Discrete-time Dynamic Term Structure Models with Generalized Market Prices of Risk

Qiang Dai, Anh Le, and Kenneth J. Singleton [1]

This draft: August 21, 2005

[1]Dai is with University of North Carolina at Chapel Hill, dq@unc.edu. Le is a graduate student at the Stern School of Business. Singleton is with the Graduate School of Business, Stanford University, Stanford, CA 94305 and NBER, ken@future.stanford.edu. We are grateful for financial support from the Gifford Fong Associates Fund, at the Graduate School of Business, Stanford University.

1 Introduction

In this paper, we formulate and empirically investigate a rich class of discrete-time, nonlinear dynamic term structure models (*DTSM*s). Key to our analysis is the insight that, by shifting to discrete time, we can both substantially extend the family of affine diffusion models that have heretofore been studied empirically *and* obtain closed-form solutions for the joint likelihoods of the bond yield data. Under the risk-neutral measure, the distribution of the state vector resides within a family of discrete-time Markov processes that nests the *exact* discrete-time counterparts of the entire class of continuous-time models in Duffie and Kan (1996) and Dai and Singleton (2000). Consequently, as in these studies, we obtain closed-form, exponential-affine expressions for zero-coupon bond prices.[1]

Where we gain considerable generality over the extant literature on affine *DTSM*s is through our parameterizations of the state-price density and market price of risk Λ_t. We specify the state-price density– essentially the Radon-Nykodym derivative linking the risk-neutral ($\mathbb{Q}$) and historical ($\mathbb{P}$) measures– as a known function of the Λ_t. Then we allow for very general dependence of Λ_t on the state X_t, requiring only that this dependence rules out arbitrage opportunities and that the $\mathbb{P}$ distribution of X satisfy certain stationarity/ergodicity conditions needed for econometric analysis. This leads to a quite flexible family of *nonlinear DTSM*s for which the $\mathbb{P}$ distribution of zero-coupon bond yields is *not* affine, but for which we have exact analytic expressions for the conditional densities of bond yields.[2] Moreover, our state-price density is chosen so that, through restricted choices of Λ_t, we nest the (discrete-time) counterparts to the extant *linear* models in which the state follows an affine process under both $\mathbb{P}$ and $\mathbb{Q}$.[3]

Both economic and econometric considerations motivate this analysis. Recent empirical studies find that the goodness-of-fits of *DTSM*s depend critically on the specification of market price of risk (e.g., Duffee (2002), Dai and Singleton (2002), Duarte (2004), and Ahn, Dittmar, and Gallant (2002)). However, the functional forms of Λ_t in these studies are still quite restrictive – reflecting the usual trade-off between generality and the tractability of estimation. By allowing the researcher almost complete freedom in specifying the dependence

[1] Our analysis extends immediately to the case of quadratic-Gaussian models of the type discussed in Beaglehole and Tenney (1991), Ahn, Dittmar, and Gallant (2002) and Leippold and Wu (2002). However, we focus on the affine case.

[2] In particular, our framework allows for nonlinearity in the conditional means of bond yields of the type examined by Ait-Sahalia (1996), Stanton (1997), Chan, Karolyi, Longstaff, and Sanders (1992), and Duarte (2004), all in a multi-factor setting. There is also evidence supporting nonlinear behavior in the interest rates from descriptive regime switching models of short-term interest rates (see, e.g., Gray (1996) and Ang and Bekaert (2002)). Recently, Naik and Lee (1997), Evans (2000), Boudoukh, Richardson, Smith, and Whitelaw (1999), Bansal and Zhou (2002), Ang and Bekaert (2003), and Dai, Singleton, and Yang (2005), among others, introduce regime switching on top of standard affine term structure models and examine its implications for the pricing of long-term bonds. Here we work exclusively within a single-regime framework, though we comment briefly on how our framework can be extended to accommodate multiple regimes in the presence of stochastic volatility in Section 6.

[3] In the continuous-time literature, this is a feature of the models examined in Dai and Singleton (2000), Duffee (2002), and Cheridito, Filipovic, and Kimmel (2003). It is also true of the discrete-time affine term structure models discussed in Ang and Piazzesi (2003) and Gourieroux, Monfort, and Polimenis (2002).

of Λ_t on the state vector, we facilitate empirical investigation of much richer specifications of risk premiums. Furthermore, the development of the exact discrete-time counterparts to the entire family of affine models examined by Dai and Singleton (2000) (hereafter DS) substantially expands the family of models within which the macroeconomic underpinnings of the latent risk factors in $DTSM$s can be tractably studied empirically. To date, the literature on integrating $DTSM$s with dynamic macroeconomic models (e.g., Rudebusch and Wu (2003), Hordahl, Tristani, and Vestin (2003), Dai and Phillipon (2004), and Ang, Dong, and Piazzesi (2005)) has focused exclusively on discrete-time Gaussian $DTSM$s thereby ruling out a role for either nonlinearity or time-varying second moments in modelling macroeconomic risks.

With regard to estimation of $DTSM$s, even when the state vector follows an affine diffusion under the physical measure, the one-step ahead conditional density of the state vector is not known in closed form, except for the special cases of Gaussian (Vasicek (1977)) and independent square-root diffusions (Cox, Ingersoll, and Ross (1985)). Accordingly, in estimation, the literature has relied on approximations, with varying degrees of complexity, to the relevant conditional $\mathbb{P}$-densities.[4] By shifting to discrete time, we are able to nest the (discrete-time counterparts to the) entire class of affine $DTSM$s classified by DS within a much larger class of nonlinear $DTSM$s, and at the same time obtain exact representations of the likelihood functions of bond yields. Therefore, no approximations are necessary in estimation.

Our development of our family of nonlinear $DTSM$s proceeds in three steps. First, we develop $N+1$ families of discrete-time affine processes $DA_M^{\mathbb{Q}}(N)$, in which m of the N risk factors drive stochastic volatility ($M = 0, \ldots, N$). Each member of $DA_M^{\mathbb{Q}}(N)$ will serve as an admissible $\mathbb{Q}$ representation of the risk factors, exactly analogously to the family $A_M^{\mathbb{Q}}(N)$ of $\mathbb{Q}$-affine models examined in *DS*. For the m volatility factors, we build upon the analysis of scalar "autoregressive gamma" processes in Gourieroux and Jasiak (2001) and Darolles, Gourieroux, and Jasiak (2001) to develop the discrete-time counterpart to the multivariate, correlated *CIR* process, Z_t in the family $DA_M^{\mathbb{Q}}(M)$. This construction is then extended to the family $DA_M^{\mathbb{Q}}(N)$ by introducing an $N-M$ dimensional state process Y_{t+1} with the property that, conditional on $X_t = (Z_t', Y_t')'$, it is normally distributed with a conditional variance that is an affine function of Z_t.

Given a $\mathbb{Q}$-affine representation of the risk factors X residing in $DA_M^{\mathbb{Q}}(N)$, for some M, the pricing of zero-coupon bonds is straightforward under the additional assumption that the one-period short-term rate is an affine function of X. Zero-coupon bond prices are exact, exponential-affine functions of X, just as in the continuous-time counterparts $A_M^{\mathbb{Q}}(N)$ examined in Duffie and Kan (1996) and DS.[5]

[4] These include the direct approximations to the conditional densities explored in Duan and Simonato (1999), Ait-Sahalia (1999, 2002), and Duffie, Pedersen, and Singleton (2003); the Monte Carlo based approximations of Pedersen (1995) and Brandt and Santa-Clara (2001)); and the simulation-based method-of-moments estimators proposed by Duffie and Singleton (1993) and Gallant and Tauchen (1996).

[5] As with DS's construction of a canonical model for the family $A_M^{\mathbb{Q}}(N)$, our canonical model for $DA_M^{\mathbb{Q}}(N)$ is the maximally flexible $\mathbb{Q}$-representation of the first-order Markov process X_t. Fixing the state space to be $\mathbb{R}^M \times \mathbb{R}^{N-M}$, Duffie, Filipovic, and Schachermayer (2003) show that DS's normalizations and constraints are necessary and sufficient to derive the maximally $\mathbb{Q}$-admissible (i.e., canonical) continuous-time affine

Second, for each family $DA_M^{\mathbb{Q}}(N)$, we specify an associated family of state-price densities, each member of which is a known function of $(X_{t+1}, X_t, \Lambda_t)$. Our particular parametrization of the state-price price density is chosen to be a natural discrete-time counterpart to the state-price density associated with affine diffusion-based, continuous-time *DTSM*s. When combined with a known $\mathbb{Q}$-affine distribution of the state X, each member of this family of state-price densities leads to a known parametric representation of the $\mathbb{P}$-distribution of X. Moreover, since bond prices are a known function of X, it follows immediately that the likelihood functions of data on zero-coupon or coupon bond prices are known exactly in closed form.

Finally, this construction of the likelihood function allows the modeler substantial flexibility in specifying the dependence of Λ_t on X_t, requiring only that the model not admit arbitrage opportunities and be econometrically identified, and that the $\mathbb{P}$ distribution of X be sufficiently regular for the maximum likelihood estimators to have well-behaved large-sample distributions. By roaming over admissible choices of Λ_t, we are effectively ranging across the entire family of admissible arbitrage-free *DTSM*s constructed under the assumption that, under $\mathbb{Q}$, X follows an affine process. While, in principle, similar flexibility arises in $A_M^{\mathbb{Q}}(N)$ models, researchers have rarely exploited this flexibility in practice because of the computational complexity arising from a non-affine $\mathbb{P}$ distribution of X. Our discrete-time formulation circumvents these computational considerations by delivering an exact likelihood function under general state-dependence of Λ_t.

To illustrate our approach, we report estimates of two different nonlinear $(DA_1^{\mathbb{Q}}(3), \Lambda)$ models. The first is Duarte (2004)'s $SASR_1(3)$ model in which the square-root of the volatility factor appears in its own drift under $\mathbb{P}$. This model is of interest in part, because the likelihood function of its continuous-time counterpart (studied by Duarte) is not known in closed form. The second is an alternative formulation of the volatility process under which its conditional $\mathbb{P}$-mean depends on its squared and cubed values. These models are compared along various dimensions, including their within and out of sample forecasting performances for bond yields.

In what is perhaps the closest precursor to our analysis, Gourieroux, Monfort, and Polimenis (2002) developed $DTSM$s based on the single-factor autoregressive gamma model (the discrete-time counterpart to a one-factor CIR model), and multi-factor Gaussian models (the counterparts of $A_0^{\mathbb{Q}}(N)$ models). In terms of coverage of models, our framework extends their analysis to all of the families of multi-factor models $DA_M^{\mathbb{Q}}(N)$, $M = 0, 1, \ldots, N$. Furthermore, Gourieroux, et. al. assumed that the market price of risk Λ is constant and, as such, they focused on the "completely" affine versions of the $DA_1^{\mathbb{Q}}(1)$ and $DA_0^{\mathbb{Q}}(N)$ models. A major focus of our analysis is the specification and estimation of discrete-time affine *DTSM*s that allow general dependence of Λ_t on X_t.

The families of models $DA_M^{\mathbb{Q}}(N)$, $M = 0, \ldots, N$, are not the only well-defined discrete-time affine *DTSM*s. Gourieroux, Monfort, and Polimenis (2002) discuss a variety of other examples that are outside the purview of our analysis (because their continuous-time coun-

model. Collin-Dufresne, Goldstein, and Jones (2004) discuss an equivalent canonical continuous-time model based on an invariant transformation of *DS*'s canonical model.

terparts do not reside in one of the families $A_M^{\mathbb{Q}}(N)$). Moreover, Ang and Piazzesi (2003) and Gourieroux, Monfort, and Polimenis (2002) illustrate (in the context of $DA_0^{\mathbb{Q}}(N)$ models) the fact that discrete-time affine *DTSM*s can be extended to include lagged values of the state. All of our representations of the $\mathbb{Q}$ distributions of X can similarly be extended to higher-order Markov processes, though we choose to focus on the case of first-order Markov processes for ease of exposition.

2 Canonical Discrete-Time Affine Processes

Following Duffie, Filipovic, and Schachermayer (2003), we will refer to a Markov process X as *affine* if the conditional Laplace transforms of X_{t+1} given X_t is an exponential-affine function of X_t:[6] under a probability measure $\mathbb{Q}$, for an $N \times 1$ state vector X,

$$\phi^{\mathbb{Q}}(u; X_t) = E^{\mathbb{Q}}\left[e^{u' X_{t+1}} \middle| X_t \right] = e^{a(u)+b(u)X_t}. \tag{1}$$

Paralleling *DS*, we focus (by choice of the $N \times 1$ vector $a(u)$ and $N \times N$ matrix $b(u)$) on the particular sub-families of discrete-time affine models $DA_M^{\mathbb{Q}}(N)$ that are formally the exact discrete-time counterparts to their families $A_M^{\mathbb{Q}}(N)$. The members of $DA_M^{\mathbb{Q}}(N)$ are well-defined affine models in their own right, and also have (by construction) the property that, as the sampling interval of the data shrinks to zero, they converge to members of the continuous-time family $A_M^{\mathbb{Q}}(N)$.

Throughout this paper, we assume that the state vector X_t is affine under the risk-neutral measure $\mathbb{Q}$, in the sense just described. Hence equation (1) constitutes a basic distributional assumption of our model. In the rest of this section, we make explicit the functional forms of $a(\cdot)$ and $b(\cdot)$ that define the $\mathbb{Q}$-affine families $DA_M^{\mathbb{Q}}(N)$, $M = 0, \ldots, N$.

2.1 $DA_0^{\mathbb{Q}}(N)$

The $DA_0^{\mathbb{Q}}(N)$ process is an $N \times 1$ vector Y that follows a Gaussian vector autoregression: conditional on Y_t, Y_{t+1} is normally distributed with conditional mean $\mu_0 + \mu_Y Y_t$, and conditional covariance matrix V. The conditional Laplace transform of Y is given by (1) with

$$a(u) = \mu_0' u + \frac{1}{2} u' V u, \quad b(u) = u' \mu_Y. \tag{2}$$

To derive the continuous-time counterpart of this family, let Δt be the length of the observation interval, and let $\mu_0 = \kappa^{\mathbb{Q}} \theta^{\mathbb{Q}} \Delta t$, $\mu_Y = I_{N \times N} - \kappa^{\mathbb{Q}} \Delta t$, and $V = \sigma \sigma' \Delta t$, where $\kappa^{\mathbb{Q}}$ and σ are $N \times N$ matrices and $\theta^{\mathbb{Q}}$ is a $N \times 1$ vector. Then in the limit $\Delta t \to 0$,

[6] See Duffie, Pan, and Singleton (2000) for a proof that continuous-time affine processes typically examined have conditional characteristic functions that are exponential-affine functions, and Gourieroux and Jasiak (2001) and Darolles, Gourieroux, and Jasiak (2001) for discussions of discrete-time affine processes related to those examined in this paper.

the process $DA_0^{\mathbb{Q}}(N)$ converges to the continuous-time $A_0(N)$ process, the N-dimensional Gaussian process:

$$dY_t = \kappa^{\mathbb{Q}}(\theta^{\mathbb{Q}} - Y_t)dt + \sigma dB_t^{\mathbb{Q}},$$

where $B_t^{\mathbb{Q}}$ is a $N \times 1$ vector of standard Brownian motions under the measure $\mathbb{Q}$.

Virtually all of the empirical work to date on multi-factor (exact) discrete-time affine models has focused on the family $DA_0^{\mathbb{Q}}(N)$. See, for example, Ang and Piazzesi (2003), Dai, Singleton, and Yang (2005), Rudebusch and Wu (2003), Hordahl, Tristani, and Vestin (2003), and Dai and Phillipon (2004).

2.2 $DA_N^{\mathbb{Q}}(N)$

Perhaps the most widely studied family of continuous-time affine *DTSM*s is the family $A_N^{\mathbb{Q}}(N)$, the multi-factor *CIR*-style models (see Dai and Singleton (2003) for a survey). Numerous authors, including Sun (1992), Gray (1996), and Bekaert, Engstrom, and Grenadier (2004), have examined discrete-time "CIR models" in which the shock to a state variable Z_t takes the form $\sigma_Z\sqrt{Z_{t-1}}\epsilon_t$, $\epsilon_t \sim N(0,1)$. The resulting term structure models are not exact, either in the pricing of bonds or in the representations of the likelihood functions, because these models are not well defined if ϵ_{t+1} is literally a normal random variable.

The $DA_N^{\mathbb{Q}}(N)$ process is the exact discrete-time equivalent of the multi-variate *correlated* square-root or *CIR* process; Z is non-negative with probability one, no approximations are required in the pricing bonds, and the associated likelihood functions are known exactly in closed-form. The scalar case $N = 1$ was explored in depth in Gourieroux and Jasiak (2001) and Darolles, Gourieroux, and Jasiak (2001). We extend their analysis to the multi-variate case of a $DA_N^{\mathbb{Q}}(N)$ process Z_t as follows.

As in the canonical $A_N^{\mathbb{Q}}(N)$ model of *DS* we assume that, conditional on Z_t, the components of Z_{t+1} are independent. To specify the conditional distribution of Z_{t+1}, we let ϱ be an $N \times N$ matrix with elements satisfying

$$0 < \varrho_{ii} < 1, \ \ \varrho_{ij} \leq 0, \ \ 1 \leq i, j \leq N.$$

Furthermore, for each $1 \leq i \leq N$, we let ρ_i be the i^{th} row of the $N \times N$ non-singular matrix $\rho = (I_{N\times N} - \varrho)$. Then, for constants $c_i > 0$, $\nu_i > 0$, $i = 1, \ldots, N$, we define the conditional density of Z_{t+1}^i given Z_t as the Poisson mixture of standard gamma distributions:

$$Z_{t+1}^i|(\mathcal{P}, Z_t) \sim gamma(\nu_i + \mathcal{P}), \quad \text{where} \ \ \mathcal{P}|Z_t \sim Poisson(\rho_i Z_t/c_i). \tag{3}$$

Here, the random variable $\mathcal{P} \in (0, 1, 2, \ldots)$ is drawn from a Poisson distribution with intensity modulated by the current realization of the state vector Z_t, and it in turn determines the coefficient of the standard gamma distribution (with scale parameter equal to 1) from which Z_{t+1}^i is drawn.

The conditional density function of Z_{t+1}^i takes the form:

$$f^{\mathbb{Q}}(Z_{t+1}^i|Z_t) = \frac{1}{c_i}\sum_{k=0}^{\infty}\left[\frac{\left(\frac{\rho_i Z_t}{c_i}\right)^k}{k!} e^{-\frac{\rho_i Z_t}{c_i}} \times \frac{\left(\frac{Z_{t+1}^i}{c_i}\right)^{\nu_i+k-1} e^{-\frac{Z_{t+1}^i}{c_i}}}{\Gamma(\nu_i + k)}\right]. \tag{4}$$

Using conditional independence, the distribution of a $DA_N^{\mathbb{Q}}(N)$ process Z_{t+1}, conditional on Z_t, is given by $f^{\mathbb{Q}}(Z_{t+1}|Z_t) = \prod_{i=1}^N f^{\mathbb{Q}}(Z_{t+1}^i|Z_t)$. Finally, it is straight-forward to show that for any u, such that $u_i < \frac{1}{c_i}$, the conditional Laplace transform of Z_{t+1} is given by (1) with

$$a(u) = -\sum_{i=1}^{N} \nu_i \log(1 - u_i c_i), \quad b(u) = \sum_{i=1}^{N} \frac{u_i}{1 - u_i c_i} \rho_i.$$

When the off-diagonal elements of the $N \times N$ matrix ϱ are non-zero, the autoregressive gamma processes $\{Z^i\}$ are (unconditionally) correlated. Thus, even in the case of correlated Z_t^i, the conditional density of Z_{t+1} is known in closed form. This is not the case for correlated Z in the continuous-time family $A_N^{\mathbb{Q}}(N)$. The nature of the correlation between Z^i and Z^j $(i \neq j)$ is constrained by our requirement that $\varrho_{ij} \leq 0$. Analogous to the constraint imposed by *DS* on the off-diagonal elements of the feedback matrix $\kappa^{\mathbb{Q}}$ in their continuous-time models, this constraint serves to assure that feedback among the Z's through their conditional means does not compromise the requirement that the intensity of the Poisson process be positive. Equivalently, it assures that we have a well-defined multivariate discrete-time process taking on strictly positive values.

The conditional mean $E_t^{\mathbb{Q}}[Z_{t+1}]$ and conditional covariance matrix $V_t^{\mathbb{Q}}[Z_{t+1}]$ implied by this conditional moment-generating function are

$$E_t^{\mathbb{Q}}[Z_{t+1}] = a^{(1)}(0) + \sum_{i=1}^{n} b_i^{(1)}(0) Z_t^i, \;\; V_t^{\mathbb{Q}}[Z_{t+1}] = a^{(2)}(0) + diag\left[\partial^2 b/\partial u_i^2(0) Z_t\right], \qquad (5)$$

where $a^{(k)}(u)$ denotes the k^{th} derivative of $a(u)$ with respect to u and $diag[\cdot]$ denotes the diagonal matrix generated by the elements in brackets. Specifically,

$$E_t^{\mathbb{Q}}[Z_{t+1}](i) = \nu_i c_i + \rho_i Z_t, \;\; V_t^{\mathbb{Q}}[Z_{t+1}](i,i) = \nu_i c_i^2 + 2 c_i \rho_i Z_t, \qquad (6)$$

and the off-diagonal elements of $V_t^{\mathbb{Q}}[Z_{t+1}]$ are all zero (correlation occurs only through the feedback matrix). Note the similarity between the affine form of these moments and those of the exact discrete-time process implied by a univariate square-root diffusion.

That this process converges to the multi-factor correlated $A_N^{\mathbb{Q}}(N)$ process[7] can be seen by letting $\rho = I_{N\times N} - \kappa^{\mathbb{Q}} \Delta t$, $c_i = \frac{\sigma_i^2}{2}\Delta t$, and $\nu_i = \frac{2(\kappa^{\mathbb{Q}}\theta^{\mathbb{Q}})_i}{\sigma_i^2}$, where $\kappa^{\mathbb{Q}}$ is a $N \times N$ matrix and $\theta^{\mathbb{Q}}$ is a $N \times 1$ vector. In the limit as $\Delta t \to 0$, the $DA_N^{\mathbb{Q}}(N)$ process converges to:

$$dZ_t = \kappa^{\mathbb{Q}}(\theta^{\mathbb{Q}} - Z_t)dt + \sigma\sqrt{diag(Z_t)}dB_t^{\mathbb{Q}},$$

where σ is a $N \times N$ diagonal matrix with i^{th} diagonal element given by σ_i.

[7] Gourieroux and Jasiak (2001) attribute the insight that the $DA_1^{\mathbb{Q}}(1)$ process is a discrete-time counterpart to the square-root diffusion to Lamberton and Lapeyre (1992).

2.2.1 $DA_M^{\mathbb{Q}}(N)$ Processes, For $0 < M < N$

We refer to an $N \times 1$ vector of stochastic processes $X_t = (Z_t', Y_t')'$ as a $DA_M^{\mathbb{Q}}(N)$ process if (i) Z_t is an autonomous $DA_M^{\mathbb{Q}}(M)$ process; and (ii) conditional on Y_t and Z_t, Y_{t+1} is independent of Z_{t+1}[8] and normally distributed with conditional mean and variance

$$\omega_{Yt}^{\mathbb{Q}} \equiv \mu_0 + \mu_Y X_t \text{ and } \Omega_{Yt} \equiv \Sigma_Y S_{Yt} \Sigma_Y', \tag{7}$$

where μ_0 is a $(N-M)\times 1$ vector, $\mu_Y \equiv \left(\mu_Y^Z \quad \mu_Y^Y \right)$ is a $(N-M)\times N$ matrix, μ_Y^Z is a $(N-M)\times m$ matrix, μ_Y^Y is a $(N-M)\times(N-M)$ matrix, Σ_Y is an $(N-M)\times(N-M)$ matrix, and S_{Yt} is a $(N-M)\times(N-M)$ diagonal matrix with i^{th} diagonal given by $\alpha_i + \beta_i' Z_t$, $1 \leq i \leq N-M$. By construction, then, the conditional density of X is given by

$$f^{\mathbb{Q}}(X_{t+1}|X_t) = f^{\mathbb{Q}}(Y_{t+1}|Y_t, Z_t) \times f^{\mathbb{Q}}(Z_{t+1}|Z_t), \tag{8}$$

with the first term being a multi-variate Gaussian density and the second term being a multi-variate autoregressive gamma density.

Let u_Z and u_Y be $M \times 1$ and $(N-M)\times 1$ vectors such that $u = (u_Z', u_Y')'$, and let h_0 and h_i, $i = 1, 2, \ldots, M$ be $(N-M)\times(N-M)$ matrices defined as the coefficients in the expansion of $\Omega_{Yt} = h_0 + \sum_{i=1}^M h_i Z_t^i$, then the conditional Laplace transform of X_{t+1} given X_t is again given by (1), with

$$a(u) = -\sum_{i=1}^{M} \nu_i \log\left(1 - u_{Z,i} c_i\right) + \mu_0' u_Y + \frac{1}{2} u_Y' h_0 u_Y, \tag{9}$$

$$b(u) = \left[\begin{array}{cc} \sum_{i=1}^M \frac{u_{Z,i}}{1-u_{Z,i}c_i}\rho_i + \left(\frac{1}{2}u_Y' h_i u_Y\right)_{i=1,2,\ldots,M} + u_Y' \mu_Y^Z & u_Y' \mu_Y^Y \end{array} \right], \tag{10}$$

provided that $u_{Z,i} < \frac{1}{c_i}$ for all $1 \leq i \leq M$.

Based on the above constructions, our first maintained assumption can be summarized as follows:

Assumption 1 (N($\mathbb{Q}$)) : *Under $\mathbb{Q}$, the state vector X_t follows a $DA_M^{\mathbb{Q}}(N)$ process, with its conditional Laplace transform given by (1), (9), and (10).*

If $M = 0$, we write $X_t = Y_t$, where Y_t is a $DA_0^{\mathbb{Q}}(N)$ process. If $M > 0$, we write $X_t = (Z_t', Y_t')'$, where Z_t is a $DA_M^{\mathbb{Q}}(M)$ process.

2.3 Bond Pricing

As in the extant literature on affine term structure models, we assume that the interest rate on one-period zero-coupon bonds is related to the state vector according to:

[8] Within a general $A_M^{\mathbb{Q}}(N)$ model the M factors driving stochastic volatility and the remaining $(N-M)$ factors may be (instantaneously) correlated. However, as discussed in Dai and Singleton (2000), within a term structure context one is free to normalize these (instantaneous) correlations to zero. Our conditional independence assumption is the discrete-time counterpart to this normalization.

Assumption 2 (N(r)) *: r_t is affine in X_t, i.e., $r_t = \delta_0 + \delta_X X_t$, where $\delta_X > 0$ is a $1 \times N$ vector.*[9]

Assumptions N($\mathbb{Q}$) and N(r) imply that zero-coupon bond yields are linear in the state vector X_t. Specifically, the time-t zero-coupon bond price with maturity of n periods is given by

$$D_t^n = E_t^{\mathbb{Q}}\left[e^{-\sum_{i=0}^{n-1} r_{t+i}}\right] = e^{-r_t} E_t^{\mathbb{Q}}\left[D_{t+1}^{n-1}\right] = e^{-A_n - B_n X_t},. \tag{11}$$

where the loadings A_n and B_n are determined by the following recursion:

$$A_n - A_{n-1} = \delta_0 + A_{n-1} - a(-B_{n-1}),, \tag{12}$$

$$B_n = \delta_X - b(-B_{n-1}), \tag{13}$$

with the initial condition $A_0 = B_0 = 0$.[10]

The linear structure to the cross-section of bond yields implied by affine *DTSM*s, including the discrete-time models examined here, is potentially restrictive. Indeed, Boudoukh, Richardson, Stanton, and Whitelaw (1998) present evidence of departures from this linear structure within a two-factor setting. Yet Litterman and Scheinkman (1991), and many subsequent papers, have shown that assuming that bond yields are linear functions of a small number of factors (e.g., principal components of yields) provides an effective means of hedging bond portfolios. Accordingly we maintain the linear yield structure implied by (11) and, thereby, preserve tractability of bond pricing.

3 Physical Distribution of Bond Yields

A standard means of constructing an affine *DTSM* in continuous time is to start with a $\mathbb{Q}$ representation of X in one of the families $A_M^{\mathbb{Q}}(N)$, introduce a market price of risk η_t for the state X, and then derive the implied $\mathbb{P}$ distributions of X and bond yields. Equivalently, in a diffusion setting, one posits a pricing kernel or Radon-Nykodym derivative

$$(d\mathbb{Q}/d\mathbb{P})_{t,t+1}^{C} = \frac{e^{-\frac{1}{2}\int_t^{t+1}\eta(s)'\eta(s)ds - \int_t^{t+1}\eta(s)'dB^{\mathbb{P}}(s)}}{E_t^{\mathbb{P}}\left[e^{-\frac{1}{2}\int_t^{t+1}\eta(s)'\eta(s)ds - \int_t^{t+1}\eta(s)'dB^{\mathbb{P}}(s)}\right]} \tag{14}$$

linking $\mathbb{P}$ to $\mathbb{Q}$, subject to the requirement that X is a $\mathbb{Q}$-affine process. In principal, this construction places minimal restrictions on the $\mathbb{P}$-drifts of X. Starting with a $\mathbb{Q}$-affine model for X, one can generate essentially any functional form for the $\mathbb{P}$ drift of X by choice of the market price of risk η, up to the weak requirement that η not admit arbitrage opportunities.

[9] If X_t is a $DA_M^{\mathbb{Q}}(N)$ process, then setting $\delta_{Xi} > 0$ for $i > M$ is a normalization, but setting $\delta_{Xi} > 0$ for $i \leq M$ is a model restriction. When $M > 0$, this restriction ensures that (i) the level of the short rate r and the factors with stochastic volatility are positively correlated; and (ii) zero-coupon bond prices are well defined for any maturity. See Footnote 10 for further elaboration on the second point.

[10] When $M > 0$, the assumption $\delta_X > 0$ ensures that the first M elements of B_n are never negative. This in turn ensures that $a(\cdot)$ and $b(\cdot)$ are always evaluated in their admissible range in the recursion.

What has led researchers to focus on relatively restrictive specifications of $\eta(X_t)$ are the computational burdens of estimation that arise when the chosen η leads to an unknown (in closed form) $\mathbb{P}$-likelihood function for the observed bond yields.

In this section we introduce a discrete-time $\mathbb{P}$-formulation of affine *DTSM*s that overcomes this limitation of continuous-time models. This is accomplished by choosing a Radon-Nykodym derivative $(d\mathbb{P}/d\mathbb{Q})^D(X_{t+1},\Lambda_t)$ satisfying

$$f^{\mathbb{P}}(X_{t+1}|X_t) = (d\mathbb{P}/d\mathbb{Q})^D(X_{t+1};\Lambda_t) \times f^{\mathbb{Q}}(X_{t+1}|X_t), \tag{15}$$

with the properties that (**P1**) it is known in closed form (so that $f^{\mathbb{P}}$ can be derived in closed-form from our knowledge of $f^{\mathbb{Q}}$ developed in Section 2); (**P2**) Λ_t is naturally interpreted as the market price of risk of X_{t+1}; and (**P3**) rich nonlinear dependence of Λ_t on X_t is accommodated. In principle, any choice of $(d\mathbb{P}/d\mathbb{Q})^D$ that is a known function of (X_{t+1},Λ_t) and for which $\mathbb{P}$ and $\mathbb{Q}$ are equivalent measures (as required by the absence of arbitrage) leads to a nonlinear *DTSM* satisfying **P1**. We proceed by adopting the following particularly tractable choice of $(d\mathbb{P}/d\mathbb{Q})^D$:

Assumption 3 (N($\mathbb{P}$)) *The conditional density of X under the physical measure $\mathbb{P}$ is given by (15) with*

$$\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)^D (X_{t+1};\Lambda_t) = \frac{e^{\Lambda_t' X_{t+1}}}{\phi^{\mathbb{Q}}(\Lambda_t;X_t)}, \tag{16}$$

where $\phi^{\mathbb{Q}}$ is the conditional Laplace transform of X under $\mathbb{Q}$, Λ_t is a $N\times 1$ vector of functions of X_t satisfying $Prob\{\Lambda_t^i c_i < 1\} = 1$, for $1 \le \forall i \le M$, and $Prob\{\Lambda_t^i < \infty\} = 1$, for $M+1 \le i \le N$.

Under Assumption $N(\mathbb{P})$, the conditional $\mathbb{P}$-Laplace transform of X_t is given by

$$\phi^{\mathbb{P}}(u;X_t) \;=\; \frac{\phi^{\mathbb{Q}}(u+\Lambda_t;X_t)}{\phi^{\mathbb{Q}}(\Lambda_t;X_t)} = e^{\mathcal{A}(u;\Lambda_t)+\mathcal{B}(u;\Lambda_t)X_t}, \tag{17}$$

where $\mathcal{A}(u;v) \equiv a(u+v) - a(v)$ and $\mathcal{B}(u;v) \equiv b(u+v) - b(v)$. It follows that the pricing kernel consistent with Assumptions N($\mathbb{Q}$) and N($\mathbb{P}$) can be written as

$$\mathcal{M}_{t,t+1} = e^{-r_t} \times \frac{e^{-\Lambda_t' X_{t+1}}}{\phi^{\mathbb{P}}(-\Lambda_t;X_t)}, \tag{18}$$

where we have invoked (17) evaluated at $u = -\Lambda_t$, which leads to $\phi^{\mathbb{P}}(-\Lambda_t;X_t) = \left[\phi^{\mathbb{Q}}(\Lambda_t;X_t)\right]^{-1}$.

To motivate this choice of Radon-Nykodym derivative– equivalently, pricing kernel $\mathcal{M}$– consider again the continuous-time formulation in (14). For a small time interval Δ, and approximate affine state process $X_{t+\Delta} \approx \mu_X(X_t)\Delta + \Sigma_X\sqrt{S_{Xt}}\epsilon_{t+\Delta}$, with $\epsilon_{t+\Delta}|X_t \sim N(0,\Delta I)$,

$$\begin{aligned}(d\mathbb{Q}/d\mathbb{P})^C_{t,t+\Delta} \;&\approx\; \frac{e^{-\frac{1}{2}\eta_t'\eta_t\Delta - \eta_t'\epsilon^{\mathbb{P}}_{t+\Delta}}}{E^{\mathbb{P}}_t\left[e^{-\frac{1}{2}\eta_t'\eta_t\Delta - \eta_t'\epsilon^{\mathbb{P}}_{t+\Delta}}\right]} = \frac{e^{-\Lambda_t'\Sigma_X\sqrt{S_{Xt}}\epsilon^{\mathbb{P}}_{t+\Delta}}}{E^{\mathbb{P}}_t\left[e^{-\Lambda_t'\Sigma_X\sqrt{S_{Xt}}\epsilon^{\mathbb{P}}_{t+\Delta}}\right]} \\ &=\; \frac{e^{-\Lambda_t' X_{t+\Delta}}}{E^{\mathbb{P}}_t\left[e^{-\Lambda_t' X_{t+\Delta}}\right]} = \frac{e^{-\Lambda_t' X_{t+\Delta}}}{\phi^{\mathbb{P}}(-\Lambda_t;X_t)},\end{aligned} \tag{19}$$

where $\Lambda_t \equiv \left(\Sigma_X \sqrt{S_{Xt}}\right)'^{-1} \eta_t$ is a transformation of the market price of risk η_t. Thus, this (approximate) continuous-time construction suggests that, for a small discrete time interval of length Δ, the kernel for pricing payoffs at date $t+\Delta$ is

$$\mathcal{M}_{t,t+\Delta} \equiv e^{-r_t} \times \frac{f^{\mathbb{Q}}(X_{t+\Delta}|X_t)}{f^{\mathbb{P}}(X_{t+\Delta}|X_t)} \approx e^{-r_t \Delta} \frac{e^{-\Lambda_t' X_{t+\Delta}}}{\phi^{\mathbb{P}}(-\Lambda_t; X_t)}. \tag{20}$$

This kernel takes exactly the same form as (18).

Importantly, in deriving our actual pricing kernel we have dispensed with the "small time interval" construction. Instead, we are assuming that t indexes the sampling interval of the data which, as is conventional in discrete-time asset pricing models, is also assumed to index the appropriate interval for the chosen specification of the pricing kernel (18). Subject to this "matching condition," no approximations are involved in deriving either $f^{\mathbb{P}}(X_{t+1}|X_t)$ in (15) or the associated pricing kernel $\mathcal{M}_{t,t+1}$ in (18).

The preceding heuristic construction of $\mathcal{M}$ from a continuous-time model does suggest that, as the sampling interval of the data shrinks to zero,

$$\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)^{D}_{t,t+\Delta} \approx \left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)^{C}_{t,t+\Delta}. \tag{21}$$

As such, the $\mathbb{P}$ distributions of the bond yields implied by our families $DA_M^{\mathbb{Q}}(N)$, and associated market prices of risk Λ, capture essentially the same degree of flexibility inherent in the families $A_M^{\mathbb{Q}}(N)$ as one ranges across all admissible (arbitrage-free) specifications of the market prices of risk $\eta(X_t)$. It is in this sense that we view our framework as the discrete-time counterpart of the entire family of arbitrage-free, continuous-time affine *DTSM*s derived under the assumption that the $\mathbb{Q}$-representation of X resides in one of the families $A_M^{\mathbb{Q}}(N)$.

The restrictions in Assumption $\mathbf{N}(\mathbb{P})$ that the products $\Lambda_{it} c_i$, $1 \leq i \leq M$, for the M volatility factors are bounded by unity are required to assure that $f^{\mathbb{P}}$ is a well-defined probability density function and that $\mathbb{P}$ and $\mathbb{Q}$ are equivalent measures. This follows from the observation that $\phi^{\mathbb{Q}}(u; X_t)$ is finite if and only if $u_i c_i < 1$. Unless $\Lambda_{it} c_i < 1$ almost surely, for $i = 1, \ldots, M$, $\phi^{\mathbb{Q}}(\Lambda_t; X_t)$ is infinite with positive probability. In this case, $f^{\mathbb{P}}$ would not integrate to unity for a set of X_t that has positive measure, and $\mathbb{P}$ and $\mathbb{Q}$ would not be equivalent. Examining these restrictions more closely, and using our mapping to the parameters of the related *CIR* process, we see that we are effectively requiring that $2/(\sigma_i^2 \Delta t) > \Lambda_{it}$, $i = 1, \ldots, M$. Typically σ_i^2 is small and, depending on the application, Δt may also be small. Therefore, these bounds are typically weak and in the applications we have encountered so far they are far from binding. As Δt approaches zero (continuous time), the only requirement is that the Λ_{it} be finite almost surely.[11]

Under these regularity conditions we have all of the information necessary to construct the likelihood function of the state, and hence the bond yields, under $\mathbb{P}$. Under Assumptions N($\mathbb{Q}$) and N(r), we effectively know $f^{\mathbb{Q}}(X_{t+1}|X_t)$ from the cross-sectional behavior of

[11]Note that, if Λ_{it} were to scale with $(\Delta t)^{-1}$, the continuous-time limit would be different from a CIR model.

bond yields.[12] Furthermore, the relationship between the observed yields y_t and the state vector X_t are also known due to the pricing equation (11), which depends only on the risk-neutral distribution $f^{\mathbb{Q}}(X_{t+1}|X_t)$. Thus, the unknown function $(d\mathbb{P}/d\mathbb{Q})^D(X_{t+1};\Lambda_t)$ can be estimated from the time-series observations of bond yields, y_t.

The form of the $\mathbb{P}$ distribution of X bears a close similarity to its form under $\mathbb{Q}$. Upon making the dependence of $a(\cdot)$ and $b(\cdot)$ on the risk-neutral parameters explicit by writing

$$\begin{aligned} a(u) &= a(u;\Theta^{\mathbb{Q}}),\ b(u) = b(u;\Theta^{\mathbb{Q}}), \\ \Theta^{\mathbb{Q}} &= (c_i, \rho_i, \nu_i; \mu_0, \mu, h_0, h_i : i = 1, 2, \ldots, M)\,, \end{aligned}$$

$\mathcal{A}(u,v)$ and $\mathcal{B}(u,v)$ can be written as

$$\begin{aligned} \mathcal{A}(u,v) &= a(u;\Theta^{\mathbb{P}}(v)),\ \mathcal{B}(u,v) = b(u;\Theta^{\mathbb{P}}(v)), \\ \Theta^{\mathbb{P}}(v) &= (c_i(v), \rho_i(v), \nu_i; \mu_0(v), \mu(v), h_0, h_i : i = 1, 2, \ldots, M)\,. \end{aligned}$$

where $v' = (v'_Z, v'_Y)$, for $M \times 1$ vector v_Z and $(N - M) \times 1$ vector v_Y, and

$$\begin{aligned} c_i(v) &= \frac{c_i}{1 - v_{Z,i}c_i},\ \rho_i(v) = \frac{\rho_i}{(1 - v_{Z,i}c_i)^2}, \\ \mu_0(v) &= \mu_0 + h'_0 v_Y,\ \mu_Y(v) = \begin{pmatrix} \mu_Y^Z + \{h'_i v_Y\}_{i=1,2,\ldots,M} & \mu_Y^Y \end{pmatrix}. \end{aligned}$$

It follows that the conditional density under $\mathbb{P}$ has exactly the same functional form as that under $\mathbb{Q}$, except that the latter is now evaluated at the (possibly time-varying) parameters $\Theta^{\mathbb{P}}(\Lambda_t)$.[13] Analogously to the continuous-time case, the volatility parameters $\{\nu_i\}_{i=1}^M$ (for the M stochastic volatility factors), and h_0 and $\{h_i\}_{i=1}^M$ (for the $N - M$ conditional Gaussian factors), are not affected by the measure change. Nevertheless, all of the conditional moments of X in model $(DA_M^{\mathbb{Q}}(N), \Lambda)$, including the one-period conditional variances, are in general affected by the measure change from $\mathbb{Q}$ to $\mathbb{P}$ (see below).

4 The Market Prices of Risk

An immediate implication of Assumption $N(\mathbb{P})$ is that, if $\Lambda_t = 0$, then $f^{\mathbb{P}}(X_{t+1}|X_t) = f^{\mathbb{Q}}(X_{t+1}|X_t)$. Thus, agents' market prices of risk are zero if and only if $\Lambda_t = 0$. In our discrete-time setting, Λ_t is not literally the market price of X risk (*MPR*), but rather the *MPR* is a nonlinear (deterministic) function of Λ_t. However, in a sense that we now make precise, Λ_t is the dominant term in the *MPR*. Accordingly, we will refer to Λ_t as *the MPR* as this will facilitate comparisons with the *MPR* in continuous-time $(A_M^{\mathbb{Q}}(N), \eta)$ models.

[12] Intuitively, taking the leading principal components as the state vector, we can estimate δ_0, δ_X, A_n, and B_n by regressing bond yields on this state vector. The parameters that characterize $f^{\mathbb{Q}}(X_{t+1}|X_t)$ can then be estimated by treating the recursions (12) and (13) as (possibly nonlinear) cross-equation restrictions.

[13] This observation is extremely useful for simulating the state process under $\mathbb{P}$: the next state can be simulated exactly using the $\mathbb{Q}$ density, with the parameters adjusted to reflect the state dependence induced by the measure change.

Notice first of all that[14]

$$\begin{aligned} E_t^{\mathbb{P}}\left[X_{t+1}\right] - E_t^{\mathbb{Q}}\left[X_{t+1}\right] &= \left[\mathcal{A}^{(1)}(0;\Lambda_t) - a^{(1)}(0)\right] + \left[\mathcal{B}^{(1)}(0;\Lambda_t) - b^{(1)}(0)\right] X_t \\ &= V_t^{\mathbb{P}}[X_{t+1}] \times \Lambda_t + o(\Lambda_t), \end{aligned}$$

where $V_t^{\mathbb{P}}[\cdot]$ is the conditional covariance matrix under $\mathbb{P}$. Ignoring the higher order terms, the above relationship is exactly what arises in diffusion-based models: Λ_t is the vector of market prices of risk underlying the adjustment to the "drift" in the change of measure from $\mathbb{Q}$ to $\mathbb{P}$. Moreover, the continuously compounded, expected excess return on the security with the payoff $e^{-c'X_{t+1}}$ is

$$\begin{aligned} E_t^{\mathbb{P}}\left[\log \frac{e^{-c'X_{t+1}}}{E_t^{\mathbb{Q}}\left[e^{-r_t}e^{-c'X_{t+1}}\right]}\right] - r_t &= -\left[a(-c) + c'a^{(1)}(\Lambda_t)\right] - \left[b(-c) + c'b^{(1)}(\Lambda_t)\right] X_t, \\ &= -c'V_t^{\mathbb{P}}[X_{t+1}] \times \Lambda_t + o(c) + o(\Lambda_t). \end{aligned} \quad (22)$$

Since c determines the exposure of this security to the factor risk X and $V_t^{\mathbb{P}}[X_{t+1}]$ measures the size of the risk, the random variable Λ_t is the dominant term in the true market price of risk underlying expected excess returns.

A notable difference between Λ_t and the market price of risk η that appears in continuous-time $(A_M^{\mathbb{Q}}(N), \eta)$ models is that Λ_t measures the price of risk per per unit of variance, whereas η measures risk in units of standard deviation. From the heuristic derivation of our choice of $(d\mathbb{P}/d\mathbb{Q})^D$ it is seen that this difference is simply a consequence of our convention that

$$\Lambda_t = \left(\Sigma_X \sqrt{S_{Xt}}\right)^{\prime -1} \eta_t. \quad (23)$$

Our strategy for developing a fully specified model $(DA_M^{\mathbb{Q}}(N), \Lambda)$ will be to specify the $\mathbb{Q}$ distribution of X; specify Λ_t through (23) by adopting a specification η_t; and then to use the resulting specification of $(d\mathbb{P}/d\mathbb{Q})^D(X_{t+1}, \Lambda_t)$ to derive the $\mathbb{P}$ distribution X and the likelihood function of the bond yields. Following this approach, the resulting model automatically satisfies **P1 - P3**. In particular, the modeler has complete freedom to specify the dependence of Λ_t on X_t (**P3**), while preserving **P1**. Moreover, by substituting (23) into (16) to construct $f^{\mathbb{P}}(X_{t+1}|X_t)$ within the model $(DA_M^{\mathbb{Q}}(N), \Lambda)$, we assure that the resulting model fully accounts for any higher-order (nonlinear) terms in the actual *MPR*.

To better understand the nature of the potential nonlinearity inherent in our modeling framework it is instructive to examine in more detail the model-implied first and second $\mathbb{P}$-moments of X. Let Λ_{Zt} and Λ_{Yt} form a conformal partition of Λ_t. The conditional mean of the i^{th} member of the M-vector of volatility factors Z_{t+1}, under $\mathbb{P}$, is given by

$$E_t^{\mathbb{P}}\left[Z_{t+1}^i\right] = \frac{\partial}{\partial u_{Zi}}\left[\mathcal{A}(u;\Lambda_t) + \mathcal{B}(u;\Lambda_t)X_t\right]\Big|_{u=0} = \frac{\nu_i c_i}{1-\Lambda_{Zt,i}c_i} + \frac{\rho_i}{(1-\Lambda_{Zt,i}c_i)^2} Z_t. \quad (24)$$

[14]The derivatives of $\mathcal{A}$ and $\mathcal{B}$ are with respect to their first arguments.

Similarly, the conditional variance of Z^i_{t+1} is given by

$$\mathrm{Var}^{\mathbb{P}}_t[Z^i_{t+1}] = \frac{\nu_i c_i^2}{(1-\Lambda_{Zt,i}c_i)^2} + \frac{2c_i\rho_i Z_t}{(1-\Lambda_{Zt,i}c_i)^3},\ \ i=1,\ldots,M. \tag{25}$$

The nonlinearity of these moments, in contrast to their affine counterparts under $\mathbb{Q}$ (see (6)), is induced by the state-dependence of $\Lambda_{Zt,i}$ through the terms $1/(1-\Lambda_{Zt,i}c_i)$.

Suppose that Λ_t is parameterized, by choice of η_t, in the same manner as in $(A^{\mathbb{Q}}_M(N),\eta)$ models. Within the canonical $A^{\mathbb{Q}}_M(N)$ model Σ_X is block-diagonal with the $M\times M$ upper block being I_M, and the lower $(N-M)\times(N-M)$ block being Σ_Y. So, in particular, (from (23)) $\Lambda_{Zt} = (diag[Z^i_t])^{-1/2}\,\eta_{Zt}$. The special case of Dai and Singleton (2000)'s "completely affine" specification of η_Z has $\eta_{Zt} = (diag[Z^i_t])^{1/2}\,\lambda_{Z1}$, for constant $M\times 1$ vector λ_{Z1}. Therefore, under their *MPR*, Λ_{Zt} is constant and (24) and (25) imply that the conditional moments of Z are affine under $\mathbb{P}$ as well as under $\mathbb{Q}$. In other words, completely affine $(A^{\mathbb{Q}}_M(N),\eta)$ and $(DA^{\mathbb{Q}}_M(N),\Lambda)$ models both imply that Z_t follows an affine process under $\mathbb{P}$. A special case of this construction is the $(DA_1(1),\Lambda)$ model examined by Gourieroux, Monfort, and Polimenis (2002).

A more general formulation of Λ_{Zt} that nests the specifications (of η_Z) adopted in Duffee (2002), Duarte (2004), and Cheridito, Filipovic, and Kimmel (2003) has

$$\Lambda_{Zt} = \left(\sqrt{diag[Z^i_t]}\right)^{-1}\left(\lambda_{Z0} + \sqrt{diag[Z^i_t]}\,(\lambda_{Z1}+\lambda_{Z2}Z_t) + \Upsilon_{Zt}\right), \tag{26}$$

where λ_{Z0} and λ_{Z1} are $M\times 1$ vectors and λ_{Z2} is an $M\times M$ matrix. Setting $\lambda_{Z0}=0$ and $\Upsilon_{Zt}=0$ yields the model in Cherdito, et. al. The special case of $\lambda_{Z2}=0$ and $\Upsilon_{Zt}=0$ gives Duarte's model. So long as either $\lambda_{Z0}\neq 0$ or $\lambda_{Z2}\neq 0$, Λ_{Zt} is state-dependent and the conditional $\mathbb{P}$-moments of Z_{t+1} show nonlinear dependence on Z_t. The term Υ_{Zt} is introduced to show illustrate that the modeler is free to add essentially any nonlinear dependence of Λ_{Zt} on X. We investigate empirically a model with additional nonlinear terms in Section 5.

Turning to the conditionally Gaussian components of X, and recalling the definitions in (7), the conditional mean of Y_{t+1} under $\mathbb{P}$ is

$$E^{\mathbb{P}}_t[Y_{t+1}] = \omega^{\mathbb{Q}}_{Yt} + \Omega_{Yt}\Lambda_{Yt}. \tag{27}$$

To interpret the consequences of alternative specifications of Λ for the functional form of $E^{\mathbb{P}}_t[Y_{t+1}]$, it is instructive to express the market price of risk for the entire state vector X in an $(DA^{\mathbb{Q}}_M(N),\Lambda)$ model as

$$\begin{aligned}\Lambda_{Xt} &= (\Sigma_X\sqrt{S_{Xt}})'^{-1}\left(\sqrt{S_{Xt}}\lambda_0 + \sqrt{S^{-1}_{Xt}}I^-_N\lambda_1X_t + \sqrt{S^{-1}_{Xt}}\Sigma^{-1}_X\sqrt{S_{Xt}}\lambda_d + \sqrt{S^{-1}_{Xt}}\Sigma^{-1}_X\Upsilon_{Xt}\right)\\ &= (\Sigma_XS_{Xt}\Sigma'_X)^{-1}\left(\Sigma_XS_{Xt}\lambda_0 + \Sigma_XI^-_N\lambda_1X_t + \sqrt{S_{Xt}}\lambda_d + \Upsilon_{Xt}\right),\end{aligned} \tag{28}$$

where I^-_N is the $N\times N$ identity matrix with the first M diagonal elements set to zero, λ_0 and λ_d are $N\times 1$ vectors of constants, λ_1 is an $N\times N$ matrix of constants, and Υ_{Xt} can be any

$N \times 1$ vector of non-linear functions of X. The term in with λ_0 captures the completely affine term from *DS*, adding the term with λ_1 gives Duffee (2002)'s essentially affine specification, and the term with λ_d incorporates Duarte (2004)'s extension of Duffee's model.[15]

Examination of the implied conditional $\mathbb{P}$-mean of Y_{t+1} is facilitated by imposing, as in the continuous-time canonical $A_M^{\mathbb{Q}}(N)$ models in *DS*, the normalization that $\Sigma_X = I_N$. With this normalization, the subvector of market prices of risk associated with Y becomes

$$\Lambda_{Yt} = S_{Yt}^{-1}\left(S_{Yt}\lambda_{Y0} + \lambda_{Y1}X_t + \sqrt{S_{Yt}}\lambda_{Yd} + \Upsilon_{Yt}\right), \tag{29}$$

where λ_{Y1} is the $(N-M)\times N$ matrix containing the last $N-M$ rows of λ_1. Substituting (29) into (27) gives

$$E^{\mathbb{P}}[Y_{t+1}|Y_t] = \omega_{Yt}^{\mathbb{Q}} + S_{Yt}\lambda_{Y0} + \lambda_{Y1}X_t + \sqrt{S_{Yt}}\lambda_{Yd} + \Upsilon_{Yt}. \tag{30}$$

It follows that the completely and essentially affine components of Λ_{Yt} contribute an affine function of X to the conditional mean of Y_{t+1}. Duarte's added term in Λ_t introduces a nonlinear term to the drift of Y; specifically, the square roots of affine functions of X. Finally, to illustrate the flexibility of specifying the conditional mean within our family of nonlinear *DTSM*s, we have added the term Υ_{Yt} and given the modeler essentially complete freedom in specifying its functional dependence on X. Note in particular that, by an appropriate choice of Υ_{Yt}, we can replicate the nonlinear dependence of the drifts documented in the non-parametric analysis of Ait-Sahalia (1996). For any choice of Υ_{Yt}, the conditional $\mathbb{P}$ distribution of X, and hence the likelihood function of the data, are known in closed form.

What our formulation of the $(DA_M^{\mathbb{Q}}(N), \Lambda)$ model does not allow is complete freedom in specifying the nonlinearity of higher order moments, once we have chosen a functional form for the conditional first moment. This is illustrated by the first two moments of the autoregressive gamma process. The conditional means and variances depend on $1/(1 - \Lambda_{Zt,i}c_i)$ in a nearly symmetric way (compare (24) with (25)). Indeed, the variance has a very similar structure to the mean, except that each term is divided by one higher power of $(1 - \Lambda_{Zt,i}c_i)$. Thus, the nonlinear dependence in the mean achieved by one's choice of Λ_{Zt} effectively determines the structure of the nonlinearity of the conditional second moments. This specialized structure, which is a consequence of Assumption N($\mathbb{P}$), is the discrete-time counterpart to the similarly special structure on moments implied by diffusion models. An interesting question for future research is the feasibility of working with even richer pricing kernels, while preserving the tractability of the resulting $(DA_M(N), \Lambda)$ models.

Though we have allowed for considerable flexibility in specifying the dependence of Λ_t on X_t, it is desirable to impose sufficient structure on Λ_t to assure that the maximum likelihood estimator of $\Theta^{\mathbb{P}}$ has a well-behaved large-sample distribution. One property of the $\mathbb{P}$ distribution of X that takes us a long ways toward assuring this is geometric ergodicity.[16]

[15]Note that, under the normalization that $\Sigma_X = I_N$, which we impose in our subsequent empirical examples, the third term in the line above (28) simplifies to I_N, just as in Duarte's specification of η_t.

[16]See Duffie and Singleton (1993) for definitions and applications of geometric ergodicity in the context of generalized method of moments estimation. General criteria for the geometric ergodicity of a Markov chain have been obtained by Nummelin and Tuominen (1982) and Tweedie (1982).

That X will not be a geometrically ergodic process for all specifications of Λ_t can be seen immediately from (24). If $\Lambda_{Zt,i}$ approaches c_i as Z_t^i increases, then the second term eventually dominates and the state variable is explosive under $\mathbb{P}$. Similarly, if $\Omega_{Yt}\Lambda_t$ in (27) sufficiently amplifies the effect of X_t on Y_{t+1}, then Y will be explosive under $\mathbb{P}$.

Such explosive behavior is ruled out by geometric ergodicity since, intuitively, the latter assures that a Markov process converges to its ergodic distribution at a geometric rate. The following proposition provides sufficient conditions for the geometric ergodicity of an autoregressive gamma processes (see Appendix A for the proof).

Proposition 1 (G.E.(Z)) *Suppose that the market prices of risk* $\Lambda_Z(Z_t)$ *is a continuous function of* Z_t*, and the eigenvalues of the matrix* ρ*,* ψ_i *(*$i = 1, 2, \ldots, M$*), satisfy* $\max_i |\psi_i| < 1$*. If, in addition,*

1. $\Lambda_Z(z) \leq 0$ *for* $\forall z \geq 0$*, or*

2. $\Lambda_Z(z) \rightarrow \bar{\lambda} \leq 0$ *and* $\rho_{ij} = 0$ *for* $0 \leq i \neq j \leq M$*,*

then Z_t *is geometrically ergodic under both* $\mathbb{Q}$ *and* $\mathbb{P}$*.*

Central to the geometric ergodicity of the $\mathbb{P}$ distribution of Z is the behavior of Λ_{Zt} for $\|Z\| > K$, for some positive constant K. Applying Proposition **G.E.(Z)** to the specification (26) of Λ_{Zt}, we note first of all that the restriction $\lambda_1 < 0$ is required to replicate the upward sloping yield curve observed historically, on average. For a one-factor model ($M = 1$), Proposition **G.E.(Z)** implies that this sign restriction and the assumption that $\sqrt{Z_t}\lambda_{Z2} + \Upsilon_{Zt}$ is a bounded function of Z are sufficient for Z_t to be geometrically ergodic. Since we are free to set the bound at a very large number, for practical purposes, once we have imposed the sign restriction on λ_1 called for by the historical data we obtain geometric ergodicity. If $M > 1$, then the correlations among the Z^i will potentially affect the geometric ergodicity of Z. Sufficient conditions for geometric ergodicity would involve a bound on some terms in Λ_{Zt} and imposition of the sign restriction $\lambda_0 < 0$, though these conditions may be stronger than necessary.

The challenge of formally establishing geometric ergodicity for the entire state vector X_t is naturally even more complex, because of the range of possible specifications of Λ_{Yt}, many of which lead to models that lie outside those considered in the literature on geometric ergodicity. For this reason researchers will most likely have to treat the issue geometric ergodicity on a case-by-case basis, as we do in the following illustrations.

5 Empirical Illustrations

In this section we illustrate the flexibility of our modeling framework by estimating nonlinear $(DA_1^{\mathbb{Q}}(3), \Lambda)$ models that nest several of the linear $(A_1^{\mathbb{Q}}(3), \eta)$ models in the published literature. In presenting these models, we adopt the notation of continuous time, leaving the mappings between these parameters and the primitive parameters of our $DA_1^{\mathbb{Q}}(3)$ models presented in Section 2 implicit. We stress that this is only for notational convenience and ease of comparison with the reported estimates in the literature on continuous-time models. Our nonlinear *DTSM*s are parameterized by writing down a continuous-time model, parameterizing the drift, diffusion term, and the market prices of risk, and then mapping these parameters to those of our discrete-time conditional Gaussian and autoregressive-gamma processes. In the end, it is the likelihood functions of these nonlinear discrete-time models that we estimate. Further, all of the subsequent calculations of moments of the processes are based on the moments of these exact discrete-time pricing models.

With these implicit mappings in the background, the model we examine is:

$$dX_t = \kappa^{\mathbb{Q}}(\theta^{\mathbb{Q}} - X_t)dt + \Sigma_X \sqrt{S(t)} dB_t^{\mathbb{Q}} \tag{31}$$

where $S(t) = diag(\alpha + \beta X_t)$ and $dB^{\mathbb{Q}}$ is an N-vector of independent standard Brownian motions under $\mathbb{Q}$; and the one-period (monthly) short rate is an affine function of X, $r_t = \delta_0 + \delta_X' X_t$, where $\delta_X = (\delta_Z', \delta_Y')'$, which implies similar linear dynamics for bonds of other maturities:

$$y_t^n = \delta_0^n + \delta_X^{n\prime} X_t \tag{32}$$

Following Dai and Singleton (2000), we impose the following normalizations for econometric identification of the models:

$$\kappa^{\mathbb{Q}} = \begin{bmatrix} \kappa^{ZZ} & 0 \\ \kappa^{YZ}_{2\times 1} & \kappa^{YY}_{2\times 2} \end{bmatrix}; \tag{33}$$

$\Sigma_X = I_3$; and

$$\theta^{\mathbb{Q}} = \begin{bmatrix} \theta^Z \\ 0_{2\times 1} \end{bmatrix},\ \alpha = \begin{bmatrix} 0 \\ 1_{2\times 1} \end{bmatrix},\ \beta = \begin{bmatrix} 1 & 0_{1\times 2} \\ \beta^{YZ}_{2\times 1} & 0_{2\times 2} \end{bmatrix} \tag{34}$$

Two different specifications of the market prices of risk, leading to two different nonlinear discrete-time *DTSM*s, were examined:

Duarte's $SASR_1(3)$ Model: The market prices of risk are given by (26) with $\lambda_{Z2} = 0$ and $\Upsilon_{Zt} = 0$ and (29) with $\lambda_{Yd} = 0$ and $\Upsilon_{Yt} = 0$.

$DA_1^{\mathbb{Q}}(3)$ **Model with Nonlinear Drift in Volatility:** The market prices of risk are given by (26) with $\lambda_{Z0} = 0$ and Υ_{Zt} being an affine function of Z^2 and Z^3, and (29) with $\lambda_{Yd} = 0$ and $\Upsilon_{Yt} = 0$.

Relying again on the notation of continuous time, these formulations of the market price of risk imply a $\mathbb{P}$-drift of the form

$$\begin{pmatrix} \kappa^{\mathbb{P}}_{1,1} & 0 & 0 \\ \kappa^{\mathbb{P}}_{2,1} & \kappa^{\mathbb{P}}_{2,2} & \kappa^{\mathbb{P}}_{2,3} \\ \kappa^{\mathbb{P}}_{3,1} & \kappa^{\mathbb{P}}_{3,2} & \kappa^{\mathbb{P}}_{3,3} \end{pmatrix} \begin{pmatrix} \theta^{\mathbb{P}}_1 & & Z_t \\ \theta^{\mathbb{P}}_2 & - & Y_{t,1} \\ \theta^{\mathbb{P}}_3 & & Y_{t,2} \end{pmatrix} + \begin{pmatrix} \lambda_d \sqrt{Z_t} + \lambda_{2(1,1)} Z_t^2 + \lambda_{3(1,1)} Z_t^3 \\ 0 \\ 0 \end{pmatrix}. \tag{35}$$

Duarte's model is the special case with $\lambda_{2(1,1)} = \lambda_{3(1,1)} = 0$, and our nonlinear model is the special case with $\lambda_d = 0$.

Exploration of alternative nonlinear formulations of the drift of the risk factor determining stochastic volatility is motivated by several observations. First, one of our goals is to expand the extant focus in the literature on discrete-time *DTSM*s from $DA_0^{\mathbb{Q}}(N)$ models to $DA_M^{\mathbb{Q}}(N)$ models with $M > 0$, so focusing on a model with stochastic volatility in the risk factors seemed natural. Second, in its continuous-time formulation, Duarte's model does not have a known likelihood function and, therefore, he had to resort to approximations in computing his *ML* estimates. The likelihood function of our discrete-time counterpart is known in closed-form and so his model provides an interesting example of the tractability obtained by shifting to discrete time. Finally, and most central to the literature on term structure modeling, the goodness-of-fit of affine *DTSM*s has been constrained both by the requirements of admissibility of $A_M^{\mathbb{Q}}(M)$ models and the standard formulations of the market prices of risk for Z. Of interest, then, is the relative fits of $DA_1^{\mathbb{Q}}(3)$ models with alternative extended formulations of Λ_{Zt} with their induced nonlinearity the the $\mathbb{P}$ distributions of Z. Even more general formulations of this model are possible by allowing Λ_t to induce nonlinearity in the drifts of all three state variables.

The models were estimated using "smoothed" Fama-Bliss monthly data on treasury zero-coupon bond yields from 1970:1 to 1995:12. This is the same data used in Backus, Foresi, and Zin (1998) and Dai and Singleton (2002). We assumed that bonds with .5, 2 and 10 years to maturity were priced without errors, while bonds with maturities of .25, 1 and 5 years were priced with serially independent Gaussian errors. Data for the period 1996-2000 was omitted from the estimation in order to examine the out-of-sample predictability of excess returns. A primary motivation for Duffee (2002)'s essentially affine model is that his more flexible specification of market prices of risk substantially improves the ability of affine models to match the persistence in excess returns. A question that we address with our nonlinear $(DA_1^{\mathbb{Q}}(3), \Lambda)$ model is whether there is a further improvement in the out-of-sample performance of $DA_1^{\mathbb{Q}}(3)$ models due to the introduction of a more general market price of volatility risk.

The *ML* estimates of the parameter of three models are displayed in Tables 1 and 2: the linear $DA_1^{\mathbb{Q}}(3)$ with all three terms inducing nonlinearity in (35) set to zero, the nonlinear $DA_1^{\mathbb{Q}}(3)$ model in which $\lambda_{2(1,1)}$ and $\lambda_{3(1,1)}$ are nonzero, and Duarte's model in which λ_d is nonzero. Focusing first on these nonlinear terms, the likelihood ratio statistic of 11.23 indicates rejection of null hypothesis that $\lambda_{2(1,1)}$ and $\lambda_{3(1,1)}$ in the nonlinear $DA_1(3)$ are zero at conventional significance levels. The significantly negative coefficient on the cubic term suggests that Z is strongly mean-reverting in more volatile times, while behaving similar to a random walk under relatively stable conditions. In Duarte's $SASR_1(3)$ model, λ_d is positive and of similar magnitude to his estimate, though it carries a relatively large standard error.

Figure 1 displays the loadings on the state variables for zero-coupon bonds with maturities ranging from six months to ten years. The loadings on the volatility factor X_1 in the nonlinear $DA^1(3)$ looks similar to those of a "curvature" factor. This is consistent with link between volatility and the convexity of the yield curve (Litterman, Scheinkman, and Weiss (1991)). In contrast, the volatility factor in model $SASR_1(3)$ has loadings reminiscent of a slope factor. The properties of X_1 are also very different under $\mathbb{Q}$. In our nonlinear $DA_1(3)$ model, X_1 has an intermediate degree of mean reversion, with factor X_2 having the slowest rate of mean reversion and X_3 having the fastest (see the diagonal elements of $\kappa^{\mathbb{Q}}$ in Table 2). On the

ML estimates	**Linear Model $DA_1(3)$**		**Nonlinear Model $DA_1(3)$**		**Duarte Model**	
$\kappa^{\mathbb{P}}(1,1)$	0.627	(6.041)	2.712	(2.801)	0.295	(1.128)
$\kappa^{\mathbb{P}}(2,1)$	0.219	(0.021)	-0.476	(-0.063)	0.000	
$\kappa^{\mathbb{P}}(3,1)$	-5.251	(-0.095)	-3.538	(-0.175)	0.000	
$\kappa^{\mathbb{P}}(1,2)$	0.000		0.000		0.000	
$\kappa^{\mathbb{P}}(2,2)$	0.224	(1.601)	0.149	(1.169)	0.416	(1.666)
$\kappa^{\mathbb{P}}(3,2)$	0.000		0.000		0.000	
$\kappa^{\mathbb{P}}(1,3)$	0.000		0.000		0.000	
$\kappa^{\mathbb{P}}(2,3)$	-1.517	(-0.024)	-1.412	(-0.081)	6.980	(0.093)
$\kappa^{\mathbb{P}}(3,3)$	1.594	(5.359)	1.539	(5.555)	2.686	(6.002)
$\theta^{\mathbb{P}}(1,1)$	5.971	(3.587)	1.460	(3.378)	0.331	
$\theta^{\mathbb{P}}(2,1)$	0.000		0.000		0.000	
$\theta^{\mathbb{P}}(3,1)$	0.000		0.000		0.000	
$\delta_0(1,1)$	0.050	(2.099)	-0.055	(-0.709)	0.046	(3.969)
$\delta_Y(1,1)$	0.003	(5.467)	0.003	(4.512)	0.002	(5.219)
$\delta_Y(2,1)$	0.000	(0.024)	0.001	(0.070)	0.001	(0.092)
$\delta_Y(3,1)$	0.002	(0.094)	0.003	(0.172)	0.004	(3.698)
$\lambda^*(1,1)$					0.914	(1.022)
$\beta(1,1)$	1.000		1.000		1.000	
$\beta(2,1)$	74.514	(0.012)	73.727	(0.035)	23.774	(0.046)
$\beta(3,1)$	12.196	(0.047)	7.942	(0.085)	0.145	(1.765)
$\lambda_2(1,1)$			0.812	(3.246)		
$\lambda_3(1,1)$			-0.071	(-2.621)		
Log Likelihood	**30.612**		**30.630**		**30.673**	
Test Ratio			**11.232**			

Table 1: *ML* estimates for two $DA_1(3)$ models with and without non-linear drift and $SASR_1(3)$ model. t-statistics are included in the parenthesis. Those without a t-statistic are implied from estimated parameters. The models are fitted to monthly observations of zero yields with maturities equal to 3, 6 months, 1, 2, 5 and 10 years.

ML estimates	**Linear Model $DA_1(3)$**		**Nonlinear Model $DA_1(3)$**		**Duarte Model**	
$\kappa^{\mathbb{Q}}(1,1)$	0.552		0.553		-0.018	
$\kappa^{\mathbb{Q}}(2,1)$	-0.108		-2.567		-0.648	
$\kappa^{\mathbb{Q}}(3,1)$	-6.400		-4.992		-0.019	
$\kappa^{\mathbb{Q}}(1,2)$	0.000		0.000		0.000	
$\kappa^{\mathbb{Q}}(2,2)$	0.003		0.003		0.543	
$\kappa^{\mathbb{Q}}(3,2)$	-0.132		0.000		0.000	
$\kappa^{\mathbb{Q}}(1,3)$	0.000		0.000		0.000	
$\kappa^{\mathbb{Q}}(2,3)$	0.007		0.972		17.211	
$\kappa^{\mathbb{Q}}(3,3)$	2.013		1.998		1.841	
$\theta^{\mathbb{Q}}(1,1)$	6.787		7.157		-5.457	
$\theta^{\mathbb{Q}}(2,1)$	750.533		841.397		-4.542	
$\theta^{\mathbb{Q}}(3,1)$	54.126		15.385		0.015	
$\lambda_0(1,1)$	-0.075	(-0.746)	-2.159	(-2.250)	-0.313	(-1.198)
$\lambda_0(2,1)$	-0.202	(-0.147)	0.000		-1.328	(-0.098)
$\lambda_0(3,1)$	2.299	(0.085)	-0.183	(-0.171)	-0.133	(-1.206)
$\lambda_1(1,1)$	0.000		0.000		0.000	
$\lambda_1(2,1)$	14.704	(0.013)	-2.092	(-0.071)	30.929	(0.031)
$\lambda_1(3,1)$	-29.188	(-0.031)	0.000		0.000	
$\lambda_1(1,2)$	0.000		0.000		0.000	
$\lambda_1(2,2)$	-0.222	(-1.676)	-0.146	(-1.144)	0.127	(0.514)
$\lambda_1(3,2)$	-0.132	(-0.024)	0.000		0.000	
$\lambda_1(1,3)$	0.000		0.000		0.000	
$\lambda_1(2,3)$	1.523	(0.024)	2.384	(0.081)	10.231	(0.092)
$\lambda_1(3,3)$	0.419	(1.307)	0.460	(1.472)	-0.845	(-1.838)

Table 2: *ML* estimates for two $DA_1(3)$ models with and without non-linear drift and $SASR_1(3)$ model. t-statistics are included in the parenthesis. Those without a t-statistic are implied from estimated parameters. The models are fitted to monthly observations of zero yields with maturities equal to 3, 6 months, 1, 2, 5 and 10 years.

other hand, the volatility factor X_1 is explosive under $\mathbb{Q}$ in model $SASR_1(3)$. X_2 has the intermediate level of mean reversion and X_3 has the slowest rate of mean reversion.

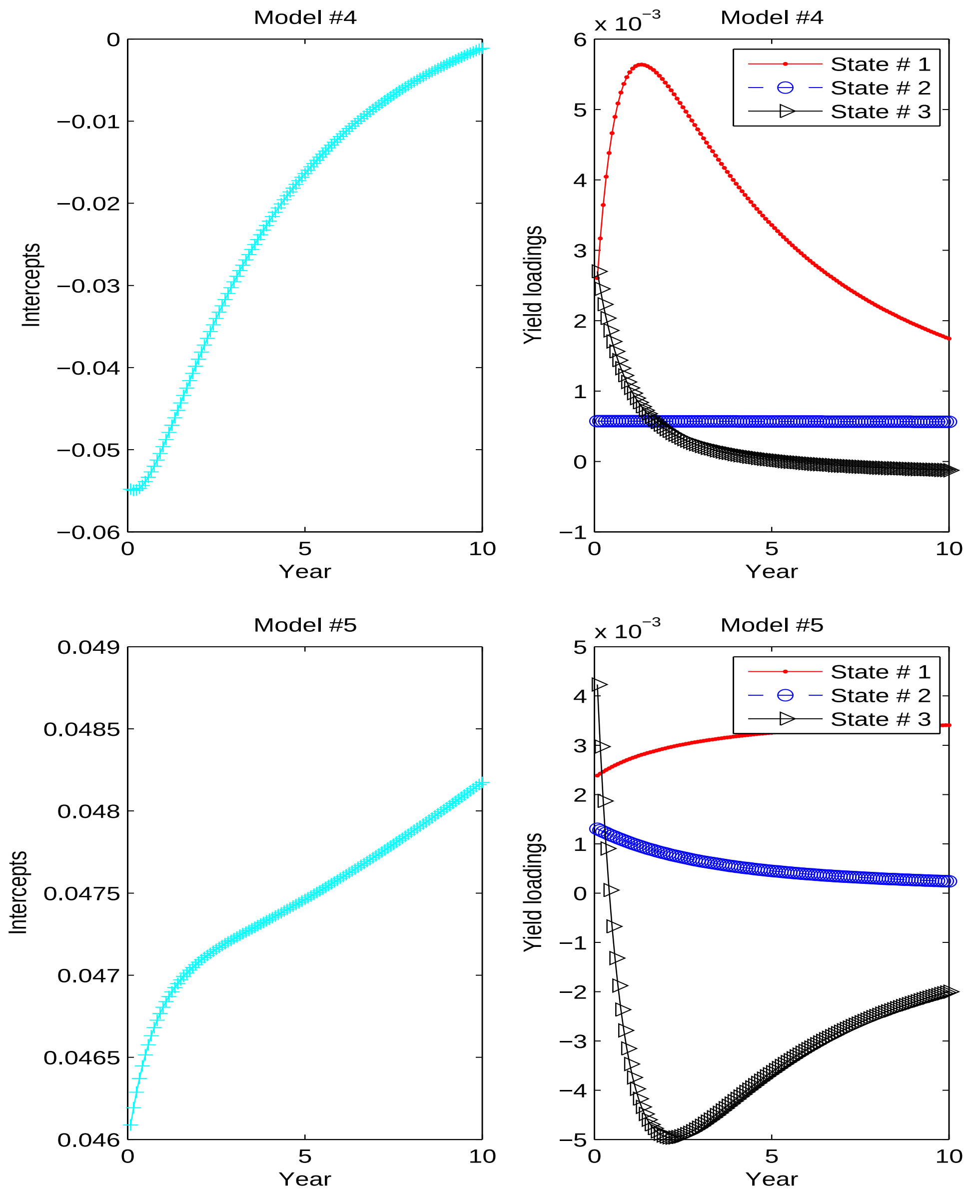
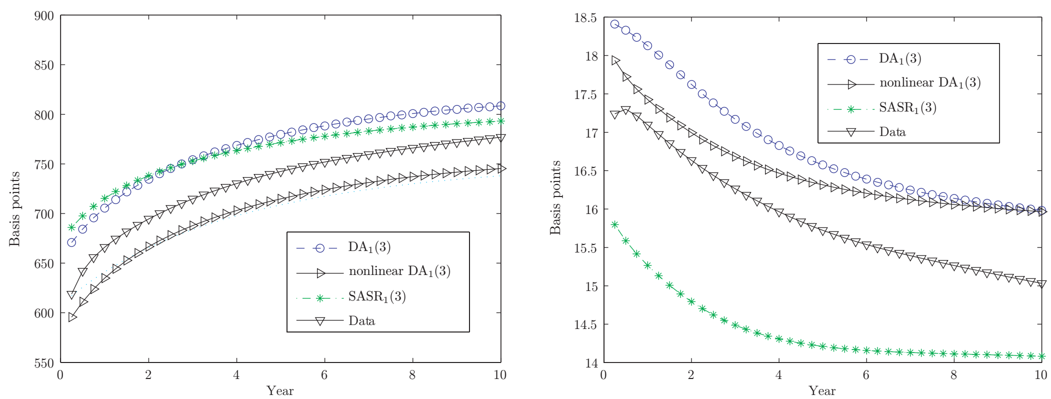


Figure 1: Intercepts (δ_0^n) and loadings (δ_Y^n) of states on bond yields of different maturities for the nonlinear $DA_1(3)$ model (top) and Duarte's model (bottom).

The implications of these different "rotations" of the factors induced by the different forms of nonlinearity for the terms structures of means and variances of bond yields can be seen from Figure 2. For each model, 100,000 months of data on X were simulated, beyond 100,000 months of burn-in to avoid sensitivity to initial conditions, and then sample means and volatilities (standard deviations) of the yields were computed at the ML estimates of the parameters. All three models match the unconditional means of the yields reasonably well. The nonlinear $DA_1(3)$ model matches somewhat better at short maturities, and the $SASR_1(3)$ model matches better at the longer maturities, but overall the fits are comparable. On the other hand, there are substantial differences in the models' fits to the term structures of volatilities. Most striking are the low volatilities implied by the $SASR_1(3)$ model, relative to those in the historical data, particularly at the short- to intermediate-term maturities. Both the linear and nonlinear $DA_1(3)$ models tend to overstate volatility at the longer end of the curve, but they match up quite well with the data at shorter maturities, with the nonlinear $DA_1(3)$ model having the closest match.



(a) Term Structure of Means

(b) Term Structure of Volatilities

Figure 2: Term structures of unconditional means and volatilities of zero-coupon bond yields implied by the linear and nonlinear $DA_1(3)$ models.

Turning next to the properties of the conditional means, the nonlinearity implied by the $SASR_1(3)$ model is quite different from that implied by the nonlinear $DA_1(3)$ model (Figure 3). Due to the quadratic shape of the mean of Z induced by the presence of $\sqrt{Z}$ in the drift, model $SASR_1(3)$ exhibits strong mean reversion for large values of Z, but essentially no mean reversion when Z is near or below its mean. This inability of model $SASR_1(3)$ to generate mean reversion at small values of Z is an inherent feature of the assumed structure of the drift of Z. Most of the descriptive evidence on bond yields suggests that they are mean reverting both at high and low values relative to their means (e.g., Ait-Sahalia (1996) and Ang and Bekaert (2002)). We follow up on this point subsequently by examining how well this model performs out-of-sample in predicting excess returns.

The drift in the nonlinear $DA_1(3)$ model resembles more closely the nonlinear shapes documented in previous descriptive studies of nonlinearity in short-term interest rates. The degree of mean reversion in Z at low values is mild compared to high values, but the characteristic "S on its side" shape is present. The left-hand-side of this graph turns up more sharply for values of Z even further below its mean, though the likelihood of observing a value of Z in this region is very small.

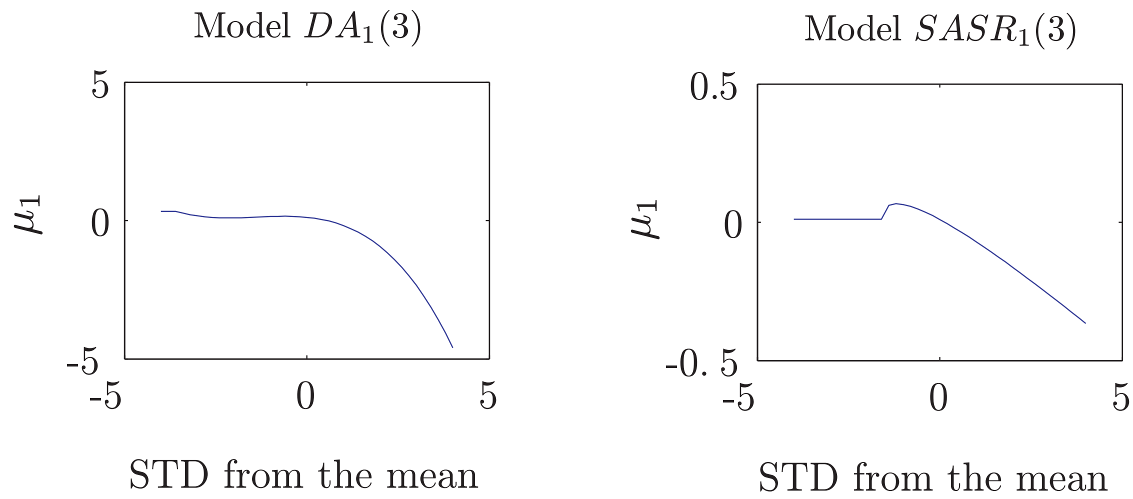


Figure 3: Conditional $\mathbb{P}$ means of Z implied by the nonlinear $DA_1(3)$ model (left) and the $SASR_1(3)$ model (right).

One means of comparing the relative fits of *DTSM*s is to compare their forecasting powers both within and out of sample. Figure 4 displays the root-mean-squared forecast errors (*RMFE*), within sample, for the three *DTSM*s over forecast horizons of eight and ten months. For comparison we have also included the *RMFE*s generated by the assumption that the zero-coupon bond yields follow random walks (*RD*). All three models perform comparably based on this metric, with our nonlinear $DA_1(3)$ somewhat outperforming the other two models. Mapping the loadings in Figure 1 to the forecasting performance in Figure 4 for each maturity, we find that the nonlinear $DA_1(3)$ tends to forecast better than its linear counterpart as the contribution of the linear slope factor decreases in importance. It is also interesting that our linear $DA_1(3)$ model forecasts slightly better than the $SASR_1(3)$ model.

When the corresponding *RMFE*s are computed using out-of-sample data for the period $1996-2000$ the results are somewhat different. The nonlinear $DA_1(3)$ model is still outperforms its linear counter part, producing estimates that are both less biased and more precise. However, there is little difference between the two models with nonlinear market price of risk specifications.

6 Concluding Remarks

In this paper we have argued that, along important dimensions, researchers can gain flexibility and tractability in analyzing *DTSM*s by switching from continuous to discrete time. We have developed a family of nonlinear *DTSM*s that has several key properties: (i) under $\mathbb{Q}$, the risk factors X follow the discrete-time counterpart of an affine process residing in one of the families $A_M^{\mathbb{Q}}(N)$, as classified by Dai and Singleton (2000), (ii) the pricing kernel is specified so as to give the modeler nearly complete flexibility in specifying the market price

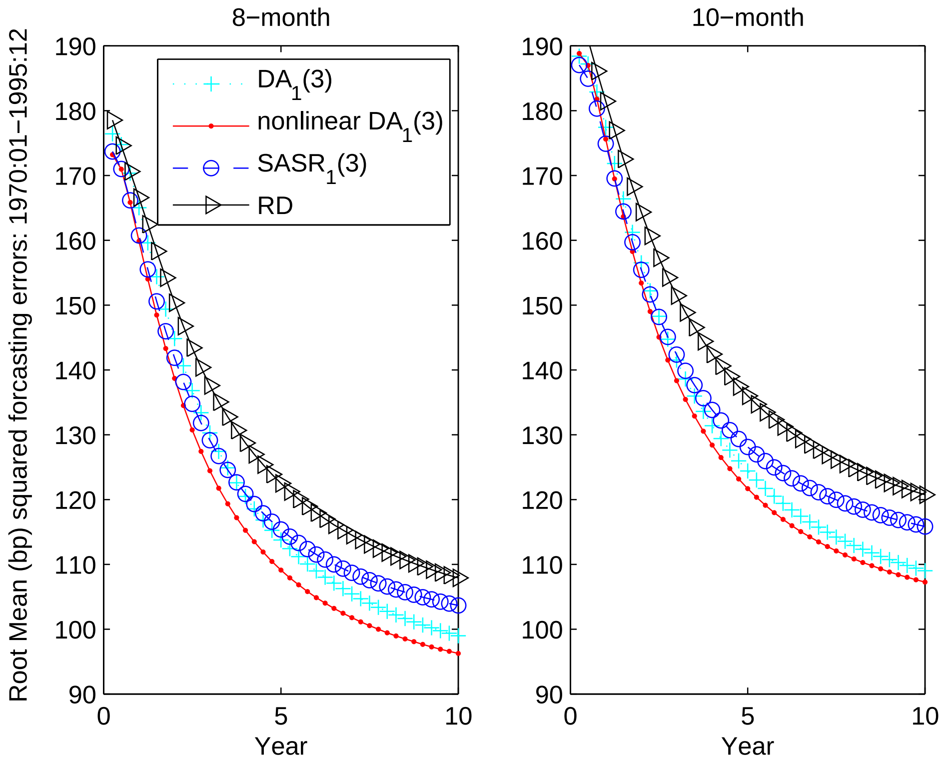


Figure 4: Root mean squared errors in sample (1970:1-1995:12) for a linear $DA_1(3)$, a nonlinear $DA_1(3)$ model with a quare and a cubic term in the $\mathbb{P}$-drifts, a nonlinear $SASR_1(3)$ model and a random walk model (RD).

of risk Λ_t of the risk factors, and (iii) for any admissible specification of Λ_t, the likelihood function of the bond yield data is known in closed form. This modeling framework was illustrated by estimating nonlinear $(DA_1^{\mathbb{Q}}(3), \Lambda)$ models with several specifications of Λ_t that give rise to nonlinear (and non-affine) representations of X under the historical measure $\mathbb{P}$. Our particular choices of Λ_t introduced powers of the volatility factor Z. However, our modeling framework allows, in principle, for a fully semi-parametric specification of Λ_t, and for possible nonlinearity in all three state variables in these $(DA_1^{\mathbb{Q}}(3), \Lambda)$ models, and not just in the volatility factor.

There are many directions in which our modeling framework can be extended. For instance, given the widespread interest in regime-switching models for interest rates, the extension to pricing models that allow for volatility processes to switch regimes may well be of interest. Ang and Bekaert (2003) and Dai, Singleton, and Yang (2005) study *DTSM*s in which X follows a regime-switching $DA_0^{\mathbb{Q}}(N)$ process, with the latter study allowing for priced regime-shift risk. Bansal and Zhou (2002) examine an approximate *DTSM* in which the risk factors follow independent processes, each of which is interpreted as a discrete-time approximation to an $A_1^{\mathbb{Q}}(1)$ process. Our framework allows for the introduction of

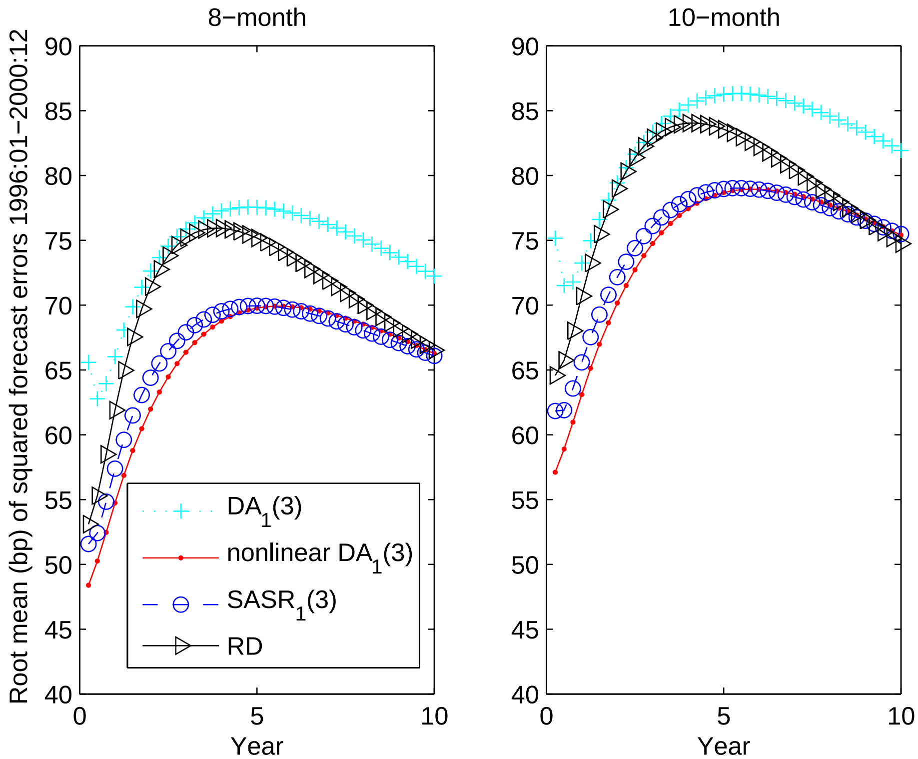


Figure 5: Root mean squared errors out of sample (1996:1-2000:12) for a linear $DA_1(3)$, a nonlinear $DA_1(3)$ model with a quare and a cubic term in the $\mathbb{P}$-drifts, a nonlinear $SASR_1(3)$ model and a random walk model (RD).

stochastic volatility into those studies that have focused on Gaussian models, and allows the stochastic volatility factors to be treated consistently as $DA_M^{\mathbb{Q}}(M)$ processes rather than as approximations to their continuous-time counterparts.

Moreover, under certain conditions analogous to those set forth in Dai and Singleton (2003) for continuous-time models, we preserve analytical bond pricing even in the presence of switching regimes. To be concrete, for the case of a regime-switching $(DA_1^{\mathbb{Q}}(N), \Lambda)$ model, what we need for analytical bond pricing is that $\kappa_Z^{\mathbb{Q}}$ and σ_Z do not switch across regimes. However, $\nu = 2\kappa_Z^{\mathbb{Q}}\theta_Z^{\mathbb{Q}}/\sigma_Z^2$ can change with the regime. This is analogous to the Gaussian case where we can allow the "constant term" to switch across regimes. These restrictions do not preclude priced regime shift risk or that the $\mathbb{P}$ distributions of the risk factors and the bond yields have a regime-switching structure. To implement a model like this, general assumption, we can specify Z_t as an autoregressive gamma process under $\mathbb{Q}$. Depending on how one parameterizes the market prices of risks, the conditional density of Z_t under $\mathbb{P}$, which is needed to construct the likelihood function, may not be known in closed form. However, at least for the case of $M = 1$, the relevant density can be computed through Fourier inversion (see, e.g., Duffie, Pan, and Singleton (2000) and Singleton (2001)).

Appendix

A Proof of Proposition G.E.(Z)

The proof follows from a lemma due to Mokkadem (1985)

Lemma 1 *(Mokkadem) Suppose $\{Z_t\}$ is an aperiodic and irreducible Markov chain defined by*

$$Z_{t+1} = H(Z_t, \epsilon_{t+1}, \theta), \tag{36}$$

where ϵ_t is an i.i.d. *process. Fix θ and suppose there are constants $K > 0, \delta_\theta \in (0,1)$, and $q > 0$ such that $H(\cdot, \epsilon_1, \theta)$ is well defined and continuous with*

$$\|H(z, \epsilon_1, \theta)\|_q < \delta_\theta \|z\|, \quad \|z\| > K. \tag{37}$$

Then $\{Z_t\}$ is geometrically ergodic.

In our setting, we can write, without loss of generality,

$$H(z, \epsilon_1, \theta) = \left[a^{(1)}(\lambda(z)) + b^{(1)}(\lambda(z))z\right] + \sqrt{\Omega(z)}\epsilon_1,$$

where ϵ_1 has a zero mean and unit variance, and $\Omega(z) = a^{(2)}(\lambda(z)) + b^{(2)}(\lambda(z))z$. Take $q = 2$, we have

$$\frac{\|H(z, \epsilon_1, \theta)\|_2}{\|z\|} \leq \frac{\|a^{(1)}(\lambda(z))\|}{\|z\|} + \frac{\|b^{(1)}(\lambda(z))z\|}{\|z\|} + \frac{\|\sqrt{\Omega(z)}\epsilon_1\|_2}{\|z\|}. \tag{38}$$

The first term on the right-hand-side of (38) satisfies

$$\frac{\|a^{(1)}(\lambda(z))\|}{\|z\|} = \frac{\|\mathrm{vec}\left[\frac{\nu_i c_i}{1-\lambda_i(z)c_i}\right]\|}{\|z\|} \leq \frac{\|\mathrm{vec}\left[\nu_i c_i\right]\|}{\|z\|} \to 0, \ \|z\| \to \infty, \tag{39}$$

where we have used the assumption (i) to obtain the inequality.

Since all elements of ρ are non-negative, if $1 - \lambda_i(z)c_i \geq 1$ for all z and i, then the second term in (38) is bounded by

$$\frac{\|b^{(1)}(\lambda(z))z\|}{\|z\|} \leq \frac{\|\rho z\|}{\|z\|} \leq \max_i |\psi_i|.$$

If, in addition, $\rho_{ij} = 0$ for $i \neq j$, the above bound is valid for each element of z when it is sufficiently large. That is, there exists a $K > 0$, such that

$$\frac{\|{b^{(1)}}_{ii}(\lambda(z))z_i\|}{\|z_i\|} \leq \frac{\|\rho_{ii} z_i\|}{\|z_i\|} \leq \rho_{ii} \leq \max_i \psi_i, \ z_i > K$$

Finally, the last term in (38) can be made arbitrarily small by choice of a sufficiently large K, because $\|\epsilon_1\|_2 = 1$ and $\sqrt{\Omega(z)}$ depends on z through terms of the form $\sqrt{z}$.[17]

The only term on the right-hand side of (38) that does not become arbitrarily small as K increases towards infinity is the second term. Since we assume that $\max_i |\psi_i| < 1$, we are free to choose δ_θ to satisfy $\max_i |\psi_i| < \delta_\theta < 1$ so that Lemma 1 is satisfied.

References

Ahn, D., Dittmar, R., Gallant, A., 2002. Quadratic term structure models: Theory and evidence. Review of Financial Studies 15, 243–288.

Ait-Sahalia, Y., 1996. Testing continuous-time models of the spot interest rate. Review of Financial Studies 9(2), 385 – 426.

Ait-Sahalia, Y., 1999. Transition densities for interest rate and other nonlinear diffusions. Journal of Finance 54, 1361–1395.

——— , 2002. Maximum-likelihood estimation of discretely-sampled diffusions: A closed-form approximation approach. Econometrica 2, 223–262.

Ang, A., Bekaert, G., 2002. Short rate nonlinearities and regime switches. Journal of Economic Dynamics and Control 26, 1243–1274.

——— , 2003. The term structure of real rates and expected inflation. Working paper, Columbia University.

Ang, A., Dong, S., Piazzesi, M., 2005. No-arbitrage taylor rules. Unpublished working paper. Columbia University.

Ang, A., Piazzesi, M., 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. Journal of Monetary Economics 50, 745–787.

Backus, D., Foresi, S., Zin, S., 1998. Arbitrage opportunities in arbitrage-free models of bond pricing. Journal of Business and Economic Statistics 16, 13–26.

Bansal, R., Zhou, H., 2002. Term structure of interest rates with regime shifts. Journal of Finance 57, 1997 – 2043.

Beaglehole, D. R., Tenney, M. S., 1991. General solutions of some interest rate-contingent claim pricing equations. Journal of Fixed Income September, 69–83.

Bekaert, G., Engstrom, E., Grenadier, S., 2004. Stock and bond returns with moody investors. Unpublished working paper. Stanford University.

[17] See Duffie and Singleton (1993) for a discussion of the geometric ergodicity of models in which volatility depends on terms of the form x^γ, for $\gamma < 1$. By using L^2 norm ($q = 2$), we can apply Mokkadem's lemma without the *i.i.d.* assumption for the state innovations.

Boudoukh, J., Richardson, M., Smith, T., Whitelaw, R. F., 1999. Regime shifts and bond returns. Working paper, New York University.

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

Brandt, M., Santa-Clara, P., 2001. Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. Working paper, Wharton School.

Chan, K. C., Karolyi, G. A., Longstaff, F. A., Sanders, A. B., 1992. An empirical comparison of alternative models of the short-term interest rate. Journal of Finance XLVII(3), 1209–1227.

Cheridito, R., Filipovic, D., Kimmel, R., 2003. Market price of risk in affine models: Theory and evidence. Working Paper, Princeton University.

Collin-Dufresne, P., Goldstein, R., Jones, C., 2004. Can interest rate volatility be extracted from the cross section of bond yields? an investigation of unspanned stochastic volatility. Unpublished working paper. NBER 10756.

Cox, J., Ingersoll, J., Ross, S., 1985. A theory of the term structure of interest rates. Econometrica 53, 385–407.

Dai, Q., Phillipon, T., 2004. Government deficits and interest rates: a no-arbitrage structural var approach. Unpublished working paper. NYU and UNC.

Dai, Q., Singleton, K., 2000. Specification analysis of affine term structure models. Journal of Finance 55, 1943–1978.

——— , 2002. Expectations puzzles, time-varying risk premia, and affine models of the term structure. Journal of Financial Economics 63, 415–441.

——— , 2003. Term structure dynamics in theory and reality. Review of Financial Studies 16, 631–678.

Dai, Q., Singleton, K., Yang, W., 2005. Regime shifts in a dynamic term structure model of u.s. treasury bond yields. Working Paper, Stanford University.

Darolles, S., Gourieroux, C., Jasiak, J., 2001. Compound autoregressive processes. Unpublished working paper. CREST.

Duan, J., Simonato, J., 1999. Estimating and testing exponential-affine term structure models by kalman filter. Review of Quantitative Finance and Accounting 13, 111–135.

Duarte, J., 2004. Evaluating an alternative risk preference in affine term structure models. Review of Financial Studies 17, 379–404.

Duffee, G. R., 2002. Term premia and interest rate forecasts in affine models. Journal of Finance 57, 405–443.

Duffie, D., Filipovic, D., Schachermayer, W., 2003. Affine processes and applications in finance. Annals of Applied Probability 13, 984–1053.

Duffie, D., Kan, R., 1996. A yield-factor model of interest rates. Mathematical Finance 6, 379–406.

Duffie, D., Pan, J., Singleton, K., 2000. Transform analysis and asset pricing for affine jump-diffusions. Econometrica 68, 1343–1376.

Duffie, D., Pedersen, L., Singleton, K., 2003. Modeling credit spreads on sovereign debt: A case study of russian bonds. Journal of Finance 55, 119–159.

Duffie, D., Singleton, K., 1993. Simulated moments estimation of markov models of asset prices. Econometrica 61, 929–952.

Evans, M., 2000. Regime shifts, risk and the term structure. Working Paper, Georgetown University.

Gallant, A. R., Tauchen, G., 1996. Which moments to match?. Econometric Theory 12, 657–681.

Gourieroux, C., Jasiak, J., 2001. Autoregressive gamma processes. Unpublished working paper. CREST.

Gourieroux, C., Monfort, A., Polimenis, V., 2002. Affine term structure models. Working paper, University of Toronto, Canada.

Gray, S., 1996. Modeling the conditional distribution of interest rates as a regime switching process. Journal of Financial Economics 42, 27–62.

Hordahl, P., Tristani, O., Vestin, D., 2003. A joint econometric model of macroeconomic and term structure dynamics. Working Paper, European Central Bank.

Lamberton, D., Lapeyre, B., 1992. Introduction au calcul stochastique applique a la finance. Unpublished working paper. Mathematique et Applications, Ellipses-Editions.

Leippold, M., Wu, L., 2002. Asset pricing under the quadratic class. Journal of Financial and Quantitative Analysis 37, 271–295.

Litterman, R., Scheinkman, J., 1991. Common factors affecting bond returns. Journal of Fixed Income 1, 54–61.

Litterman, R., Scheinkman, J., Weiss, L., 1991. Volatility and the yield curve. Journal of Fixed Income 1, 49–53.

Mokkadem, A., 1985. Le modele non lineaire ar(1) general. ergodicite et ergodicite geometrique. Comptes Rendues Academie Scientifique Paris 301, Serie I, 889–892.

Naik, V., Lee, M. H., 1997. Yield curve dynamics with discrete shifts in economic regimes: Theory and estimation. Working paper, Faculty of Commerce, University of British Columbia.

Nummelin, E., Tuominen, P., 1982. Geometric ergodicity of harris recurrent markov chains with applications to renewel theory. Stochastic Processes and Their Applications 12, 187–202.

Pedersen, A., 1995. A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. Scand J Statistics 22, 55–71.

Rudebusch, G., Wu, T., 2003. A no-arbitrage model of the term structure and the macroeconomy. Working Paper, Federal Reserve Bank of San Francisco.

Singleton, K., 2001. Estimation of affine asset pricing models using the empirical characteristic function. Journal of Econometrics 102, 111–141.

Stanton, R., 1997. A nonparametric model of term structure dynamics and the market price of interest rate risk. Journal of Finance 52, 1973–2002.

Sun, T., 1992. Real and nominal interest rates: A discrete-time model and its continuous-time limit. Review of Financial Studies 5(4), 581–611.

Tweedie, R., 1982. Criteria for rates of convergence of markov chains, with applications to queuing and storage theory. in *Probability, Statistics, and Analysis*, ed. by J. Kingman, and G. Reuter. Cambridge: Cambridge University Press.

Vasicek, O., 1977. An equilibrium characterization of the term structure. Journal of Financial Economics 5, 177–188.