics. We use five subintervals for each week, and take every fifth simulated observation to construct a simulated data set of size 50000. The values of the simulated states are adjusted, if necessary, so that the $\alpha_i + \beta_i Y(t)$ are always nonnegative. Furthermore, the requirements of the Existence Condition are explicitly imposed.[12]

Gallant and Tauchen (1996b) showed that the simulated SNP scores (i.e., the SNP score function evaluated at the converged parameter values of the SNP parameters and the converged SMM estimators of the structural parameters) are asymptotically 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.[13]

[12]The Existence Condition can be reduced to a set of state-independent constraints on the model parameters. Details on how these constraints are implemented for the models studied here may be requested directly from the authors.

[13]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. Though ϵ_0 is identified if the scale of $h(z|x)$ is fixed, Gallant and Long (1997) encountered numerical instability in estimating SNP models with ϵ_0 treated as a free parameter. Therefore, we chose to fix both ϵ_0 and the constant term of $h(z|x)$ at non-zero constants. It appears that ϵ_0 was set to zero in previous implementations of the SNP model. However, this choice may introduce numerical problems in SMM estimation, because the SNP density is not guaranteed to be positive definite with $\epsilon_0 \neq 0$. In our implementations of SMM, we often found that some simulated observations were close to the zeros in the density function. In such cases (even if it is only for one simulated observation), the SNP scores became nearly singular. This, in turn, caused spurious random spikes in the SMM objective function. This problem was eliminated by setting ϵ_0 to a positive number that is sufficiently small to leave the estimated parameters of the auxiliary model essentially unchanged. All of the empirical results reported in this paper are obtained with $\epsilon_0 = 0.01$.

III.B Picking the Auxiliary Model

In selecting our SNP models, we sought to accommodate known features of both the conditional density of y implied by the $AYD(3,3)$ model and of the empirical distribution of swap yields, while adhering to the principal of parsimony. There were several considerations that influenced our final choice of SNP model. For ease of notation in the subsequent discussion, we summarize SNP models in terms of the notation $sL_\mu L_r L_p K_z I_z K_x I_x$.[14]

Consideration 1 (Conditional Means) *An implication of the assumption that the state vector Y follows the affine diffusion (6) is that its conditional mean is linear*[15]

$$E[Y(t)|Y(t-1),Y(t-2),\ldots] = C_o + e^{-\mathcal{K}}Y(t-1), \tag{34}$$

where C_o is a vector of constants.

This model-implied property of Y, and the empirical observation that most of the serial dependence of swap yields is well described by a first-order VAR, motivate our choice of $L_\mu = 1$.

The $AYD(3,3)$ model assumes that the eigenvalues of $\mathcal{K}$ are positive real numbers. This implies that the roots of the polynomial $I - e^{-\mathcal{K}}$ in (34) are real and lie outside the unit circle. The roots of $I - e^{-\mathcal{K}}$, evaluated at the SMM estimates, did in fact lie outside the unit circle. However, in many of the SNP models with $L_\mu = 1$, including the chosen $s1214300$ auxiliary model, the roots were complex. This is not necessarily inconsistent with the $AYD(3,3)$ model, because there is a nonlinear transformation between the state variables and swap yields. Whether or not it is consistent, allowing for complex roots provided an additional dimension along which the swap data could inform us about the validity of the family of affine term structure models we consider. That is, a central role for complex roots in characterizing

[14] L_μ is the order of the autoregressive specification of the conditional mean; L_r is the order of the ARCH specification of the conditional variance; L_p is the number of lags included in the conditioning set for the Hermite polynomials; K_z is the order in z of the polynomial $h(z|x)$, with positive value indicating non-Gaussian behavior; positive values of I_z indicate suppression of cross-terms in z; K_x is the order in x of the polynomial $h(z|x)$, with positive values indicating heterogeneity in the conditional density; finally, positive values of I_x indicate suppression of cross-terms in x.

[15] See, for example, Fisher and Gilles (1996).

the conditional distribution of y may be manifested in larger values of the goodness-of-fit statistics.

Most of the SNP models estimated with $L_\mu = 2$ implied non-stationary roots in the transformation from y to z. This is another reason we did not pursue SNP models with $L_\mu > 1$ any further.

There are at least two reasons why the linear structure (34) may not, in fact, be a good approximation to the conditional means of the swap yields y. First, swap yields are related to zero-coupon yields according to the expression (Duffie and Singleton (1996))

$$y_t^n = \frac{1 - P(t,n)}{\sum_{j=1}^{2n} P(t,.5j)} \tag{35}$$

and, hence, they are not linear functions of Y. Of course swap yields can be approximated by linear functions of Y with state-variable durations as weights. Our empirical findings suggest that this may be a reasonably good approximation for characterizing the properties of Y. However, it is an approximation that is not imposed *a priori*.

Second, the true conditional means of the swap yields and underlying zero yields may be nonlinear. That is, the affine term structure model may be mis-specified. Evidence for nonlinearity in univariate models was presented in Ait-Sahalia (1996). Andersen and Lund (1996), on the other hand, found little evidence for non-linearity in the drift of their multi-factor analysis of a short rate alone.

If $K_z = 0$, then the conditional mean of z_t, and hence of y_t, is linear in the SNP model, regardless of the order of K_x. Therefore, to accommodate a nonlinear conditional mean for y, we start by allowing $K_z > 0$. With $K_z > 0$ and $K_x = 0$, the structure of SNP model is that of an ARCH-in-Mean model, where the ARCH is of order L_r.

More complex forms of nonlinearity for the mean can be accommodated by having both K_z and K_x nonzero. However, with $K_z > 0$, incrementing K_x by one increases the number of free parameters in the SNP model by $3(3K_z + 1)$. Thus, parsimony suggests some caution in setting these parameters.

Consideration 2 (Conditional Second Moments) *The conditional variances of the state variables in the $AYD(3,3)$ model are time varying and conditional correlations of the diffusions are nonzero.*

Though the conditional variances of the state variables in affine models are linear functions of the current state, the pricing relation (35) gives little guidance on the structure of the conditional variances of the swap yields. To accommodate persistence in the volatilities of the swap rates, (32) includes a ARCH(2) transformation of the swap yields ($L_r = 2$).[16] Additionally, the assumption that $K_z > 0$ allows for more general dependence of the conditional variances on the lagged values of the swap yields than in (32), though still of the ARCH type. Again, complete generality is obtained by letting both K_z and K_x exceed zero.

Precisely how the relaxation of the assumption of uncorrelated diffusions in affine models will be manifested in the conditional distribution of y is difficult to say. The conditional correlations of the swap yields in the SNP model are influenced by the parameters of $R_{x,t-1}$ and of the polynomial $h(z)$. Therefore, we expect that freeing up the restrictions $\sigma_{r\theta} = 0$ and $\sigma_{rv} = 0$ will result in improved diagnostics with regard to the scores for the SNP parameters A_j and τ_j associated with $h(z)$ and $R_{x,t-1}$ (see Section III.C).

Consideration 3 (Non-normal Innovations) *Though the conditional means of discretely sampled affine diffusions are linear, the innovations are non-normal. For instance, it is well known that the innovations implied by the square-root model are non-central chi-square. To capture this non-normality, we choose* $K_z > 0$.

Given the importance of $K_z > 0$ for accommodating the non-normality of the normalized swap yields (z_t) and allowing for some non-linearity of the conditional mean of y, we set $K_z = 4$. To keep the number of free parameters manageable, we also set suppressed all of the cross terms in the polynomials in z. As noted above, setting $K_x > 0$ with $K_z > 0$ substantially increases the dimensionality of the parameter space. In fact, using the Schwarz model selection criterion (BIC), the SNP model s1214300 ($BIC = -3.93100$) was clearly preferred over s1214310 ($BIC = -3.778$). Therefore, we chose to set $K_x = 0$.

The deterioration in BIC with $K_x > 0$ was not a consequence of having $K_z = 4$. A similar deterioration occurred with $K_z = 2$ when K_x was increased from 0 to 1. This lends further support to our view that $K_x > 0$ was not essential for characterizing the conditional distribution of swap yields for our sample period. Nevertheless, to provide some assurance that our conclusions are insensitive to the assumption $K_x = 0$, we also fit all of the affine models

[16]Notice that the formulation of ARCH is in terms of the absolute values of the VAR innovations and not their squared values.

using the SNP model $s1111010$ ($K_z = 1, K_x = 1$). All of the specification tests were qualitatively identical to those with the model s1214300.

In the light of all of these considerations, we chose to report results for the auxiliary model s1214300. Maximum likelihood estimators for the parameter vector

$$\begin{aligned}\phi_0 &= (A_j : 2 \leq j \leq 13;\ \psi_j : 1 \leq j \leq 12; \\ &\quad \tau_j : j = 1, 2, \ldots, 7, 15, 24, 25, 33, 42)\end{aligned} \tag{36}$$

are given in Table I. Note that A_1 is normalized to 1. Several of the A_j parameters governing the non-normality of the "innovation" z_t and the τ_j governing conditional heteroskedasticity are significantly different from zero at conventional significance levels.

III.C Specification Tests

Table II presents the SMM estimates and the estimated standard errors for three nested affine models; Chen is nested in $AYD(3,2)$ which is nested in $AYD(3,3)$. The last row of the table gives the χ^2–statistic for an overall test of the model's ability to explain the over-identifying restrictions supplied by the score generator. The degrees of freedom for this statistic is equal to the number of free parameters in the auxiliary model minus the number of free parameters in the structural model.

Associated with each element of the parameter vector ϕ_0 of the auxiliary model for y is an element of the sample score vector (28). Under the null hypothesis that the affine model correctly prices the swaps, each of these sample scores should be close to zero when $\hat{y}$ is simulated at the estimated values of the parameters. Table III displays the values of these sample scores and the associated t–ratios for the null hypotheses that the scores are zero. The first column lists the free parameters of the auxiliary model (see (31) and (32)) that index the elements of the score vector (28).

Consider first the results for the Chen benchmark model. The value of the test statistic $\chi^2(26) = 129.027$ is large relative to its degrees of freedom, which suggests that this model does not adequately describe the full term structure of swap rates. The t–statistics for the scores in Table III, column 3 reveal that this model does a poor job of fitting the conditional means and variances and the non-normality of innovations of swap rates as captured by the SNP model. The t-statistics for half of the scores have absolute values larger than 2.. Among these eighteen statistics, three are associated non-normality of z (A_j), five are associated with the conditional mean (Φ_j), and eight are associated with the conditional second moments (τ_j). The model appears to have the greatest difficulty fitting the conditional second moments of swap rates.

Relaxing the zero restrictions $\sigma_{rv} = \sigma_{r\theta} = 0$ leads to a just-identified $Ar(3,2)$ (equivalently $AYD(3,2)$) model[17]. Results for this model are displayed in columns 4–5 of Tables II and III. The p-value of the χ^2 statistic (30.224 with 23 degrees of freedom) is 14.308%. Thus, the $AYD(3,2)$ model is not rejected at conventional significance levels. The t–statistics in column 5 of Table III show that the $AYD(3,2)$ model does a much better job of describing the conditional second moments of swap yields then the model with $\sigma_{rv} = \sigma_{r\theta} = 0$. The sample scores for the $AYD(3,2)$ model are uniformly smaller (except for A_8, which is already very small in the *Chen* model) relative to their estimated standard errors. Only three of the statistics have absolute values larger than 2 and none of these is associated with the SNP parameters characterizing the conditional second moments of swap rates. Thus, allowing for nonzero correlations among the diffusions helps not only in explaining the conditional second moments, but also in explaining the conditional first moments and the non-normality of z.

We also conducted a test of the Chen benchmark model against the alternative of the $AYD(3,2)$ model using the χ^2–test proposed by Newey and West (1987)and Gallant and Tauchen (1996a). Under the null, the difference between the minimized values of the SMM criterion functions for two nested models is asymptotically distributed as $\chi^2(q)$, where q is the number of parametric restrictions in the constrained model (the null). The Chen model is strongly rejected against the $AYD(3,2)$ model ($\chi^2(3) = 98.803$) and, by implication, against the $AYD(3,3)$ model.

Both the Chen benchmark and $AYD(3,2)$ models presume that Y_3 does not affect r directly; $\delta_3 = 0$ so that $r(t) = Y_1(t) + Y_2(t)$. The $AYD(3,3)$ offers a potential further improvement in fit over the $AYD(3,2)$ model by relaxing this constraint. From Table II it is seen that $\hat{\delta}_3 = 112.253$ with a standard error of 49.1.[18] Consistent with this finding, the difference between

[17] $\hat{a}$ is also relaxed. See Section III.D.3 for some relevant discussion.

[18] The scale of δ_3 is roughly two orders of magnitude larger than $\delta_1 = \delta_2 \equiv 1$, because (for ease of comparison with extant affine models) we normalized $[\beta_1]_3$ to 1. This value for $[\beta_1]_3$ is two orders of magnitude larger than the scale for the elements of β_j associated with the first and second state variables.

the χ^2 goodness-of-fit statistics for the $AYD(3,3)$ and $AYD(3,2)$ models (see Table II) is 3.769, which is significantly different from zero at the 5% confidence level.

In going from the most to the least restrictive affine model, we relaxed the covariance restrictions in Σ first and then relaxed the restriction $\delta_3 = 0$. The incremental declines in the overall χ^2 statistics suggests the the largest gain in fit came from relaxing the covariance restrictions. This is indeed the case as we find substantial evidence against the model with $\delta_3 = 0$, but σ_{rv} and $\sigma_{r\theta}$ nonzero.

III.D Interpreting the Results

In order to interpret the evidence against the over-identifying restrictions imposed in Ar models, it is instructive to examine in more depth the nature of the factors in the AYD and Ar representations of affine models. Toward this end we first relate the fitted Y's to movements in the historical swap rates. Then we exploit the fact that every Ar model has an equivalent AYD representation to interpret the state variables in the Ar model.

III.D.1 The Risk Factors Y

The factors in $AYD(2,2)$ models are typically interpreted as "level" and "slope" based either on the factor loadings in principal components analyses (e.g., Litterman and Scheinkman (1991)) or on the properties of the implied state variables (e.g., Duffie and Singleton (1996)). Litterman and Scheinkman (1991) found that their third principal component had loadings that are suggestive of a "curvature" factor. To confirm that the Y's in our $AYD(3,3)$ model have similar interpretations, we computed the implied state variables $\hat{Y}$ and compared the $\hat{Y}$ to various linear combinations of the swap yields (Figures 1–3).[19]

$\hat{Y}_2$ is plotted against *Level*, defined as the ten-year swap yield. $\hat{Y}_3$ is plotted against *Slope*, defined as the difference between the ten- and two-

[19]The implied state variables from a given model are the particular realizations of the state variables that let the model price the six-month, two-year and ten-year yields exactly, using the SMM parameter estimates. If the model is correctly specified, then the implied state variables associated with the model embody the correct assumed factor dynamics, except for sampling errors. We report results for the implied state variables computed using parameter estimates from the $AYD(3,3)$. The results are qualitatively the same when the $AYD(3,2)$ estimates are used.

year swap rates. The third observed factor we considered was $Butterfly$, defined as the residual from the regression of the two-year swap yield on the six-month LIBOR and ten-year swap yields. In order to compare $Butterfly$ with a comparable fitted state variable, we examined the residual from the regression of $\hat{Y}_1$ on $\hat{Y}_2$ and $\hat{Y}_3$. All time series are standardized by subtracting their means and scaling by their standard deviations.

Figure 1 shows that (the orthogonalized) $\hat{Y}_1$ is highly correlated with $Butterfly$ and, hence, represents a curvature factor.[20] From the other two figures we see that $\hat{Y}_2$ behaves like a *Level* factor, and $\hat{Y}_3$ behaves like a *Slope* factor. Thus, the three state variables are the dynamic counterparts to the risk factors typically identified in principal component analyses.

III.D.2 The State Variables in the Ar Models

Given an identified $AYD(3,3)$ model, our equivalence result implies the existence of an $Ar(3,3)$ model with a terraced drift structure: the second state variable is the long-run mean of r and the third state variable is the long-run mean of the second state variable. However, this particular Ar representation of the AYD model is not the most convenient for interpreting the state variables in extant Ar models, because in these models the second, central tendency factor does not have a stochastic long-run mean.

Fortunately, there is another equivalent $Ar(3,3)$ model that preserves this feature of extant $Ar(3,2)$ models and, thereby, facilitates interpretation of the factors. Namely, consider the Ar representation of the $AYD(3,3)$ model obtained by the transformation (23) with the first row of L set at $(1,1,\delta_3)$. This transformation maps $Y_2(t)$ to $\theta(t)$ and $Y_3(t)$ to $v(t)$. Moreover, the drift of $(\theta(t), v(t))$ is identical to that in (16), while the drift of r has an additional term due to δ_3 being nonzero,

$$\kappa(\theta(t) - r(t))\,dt + \delta_3(\kappa - \mu)v(t)\,dt. \tag{37}$$

Thus, one component of the drift of r has r mean reverting to a long-term swap rate – the *Level* of the swap curve. However, with $\delta_3 \neq 0$, $\theta(t)$ is not the long-run mean of r (see Section III.D.4). Perhaps most surprisingly,

[20]Within the $AYD(3,3)$ model, the correlation between $\Delta\hat{Y}_1$ and $\Delta Butterfly$ is -0.991, whereas the correlations between $\Delta\hat{Y}_1$ and Δ*Slope* and Δ*Level* are 0.633 and -0.472, respectively. The correlations of $\Delta\hat{Y}_2$ with $\Delta Butterfly$, Δ*Slope*, and Δ*Level* are 0.280, 0.065, and 0.969, respectively. $\Delta\hat{Y}_3$ has a correlation of -0.899 with Δ*Slope*, and correlations of 0.832 and 0.245 with $\Delta Butterfly$ and Δ*Level*, respectively.

$v(t) = Y_3(t)$, which is the volatility of the instantaneous short rate, closely tracks the slope of the swap curve!

Why is the third factor well proxied by the slope of the yield curve instead of a proxy for short-rate volatility as was presumed in extant specifications of Ar models? The identification of the $AYD(3,3)$ model was driven largely by the assumed structure of the diffusion matrices, which was dictated by our desire to nest previous $Ar(3,2)$ models as special cases. A key assumption in the $AYD(3,3)$, as well as the nested $AYD(3,2)$, models is that the correlation between $\theta(t)$ and $v(t)$ is zero. In our sample, the correlations of $\Delta Butterfly$ with $\Delta Slope$ and $\Delta Level$ are -0.567 and 0.423, respectively, while the correlation of $\Delta Level$ and $\Delta Slope$ is only -0.178. The relatively small correlation of $\Delta Level$ and $\Delta Slope$ suggests that $v(t) = Y_3(t)$ is proportional to either *Level* or *Slope*.

In addition to the zero correlation between v and θ, the conditional variances of r and v are proportional and the drift of r includes $\kappa(\theta(t) - r(t))\,dt$. Both of these assumptions affect the dynamic properties of the implied term structure and, thereby, influence the selection by the SMM criterion function of *Level* and *Slope* as the second and third state variables, respectively. With $\theta(t) \propto Y_2(t)$, the first term in (37) serves to pull the instantaneous short rate toward the long-term rate. This is consistent with previous $Ar(3,2)$ models that had r mean reverting to a process that itself is slowly mean reverting (see, e.g., Andersen and Lund (1996)). Furthermore, $Y_1(t) = [(r(t) - Y_2(t)) - \delta_3 Y_3(t)]$, and $(r(t) - Y_2(t))$ behaves much like minus the slope between the ten-year and instantaneous swap rates. Thus, if $Y_3(t)$ closely tracks the $2-10$ slope of the swap curve, then $Y_1(t)$ will behave much like (minus) a linear combination of two measures of the slope of the swap curve. Thus, these assignments assure that, within the equivalent $AYD(3,3)$ model, the volatilities of $Y_1(t)$ and $Y_3(t) = v(t)$ will be approximately proportional.

Clearly, the properties of the joint conditional distribution of the level, slope, and curvature factors play a central role in these interpretations of θ and v, and in whether or not this particular branch of the affine family of models describes the U.S. swap curve. Singleton (1994), for example, found that the properties of the level and slope risk factors were very different for Japanese and U.S. government bond markets. Consistent with this finding, we would expect that different institutional and macro-economic conditions will lead one to explore different branches of the affine class.

III.D.3 Evidence Against the Covariance Restrictions

These observations "explain" the large chi-square statistics obtained when the covariance restrictions $\sigma_{r\theta} = \sigma_{rv} = 0$ are imposed. The zero restrictions essentially state that the instantaneous short rate is conditionally uncorrelated with the level (long-term rate) and slope of the swap yield curve. In the true probability model, these correlations are evidently non-zero, which is not surprising.

Though not stressed in our preceding discussion, the $Ar(3,3)$ model also relaxes the assumption that $\hat{a} = 0$. The results in Table II suggest that the estimated value of this parameter is significantly different from zero at conventional significance levels. This implies that the *Level* factor $\theta(t)$ does not follow a standard square-root diffusion. Rather, its volatility is of the extended square-root form $\sqrt{\hat{a} + \zeta^2\theta(t)}$. The extension, however, is responsible for only a small amount of the reduction in χ^2. The rejection of the Chen model is due entirely to the covariance restrictions.

III.D.4 Evidence Against the Null Hypothesis $\delta_3 = 0$

One interpretation of the evidence that $\delta_3 \neq 0$ is provided by the equivalent $Ar(3,3)$ with a terraced drift structure. Whereas extant $Ar(3,2)$ models assume that the long-run mean of the central tendency is constant, the evidence suggests that a better model for the swap curve has the third state variable mean reverting to the central tendency of r. Alternatively, within the $Ar(3,3)$ representation in which the drift of r is (37) and the drifts of $\theta(t)$ and $v(t)$ are constants, $\delta_3 \neq 0$ implies that the third state variable affects the drift of r. The key feature shared by both of these Ar representations is that $v(t)$ affects r through the drift of the multivariate diffusion $(r(t), \theta(t), v(t))$ and not just through the specification of volatility. In other words, $v(t)$ is not a pure volatility factor, but rather has a direct effect on r. Such a direct effect of all three state variables on r is assumed in multi-factor CIR-style models.

III.E Model Diagnostics with Simulated Moments

There is a potentially important difference between the specification tests conducted here and those based on the implied bond yields in previous studies of affine models. In the case of *CIR*-style models, square-root diffusions were estimated by maximum likelihood, the model was "inverted" to obtain fitted state variables as functions of the data and the maximum likelihood parameter estimates, and then the moments of the implied bond yields were compared to the corresponding moments of the actual yields (the data). For example, Pearson and Sun (1994) and Duffie and Singleton (1996) assess the goodness-of-fit in their models by regressing actual bond yields on the implied bond yields and testing whether or not the intercept is zero and slope coefficient is unity.

The diagnostics discussed in Section III.C are all based on simulated moments. Using the SMM parameter estimates, long time-series of the state variables are simulated, and the associated values of the bond yields are computed using the affine pricing model. Then, the scores of the sample log-likelihood function of the auxiliary model are computed using simulated yields and compared to zero, their population value if the model is correct.

In the context of an N-factor affine model in which N of the bond yields are assumed to be priced exactly, the implied state variables will, by construction, exactly price N of the bond yields.[21] So the empirical and implied distributions of these N yields must be identical. Moreover, when the number of yields M is larger than N, the information in the data enters the implied distributions of the other $M - N$ yields in two ways: indirectly through the ML estimates of the model, and directly through the inversion of the model, observation by observation, to compute the implied state variables from the N yields that are fit exactly.

In contrast, the empirical and simulated distributions will generally differ for all M yields. This is because the information in the actual data enters only indirectly through the SMM parameter estimates. The values of the simulated moments are otherwise determined only by the structure of the state-variable process and the choice of risk premiums. For the purpose of evaluating the characteristics of the distributions of bond yields implied by an affine model, it is the simulated distribution that is most relevant. At a practical level, it is a close correspondence between the simulated and actual distributions that is desirable for pricing options on bonds by Monte Carlo.

Another potentially informative use of simulated bond yields is an assessment of the effects of relaxing parameter constraints on the distributions of

[21] This was true of the models in Chen and Scott (1993), Pearson and Sun (1994), and Duffie and Singleton (1996), and is also true of our model as the number of state variables equals the number of bond yields.

yields implied by the model. We illustrate this possibility, as well as the fact that simulated and empirical distributions are different even with $M = N$, by examining the mean swap rates. Figure 4 plots the means of the simulated swap rates for the *Chen*, $AYD(3,2)$, and $AYD(3,3)$ models against the observed mean swap yield curve. Consistent with the overall goodness-of-fit statistics, the differences between sample and simulated mean swap rates are smallest for the $AYD(3,3)$ model and largest for the *Chen* model.

While the average ten-year yield is fit almost exactly, the models overstate the average slope of the swap yield curve. This is evidently linked to the dual role of the third state variable as both the "volatility" factor and the "slope" factor. Since the volatilities are more precisely estimated than the average yields, the SMM objective function assigns more weight to fitting the moment conditions associated with the conditional variance than those associated with the conditional mean. This weighting influences the estimated value of the long-run mean of the third state variable, $\bar{v}$. The flexibility from relaxing the correlation restrictions in the *Chen* model allows $\bar{v}$ to increase (from 0.0002456 to 0.0004378), since the increased variance of the short rate induced by a higher $\bar{v}$ is offset by the (negative) covariance of the short rate with other state variables. Since a higher $\bar{v}$ implies a flatter yield curve, relaxing the correlation restrictions leads to a better fit to the average slope.

Relaxing the constraint $\delta_3 = 0$ in the $AYD(3,2)$ model adds a direct contribution of $v(t)$ to the drift and the conditional variance of the short rate. These effects, in combination, lead on average to a further flattening of the swap curve.

The remaining gap in Figure 4 between the means of the simulated and actual two-year swap rates for the $AYD(3,3)$ model suggests that this model does less well at explaining the level of the two-year rate compared to the ten-year rate. This may underlie the relatively large values of the simulated SNP scores associated with ψ_2, ψ_5, ψ_8, and ψ_{11} (see Table III). All of these SNP parameters are associated with the conditional mean of the two-year swap rate.

III.F Level Effects on Volatility

Though the parameterization of the Ar models examined here do not allow r to directly affect its volatility ($[S_z(t)]_{11} = \sqrt{v(t)}$), there is nevertheless a "level effect" of r on volatility in these models. The level effect enters indirectly through the correlation among the state variables. More precisely, the conditional variance of r, $\sigma_r^2(t)$, implied by the $Ar(3,2)$ model is[22]

$$\sigma_r^2(t) = \sigma_{r\theta}^2 \hat{a} + (1 + \sigma_{rv}^2 \eta^2) v(t) + \sigma_{r\theta}^2 \zeta^2 \theta(t). \tag{38}$$

The level effect of the short rate on its own volatility in this Ar model takes the form of a non-zero weight on $\theta(t)$, the long-run mean of r. Increases in the long-run mean imply an increase in short-rate volatility.

Evaluating the weights on $v(t)$ and $\theta(t)$ in (38) at the point estimates from the $Ar(3,2)$ model gives $2.44v(t)$ and $0.00072\theta(t)$. The standard deviations of the implied $v(t)$ and $\theta(t)$ are 0.00030 and 0.01406, respectively. Combining these observations, it follows that the variances of $(1+\sigma_{rv}^2\eta^2)v(t)$ and $\sigma_{r\theta}^2\zeta^2\theta(t)$ are 5.35×10^{-7} and 1.04×10^{-10}, respectively. We conclude that, though the model accommodates a level effect through the non-zero correlations, the volatility of r is driven almost entirely by the third state variable $v(t)$. Different branches of affine models satisfying the existence condition might, of course, show stronger or weaker level effects than this particular class of models.

[22] The relation (38) is the $1-1$ element of the matrix $\Sigma_z S_z(t) S_z(t)' \Sigma_z'$. Similar results are obtained using the estimates from the $AYD(3,3)$ model.

IV Concluding Remarks

In this paper we placed two important strands of the empirical term structure literature on a common footing by showing that models that focus directly on the representation of the short rate are equivalent to models that describe the short rate as a linear combination of unobserved states. This facilitated comparisons and interpretations of the affine models in the literature. Furthermore, we addressed the identification problem for general affine models in order to assess the extent to which the over-identifying restrictions typically imposed can be relaxed. In applying our results to an evaluation of popular three-factor affine models of the instantaneous short rate, we found that the over-identifying restrictions 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 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. The diagnostic evidence suggests that relaxation of the restrictions on the conditional second moments of the state variables is important for simultaneously explaining movements in the short and long ends of term structures. In addition, the findings suggest that the drift of the instantaneous short rate is more complicated than simply the short rate mean reverting to a stochastic long-run mean.

We also highlighted the interplay between conditions for econometric identification of affine models and conditions for the existence of solutions to the PDE describing bond prices. Different specifications of the diffusion coefficients may lead to distinct non-nested families of affine models. To make this point more concretely, consider the the over-identifying restrictions on the diffusion matrices within the family of three-factor models that we examined empirically. The most flexible affine model examined relaxed these constraints as much as was possible, given proportionality of the variances of the first and third state variables and the requirements of the existence conditions. The reported diagnostics suggested that relaxing the constraints substantially improved the goodness of fit. Greater flexibility in fitting the joint distribution of swap rates can not be achieved within this particular class of affine models. The comparative properties of affine models along different branches will be explored in future research.

Appendices

A Proofs

A.1 Proof of Proposition II.1

To show that the bond pricing formula is invariant under an SPR, it suffices to show that $A(\tau)$ and $B(\tau)$ transform as

$$A(\tau) \rightarrow \hat{A}(\tau) = A(\tau), \tag{39}$$

$$B(\tau) \rightarrow \hat{B}(\tau) = (X')^{-1}\, B(\tau); \tag{40}$$

i.e., $\hat{A}(\tau)$ and $\hat{B}(\tau)$ defined above satisfy the Ricatti equations for $\hat{Y}(t) = X \times Y(t)$. Notice that

$$\begin{aligned} & \frac{d\hat{A}(\tau)}{d\tau} - \left[-\hat{\tilde{\theta}}'\, \hat{\tilde{\mathcal{K}}}'\hat{B}(\tau) + \frac{1}{2}\sum_{i=1}^{N} \left[\hat{\Sigma}'\hat{B}(\tau)\right]_i^2 \hat{\alpha}_i \right] \\ = \; & \frac{dA(\tau)}{d\tau} + (X\tilde{\theta})'\,(X\mathcal{K}X^{-1})'\,X'^{-1}\,B(\tau) \\ & - \frac{1}{2}\sum_{i=1}^{N} \left[(X\Sigma)'X'^{-1}\,B(\tau)\right]_i^2 \alpha_i \\ = \; & \frac{dA(\tau)}{d\tau} - \left[-\tilde{\theta}'\,\mathcal{K}\,B(\tau) + \frac{1}{2}\sum_{i=1}^{N} [\Sigma'\,B(\tau)]_i^2\, \alpha_i \right] = 0, \end{aligned} \tag{41}$$

$$\begin{aligned} & \frac{d\hat{B}(\tau)}{d\tau} - \left[\hat{\tilde{\mathcal{K}}}'\,\hat{B}(\tau) - \frac{1}{2}\sum_{i=1}^{N} \left[\hat{\Sigma}'\hat{B}(\tau)\right]_i^2 \hat{\beta}_i + \hat{\delta} \right] \\ = \; & \frac{d\,(X'^{-1}\,B)(\tau)}{d\tau} + (X\tilde{\mathcal{K}}X^{-1})'\,X'^{-1}\,B(\tau) \\ & + \frac{1}{2}\sum_{i=1}^{N} \left[(X\Sigma)'\,X'^{-1}\,B(\tau)\right]_i^2 X'^{-1}\,\beta_i - X'^{-1}\,\delta \\ = \; & X'^{-1}\left[\frac{dB(\tau)}{d\tau} - \left[-\tilde{\mathcal{K}}'\,B(\tau) - \frac{1}{2}\sum_{i=1}^{N} [\Sigma'\,B(\tau)]_i^2\, \beta_i + \delta \right]\right] = 0, \end{aligned} \tag{42}$$

where we have used the fact that $A(\tau)$ and $B(\tau)$ satisfy the Ricatti equations under the original model, and the model parameters transform in the following manner[23]

$$\begin{aligned} \mathcal{K} \to \hat{\mathcal{K}} = X\mathcal{K}X^{-1}, \quad \tilde{\mathcal{K}} \to \hat{\tilde{\mathcal{K}}} = X\tilde{\mathcal{K}}X^{-1}, \quad \Theta \to \hat{\Theta} = X\Theta, \\ \tilde{\theta} \to \hat{\tilde{\theta}} = X\tilde{\theta}, \qquad \Sigma \to \hat{\Sigma} = X\Sigma, \qquad \alpha \to \hat{\alpha} = \alpha, \\ \beta' \to \hat{\beta}' = \beta' X^{-1}, \qquad \lambda \to \hat{\lambda} = \lambda, \qquad \delta' \to \hat{\delta}' = \delta' X^{-1}. \end{aligned} \tag{43}$$

A.2 Proof of Proposition II.2

First, the conditional density $\hat{f}(\hat{y}|\hat{x})$ of the transformed model satisfies the Kolmogorov forward equation:

$$\begin{aligned} &\frac{\partial}{\partial t}\hat{f}(\hat{y}|\hat{x}) + \left[\hat{\mathcal{K}}(\hat{\Theta} - \hat{y})\right]' \frac{\partial}{\partial \hat{y}}\hat{f}(\hat{y}|\hat{x}) \\ &+ \frac{1}{2}Tr\left[(\hat{\Sigma}\hat{S})(\hat{\Sigma}\hat{S})' \frac{\partial^2 \hat{f}(\hat{y}|\hat{x})}{\partial \hat{y}\partial \hat{y}'}\right] = 0. \end{aligned} \tag{44}$$

Changing the variables from $\hat{y}$ to $y = X^{-1}\hat{y}$ and from $\hat{x}$ to $x = X^{-1}\hat{x}$, defining $g(y|x) = \hat{f}(\hat{y}|\hat{x}) = \hat{f}(Xy|Xx)$, and using (43), we obtain

$$\begin{aligned} &\frac{\partial}{\partial t}g(y|x) + \left[\mathcal{K}(\Theta - y)\right]' \frac{\partial}{\partial y}g(y|x) \\ &+ \frac{1}{2}Tr\left[(\Sigma S)(\Sigma S)' \frac{\partial^2 g(y|x)}{\partial y\partial y'}\right] = 0, \end{aligned} \tag{45}$$

which is the PDE satisfied by the conditional density $f(y|x)$ under the original model. It follows that $f(y|x) = C\,g(y|x)$, where the constant C is determined by the requirement that the density integrate to unity. In this case, since $\int g(y|x)\,d\hat{y} = 1$, we have $C = |X|$, where $|X|$ is the absolute value of the determinant of X. Thus,

$$\hat{f}(\hat{y}|\hat{x}) = |X|^{-1} f(X^{-1}\hat{y}|X^{-1}\hat{x}). \tag{46}$$

The conditional density of the observed zero-coupon bond prices (with maturities τ_1, τ_2 and τ_3), is given by

$$f_p(p_t|p_{t-1}) = J^{-1} f(y_t|y_{t-1}) \tag{47}$$

[23]It is not always necessary, but convenient for the sake of conformity to rescale $\hat{\Sigma}$, $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\lambda}$ so that the diagonal elements of $\hat{\Sigma}$ are unity.

where $J = p_t|(B(\tau_1), B(\tau_2), B(\tau_3))|$ is the Jacobian for the transformation from y_t to p_t, and y_t should be interpreted as the implied state variables here. We can now verify the invariance of f_p:

$$
\begin{aligned}
& J_{\hat{y}}^{-1}\, \hat{f}(\hat{y}_t|\hat{y}_{t-1}) \\
= \; & \left[p_t\, |(\hat{B}(\tau_1), \hat{B}(\tau_3), \hat{B}(\tau_3))|\right]^{-1} |X|^{-1} f(X^{-1}\hat{y}_t|X^{-1}\hat{y}_{t-1}) \\
= \; & \left[p_t\, |X|^{-1}\, |(B(\tau_1), B(\tau_2), B(\tau_3))|\right]^{-1} |X|^{-1} f(y_t|y_{t-1}) \\
= \; & J^{-1} f(y_t|y_{t-1}).
\end{aligned} \tag{48}
$$

B Equivalence of $AYD(N,n)$ and $Ar(N,n)$

B.1 From $AYD(N,n)$ to $Ar(N,n)$

An arbitrary $AYD(N,n)$ model may be parameterized as follows: $r(t) = \sum_{i=1}^{n} \delta_i Y_i(t)$,

$$dY(t) = \begin{pmatrix} \kappa_1 (\theta_1 - Y_1(t)) \\ \vdots \\ \kappa_N (\theta_N - Y_N(t)) \end{pmatrix} + \Sigma S(t) dW(t), \tag{49}$$

where $S_{ii}(t) = \sqrt{\alpha_i + \beta_i' Y(t)}$. Without loss of generality, we may normalize the δ_i's to 1, because we have assumed that they are non-zero. However, for the sake of generality, we will keep them in the formulae in this section.

To show that an $Ar(N,n)$ model characterized by (19) can be derived from the $AYD(N,n)$ model defined above, we first define the following. Let

$$Z(t) = \theta_z + LY(t), \tag{50}$$

where L is a block diagonal matrix[24]

$$L = \begin{pmatrix} L^{(n)} & 0_{n\times(N-n)} \\ 0_{(N-n)\times n} & I_{(N-n)\times(N-n)} \end{pmatrix}, \tag{51}$$

and $L_{ji}^{(n)} = \gamma_i^{(j)}$. For $i < j$, $1 \le j \le n$, $\gamma_i^{(j)} \equiv 0$. For $i \ge j$, $1 \le j \le n$, $\gamma_i^{(j)}$ are defined through the following recursion:

$$\gamma_i^{(1)} = \delta_i, \qquad \gamma_i^{(j)} = \gamma_i^{(j-1)}(1 - \frac{\kappa_i}{\kappa_{j-1}}),\ 2 \le j \le n. \tag{52}$$

Furthermore, $\theta_z = (\theta_{z1}, \theta_{z2}, \ldots, \theta_{zn}, 0, \ldots, 0)'$, where

$$\theta_{z1} \equiv 0,\ \theta_{zj} = \theta_{z\,j-1} + \sum_{i=j-1}^{n} \gamma_i^{(j-1)} \frac{\kappa_i}{\kappa_{j-1}} \theta_i,\ 2 \le j \le n. \tag{53}$$

A telescoping sum gives

$$\theta_{zj} = \sum_{l=1}^{j-1} \sum_{i=l}^{n} \gamma_i^{(l)} \frac{\kappa_i}{\kappa_l} \theta_i,\ 2 \le j \le n. \tag{54}$$

[24]Note that the partition of L and other matrices are conformal with the partition of the state vector into the n primary factors and the $N - n$ auxiliary factors.

It is now straightforward to show that the drift of $Z(t)$ defined in (17) is given by (19) by checking three cases.

First,

$$\begin{aligned}\mu_{Z_1(t)} &= \mu_{r(t)} = \sum_{i=1}^{n} \delta_i \kappa_i (\theta_i - Y_i(t)) \\ &= \kappa_1 \theta_{z2} - \kappa_1 Z_1(t) - \sum_{i=2}^{n} \delta_i (\kappa_1 - \kappa_i) Y_i(t) \\ &= \kappa_1 (Z_2(t) - Z_1(t))\end{aligned} \tag{55}$$

Secondly, for $1 \leq j \leq n-1$,

$$\begin{aligned}\mu_{Z_j(t)} &= \sum_{i=j}^{n} \gamma_i^{(j)} \mu_{Y_i(t)} \\ &= \sum_{i=j}^{n} \gamma_i^{(j)} \kappa_i \theta_i - \kappa_j (Z_j(t) - \theta_{zj}) + \sum_{i=j+1}^{n} \gamma_i^{(j)} (\kappa_j - \kappa_i) Y_i(t) \\ &= \kappa_j (Z_{j+1}(t) - Z_j(t))\end{aligned} \tag{56}$$

Thirdly,

$$\begin{aligned}\mu_{Z_n(t)} &= \gamma_n^{(n)} \mu_{Y_n(t)} = \gamma_n^{(n)} \kappa_n \theta_n - \kappa_n (Z_n(t) - \theta_{zn}) \\ &= \kappa_n (\bar{\theta} - Z_n(t)),\end{aligned} \tag{57}$$

where

$$\begin{aligned}\bar{\theta} &= \theta_{zn} + \gamma_n^{(n)} \theta_n = \sum_{j=1}^{n} \sum_{i=j}^{n} \gamma_i^{(j)} \frac{\kappa_i}{\kappa_j} \theta_i \\ &= \sum_{i=1}^{n} \left[\theta_i \sum_{j=1}^{i} \gamma_i^{(j)} \frac{\kappa_i}{\kappa_j} \right], \text{ (change-of-order)} \\ &= \sum_{i=1}^{n} \left[\theta_i \sum_{j=1}^{i} \left(\gamma_i^{(j)} - \gamma_i^{(j+1)} \right) \right], \text{ (use (52))} \\ &= \sum_{i=1}^{n} \left[\left(\gamma_i^{(1)} - \gamma_i^{(i+1)} \right) \theta_i \right] = \sum_{i=1}^{n} [\delta_i \theta_i].\end{aligned} \tag{58}$$

Thus, $\bar{\theta}$ is the steady state mean of $r(t)$.

We end this section by making the following two observations. First, $Z(t)$ is linked to $Y(t)$ through a sum-preserving shift and a non-singular rotation, and the transformations depend only on the free parameters of the original $AYD(N,n)$ model and known constants. In other words, the $Ar(N,n)$ model is simply a different but equivalent "basis" representation of the $AYD(N,n)$ model. To see this, we note that, since $L^{(n)}$ is generically non-singular, (17) may be rewritten as

$$Z(t) = L\hat{Y}(t), \tag{59}$$

where

$$\hat{Y}(t) = Y(t) + \vartheta, \tag{60}$$

and

$$\vartheta = \begin{pmatrix} \vartheta^{(n)} \\ 0_{(N-n)\times 1} \end{pmatrix} = L^{-1}\theta_z. \tag{61}$$

Since the first row of L is δ', and $\theta_{z1} \equiv 0$, it follows that $\delta'\vartheta = 0$.

Secondly, some of the normalizations required to identify the $AYD(N,n)$ models are built-in in the $Ar(N,n)$ representation. To see this, note that we started with n θ_i's in the conditional mean of the primary factors in the $AYD(N,n)$ representation, and ended up with only one parameter, $\bar{\theta}$, in the conditional mean of the primary factors in the $Ar(N,n)$ representation. The other $n-1$ parameters are absorbed into the definition of α_{zi}. Also note that all of the δ_i's are absorbed into other parameters in the $Ar(N,n)$ representation.

B.2 From $Ar(N,n)$ to $AYD(N,n)$

Intuitively, starting from an $Ar(N,n)$ model, a reverse transformation using L^{-1} will take us back to an $AYD(N,n)$ representation. That is, define $Y(t) = L^{-1}Z(t) - \vartheta$, where L is given by (18) and ϑ is given by (61), then, $r(t) = \sum_{i=1}^{n} \delta_i Y_i(t)$, and $Y(t)$ is governed by (49). However, both L and ϑ depend on parameters of the $AYD(N,n)$ representation. In order to achieve the reverse transformation, we need to express L in terms of the parameters of the $Ar(N,n)$ model. In addition, different choices of δ_i and θ_i, treated as normalized constants,[25] would lead to different, but equivalent $AYD(N,n)$

[25] Subject to the constraint that $\sum_{i=1}^{n} \delta_i\theta_i = \bar{\theta}$, where $\bar{\theta}$ is a primitive free parameter of the $Ar(N,n)$ model.

models (this is related to the discussion at the end of the last section). Thus, we fix $\delta_i = 1$, for $1 \leq i \leq n$, $\theta_i = 0$, for $1 \leq i < n$, and let $\theta_n = \bar{\theta}$. Then (52) and (53) become

$$\gamma_i^{(1)} = 1, \qquad \gamma_i^{(j)} = \gamma_i^{(j-1)}(1 - \frac{\kappa_i}{\kappa_{j-1}}),\ 2 \leq j \leq n \tag{62}$$

$$\theta_{z1} = 0, \qquad \theta_{zj} = \sum_{l=1}^{j-1} \gamma_n^{(l)} \frac{\kappa_n}{\kappa_l} \bar{\theta},\ 2 \leq j \leq n. \tag{63}$$

Now L and ϑ are defined entirely in terms of the primitive free parameters κ_i and $\bar{\theta}$ of the $Ar(N,n)$ model and known constants. The resulting $AYD(N,n)$ model is given by $r(t) = \sum_{i=1}^{n} Y_i(t)$,

$$dY(t) = \begin{pmatrix} \kappa_1 \left(0 - Y_1(t)\right) \\ \kappa_2 \left(0 - Y_2(t)\right) \\ \vdots \\ \kappa_{n-1} \left(0 - Y_{n-1}(t)\right) \\ \kappa_n \left(\bar{\theta} - Y_n(t)\right) \\ \kappa_{n+1} \left(\theta_{n+1} - Y_{n+1}(t)\right) \\ \vdots \\ \kappa_N \left(\theta_N - Y_N(t)\right) \end{pmatrix} + \Sigma S(t) dW(t), \tag{64}$$

with Σ, α_i, and β_i properly defined in terms of Σ_z, α_{zi}, β_{zi}, κ_i's and $\bar{\theta}$. The reverse rotation of (59) is achieved by using (43) (setting $X = L^{-1}$. The reverse shift of (60) is achieved by transforming θ and α in the following manner:

$$\theta \to \hat{\theta} = \theta - \vartheta, \tag{65}$$

and

$$\alpha_i \to \hat{\alpha}_i = \alpha_i + \beta_i' \vartheta, \quad 1 \leq i \leq N, \tag{66}$$

where $\vartheta = L^{-1}\theta_z$.

References

Ait-Sahalia, Y. 1996. Testing continuous-time models of the spot interest rate. *Review of Financial Studies* 9.

Andersen, T. G., and J. Lund, 1996. Stochastic volatility and mean drift in the short term interest rate diffusion: Sources of steepness, level and curvature in the yield curve. Working paper.

Backus, D., S. Foresi, and C. Telmer, 1996. Affine models of currency prices. Working paper, New York University.

Backus, D. K., and S. E. Zin, 1994. Reverse engineering the yield curve. NBER Working Paper #4676.

Bakshi, G. S., and Z. Chen, 1997. Asset pricing without consumption or market portfolio data. Working paper.

Balduzzi, P., S. R. Das, and S. Foresi, 1995. The central tendency: A second factor in bond yields. Working paper.

Balduzzi, P., S. R. Das, S. Foresi, and R. Sundaram. 1996. A simple approach to three factor affine term structure models. *Journal of Fixed Income* 6.

Beaglehole, D. R., and M. S. Tenney. 1991. General solutions of some interest rate-contingent claim pricing equations. *Journal of Fixed Income* .

Chen, L. 1996. *Stochastic Mean and Stochastic Volatility – A Three-Factor Model of the Term Structure of Interest Rates and Its Application to the Pricing of Interest Rate Derivatives*. Blackwell Publishers.

Chen, R., and L. Scott. 1993. Maximum likelihood estimation for a multi-factor equilibrium model of the term structure of interest rates. *Journal of Fixed Income* 3.

Cox, J. C., J. E. Ingersoll, and S. A. Ross. 1985. A theory of the term structure of interest rates. *Econometrica* 53.

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

Duffie, D., and R. Kan. 1996. A yield-factor model of interest rates. *Mathematical Finance* .

Duffie, D., and K. Singleton. 1993. Simulated moments estimation of markov models of asset prices. *Econometrica* 61.

Duffie, D., and K. Singleton, 1996. An econometric model of the term structure of interest rate swap yields. Journal of Finance, forthcoming.

Fisher, M., and C. Gilles, 1996. Estimating exponential affine models of the term structure. Working paper.

Gallant, A. R., and J. R. Long, 1997. Estimation stochastic differential equatiosn efficiently by minimum chi-square. *Biometrika*, forthcoming.

Gallant, A. R., and G. Tauchen, 1996a. Specification analysis of continuous time models in finance. *Modeling Stock Market Volatility: Bridging the Gap to Continuous Time*, Peter Rossi, ed.

Gallant, A. R., and G. Tauchen. 1996b. Which moments to match? *Econometric Theory* 12.

Hansen, L. P., and S. F. Richard. 1987. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica* 55.

K. Newey, W., and K. D. West. 1987. Hypothesis testing with efficient method of moment estimation. *International Economic Review* 28.

Litterman, R., and J. Scheinkman. 1991. Common factors affecting bond returns. *Journal of Fixed Income* 1.

Nielsen, L. T., and J. Saá-Requejo, 1993. Exchange rate and terms structure dynamics and the pricing of derivative securities. INSEAD working paper.

Pearson, N. D., and T. Sun. 1994. Exploiting the conditional density in estimating the term structure: An application to the Cox, Ingersoll, and Ross model. *Journal of Finance* XLIX.

Singleton, Kenneth J. 1994. Yield curve risk in japanese government bond markets. *Japanese Journal of Financial Economics* 1.

Table I: Estimators for SNP Parameters – s1214300

ϕ_0	Estimate	STD	t-ratio	p-value
A_2	0.14072	0.15618	0.90100	36.76%
A_3	0.04042	0.05218	0.77500	43.83%
A_4	0.09981	0.07455	1.33900	18.06%
A_5	-0.29095	0.09579	-3.03700	0.24%
A_6	0.02513	0.03418	0.73500	46.23%
A_7	-0.10982	0.04492	-2.44500	1.45%
A_8	-0.01392	0.03543	-0.39300	69.43%
A_9	0.00340	0.00979	0.34800	72.78%
A_{10}	-0.00818	0.00925	-0.88300	37.72%
A_{11}	0.02178	0.01547	1.40800	15.91%
A_{12}	0.00843	0.00590	1.42900	15.30%
A_{13}	0.01519	0.00422	3.59900	0.03%
ψ_1	-0.01931	0.00875	-2.20800	2.72%
ψ_2	-0.02615	0.01832	-1.42700	15.36%
ψ_3	-0.02708	0.02375	-1.14000	25.43%
ψ_4	0.92663	0.01877	49.36800	0.00%
ψ_5	-0.00359	0.02945	-0.12200	90.29%
ψ_6	0.01104	0.03406	0.32400	74.59%
ψ_7	0.09446	0.02626	3.59700	0.03%
ψ_8	1.00261	0.04092	24.50400	0.00%
ψ_9	-0.01008	0.04758	-0.21200	83.21%
ψ_{10}	-0.02676	0.01217	-2.19900	2.79%
ψ_{11}	-0.00685	0.01769	-0.38700	69.88%
ψ_{12}	0.99160	0.02014	49.24200	0.00%
τ_1	0.03544	0.00373	9.51000	0.00%
τ_2	0.02612	0.00290	8.99200	0.00%
τ_3	0.03368	0.00372	9.05100	0.00%
τ_4	0.06457	0.00725	8.90500	0.00%
τ_5	0.13351	0.01372	9.72800	0.00%
τ_6	0.16425	0.01737	9.45500	0.00%
τ_7	0.16619	0.05915	2.80900	0.50%
τ_{15}	0.01944	0.02152	0.90300	36.65%
τ_{24}	-0.01995	0.03042	-0.65600	51.18%
τ_{25}	0.12629	0.05575	2.26500	2.35%
τ_{33}	0.02938	0.02403	1.22300	22.13%
τ_{42}	0.03229	0.03335	0.96800	33.30%

Table II: EMM Estimators

Model	*Chen*		$AYD(3,2)$		$AYD(3,3)$	
ψ_0	Estimate	STD	Estimate	STD	Estimate	STD
δ_3	-	-	-	-	112.253	49.162
κ	2.596	0.322	4.570	1.309	2.820	0.296
ν	0.124	0.031	0.179	0.046	0.148	0.042
μ	1.511	0.454	0.694	0.178	1.045	0.001
θ	0.034	0.005	0.026	0.007	0.014	0.009
$\bar{v}$	$2.456\,e^{-4}$	$0.364\,e^{-4}$	$4.378\,e^{-4}$	$1.895\,e^{-4}$	$2.298\,e^{-4}$	$0.126\,e^{-4}$
$\sigma_{r\theta}$	-	-	-0.481	0.303	-0.139	0.090
σ_{rv}	-	-	-79.934	35.085	-200.541	47.680
$\hat{a}$	-	-	0.00005	0.00003	0.0001	0.00003
ζ	0.056	0.002	0.056	0.003	0.053	0.012
η	0.022	0.003	0.015	0.002	0.012	0.000
λ_1	-114.734	28.532	-345.707	141.219	-251.080	175.057
λ_2	-38.788	6.829	-53.162	6.103	-44.856	8.720
λ_3	-1950.479	448.798	1693.901	2812.175	1705.858	4214.659
χ^2	129.027 (26)	0.000%	30.224 (23)	14.308%	26.455 (22)	23.266%

Columns 2, 4, and 6 of the last row indicate χ^2 test statistics and (in parentheses) the associated degrees of freedom for the respective models. Columns 3, 5, and 7 are the corresponding p-values. Relations between the parameters presented in the table and the primitive parameters of the $AYD(N,n)$ model are: $\sigma_{r\theta} = \frac{\kappa}{\kappa-\nu}(\sigma_{12}+1)$, $\sigma_{rv} = \sigma_{13}$, $\hat{a} = (\frac{\kappa-\nu}{\kappa})^2(\alpha_2 - \frac{\nu}{\kappa-\nu}\bar{\theta}b_{22})$, $\zeta^2 = \frac{\kappa-\nu}{\kappa}b_{22}$, and $\eta^2 = b_{33}$.

Table III: SNP Scores

Model	*Chen*		$AYD(3,2)$		$AYD(3,3)$	
ϕ_0	Score	t-ratio	Score	t-ratio	Score	t-ratio
A_2	0.524	0.272	-0.096	-0.051	0.043	0.023
A_3	6.132	2.687*	4.396	2.033*	3.809	1.767
A_4	2.071	0.997	1.621	0.906	-0.523	-0.321
A_5	3.772	1.874	0.352	0.197	1.121	0.660
A_6	7.893	3.770*	4.735	2.753*	4.365	2.631*
A_7	7.367	3.371*	-0.008	-0.004	2.088	1.093
A_8	-1.114	-0.227	-1.957	-0.410	-0.973	-0.198
A_9	21.163	2.338*	13.648	1.575	10.464	1.216
A_{10}	13.073	1.507	5.540	0.734	-2.791	-0.389
A_{11}	7.208	0.775	0.464	0.053	3.167	0.363
A_{12}	28.332	1.963	16.241	1.245	17.861	1.400
A_{13}	41.161	1.954	9.423	0.497	22.206	1.135
ψ_1	-13.408	-0.674	1.164	0.066	-14.755	-0.939
ψ_2	45.011	1.726	33.874	1.396	39.618	1.618
ψ_3	-31.728	-1.578	-29.715	-1.545	-27.521	-1.434
ψ_4	29.209	1.840	21.513	1.202	22.766	1.357
ψ_5	-67.102	−3.246*	-43.163	−2.184*	-39.472	−2.105*
ψ_6	35.431	2.055*	24.567	1.540	17.755	1.102
ψ_7	33.837	2.105*	13.985	0.788	21.728	1.252
ψ_8	-65.471	−2.944*	-39.752	-1.862	-42.950	−2.053*
ψ_9	30.380	1.660	22.348	1.282	17.543	1.013
ψ_{10}	15.895	0.889	2.060	0.112	17.974	0.966
ψ_{11}	-49.125	−2.054*	-29.329	-1.252	-40.383	-1.748
ψ_{12}	18.126	0.937	13.276	0.696	10.984	0.585
τ_1	45.026	3.105*	-9.378	-0.724	1.557	0.108
τ_2	-217.587	−8.600*	-13.143	-1.230	-11.956	-0.960
τ_3	226.365	7.513*	39.029	1.972	30.767	1.541
τ_4	-57.585	−3.841*	-7.450	-0.722	-3.757	-0.356
τ_5	47.061	2.595*	8.234	0.571	7.071	0.502
τ_6	8.111	0.650	6.647	0.603	8.773	0.863
τ_7	3.055	2.392*	1.319	1.016	1.925	1.414
τ_{15}	11.681	3.736*	2.447	0.880	2.152	0.769
τ_{24}	-2.011	-1.032	-1.632	-0.876	-1.293	-0.689
τ_{25}	2.274	2.084*	0.667	0.589	1.316	1.085
τ_{33}	12.217	4.796*	2.911	1.364	2.534	1.175
τ_{42}	-0.855	-0.356	-0.411	-0.175	-0.109	-0.046

Numbers indicated by * have absolute values larger than 2.

Figure 1: First State - $AYD(3,3) - -s1214300$

(All time series are standardized.)

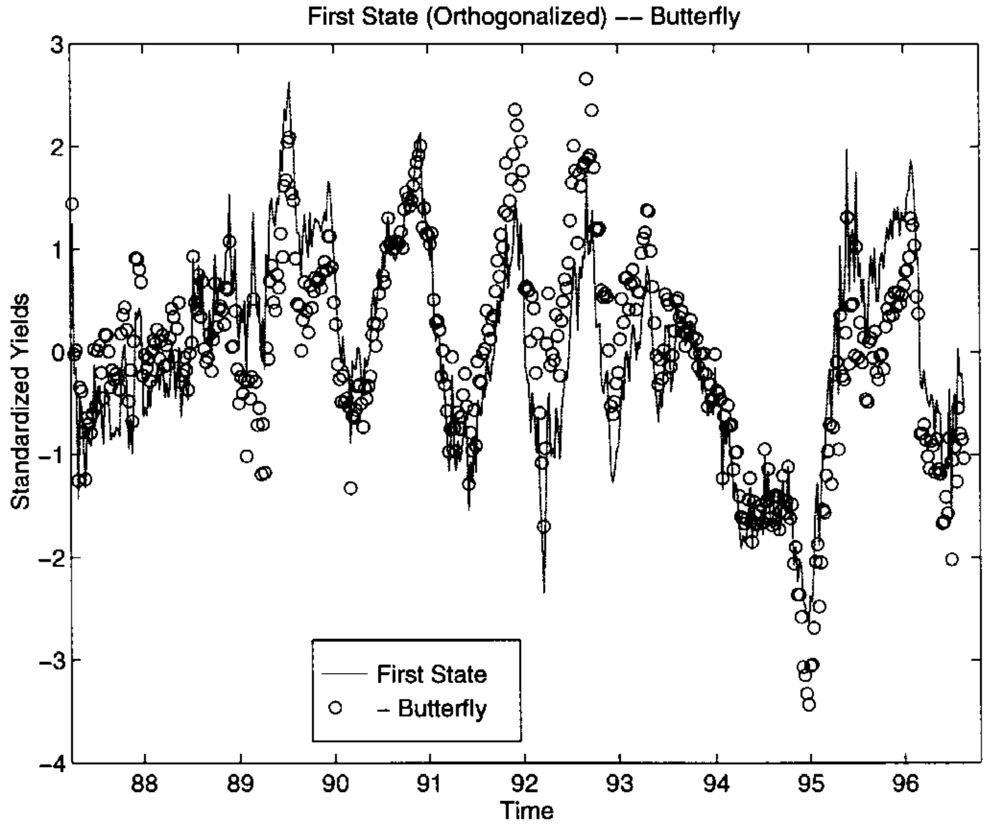

Figure 2: Second State – $AYD(3,3) - -s1214300$

(All time series are standardized.)

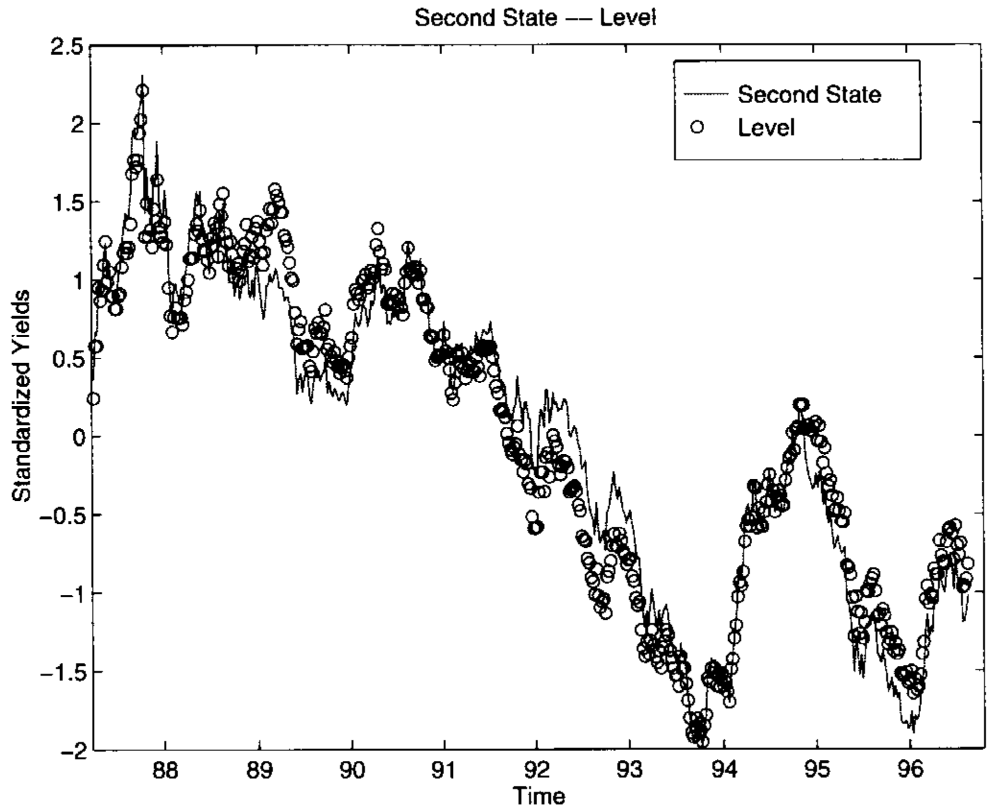

Figure 3: Third State - $AYD(3,3)$ - $-s1214300$
(All time series are standardized.)

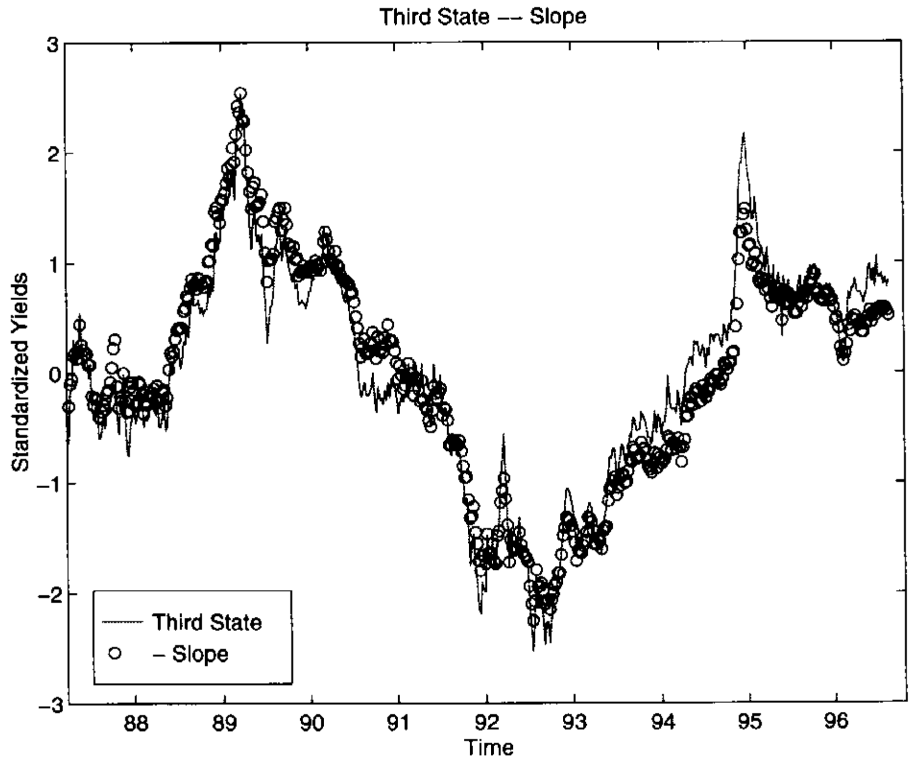

Figure 4: Simulated Mean Swap Curves