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

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

This draft: January 11, 2008

[1]Dai is with Capula Investment Management LLP. Le is with the University of North Carolina at Chapel Hill, anh_le@unc.edu. 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.

Abstract

This paper develops a rich class of discrete-time, *nonlinear* dynamic term structure models ($DTSM$s). Under the risk-neutral measure, the distribution of the state vector X_t resides within a family of discrete-time affine 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). Moreover, we allow the market price of risk Λ_t, linking the risk-neutral and historical distributions of X, to depend generally on the state X_t. The conditional likelihood functions for coupon bond yields for the resulting nonlinear models under the historical measure are known exactly in closed form. As an illustration of our approach, we estimate a three factor model with a cubic term in the drift of the stochastic volatility factor and compare it to a model with a linear drift. Our results show that inclusion of a cubic term in the drift significantly improves the models statistical fit as well as its out-of-sample forecasting performance.

1 Introduction

This paper develops a rich class of discrete-time, *nonlinear* dynamic term structure models ($DTSM$s) in which zero-coupon bond yields and their conditional densities are known exactly in closed form.[1] Under the risk-neutral measure, the distribution of the state vector X_t resides within a family of discrete-time affine 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).[2] Moreover, we allow the market price of risk Λ_t, linking the risk-neutral ($\mathbb{Q}$) and historical ($\mathbb{P}$) distributions of X, to depend generally 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 flexibility in specifying Λ_t leads to a family of $DTSM$s in which the conditional $\mathbb{P}$-means of X_{t+1} and bond yields can show very rich *nonlinear* dependence on X_t. In this manner we substantially extend the *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. The goodness-of-fits of $DTSM$s depend critically on the specification of the market price of risk (see, e.g., Duffee (2002), Dai and Singleton (2002), Duarte (2003), and Ahn, Dittmar, and Gallant (2001)). However, the functional forms of Λ_t in these studies are quite restrictive – reflecting a trade-off in continuous-time formulations of $DTSM$s between generality in pricing and tractability of estimation. By allowing the researcher almost complete freedom in specifying the dependence of Λ_t on the state vector, we facilitate empirical investigation of much richer specifications of risk premiums than have heretofore been examined empirically.

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 neo-Keynesian macroeconomic models[4] has focused exclusively on discrete-time Gaussian $DTSM$s. Arbitrage-free $DTSM$s are overlaid onto log-linear macro models with Gaussian, homoskedastic shocks and constant inflation risk premiums. Our formulation allows for the immediate extension of such models to the case of time-varying macroeconomic risks. Similarly, the reduced-form, Gaussian $DTSM$s incorporating fiscal and monetary policy

[1] 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 (2003), all in a multi-factor setting.

[2] Our analysis extends immediately to the case of quadratic-Gaussian models discussed in Beaglehole and Tenney (1991), Ahn, Dittmar, and Gallant (2001) and Leippold and Wu (2002). This can be seen from the work of Cheng and Scaillet (2002) who show that quadratic-Gaussian models can be reinterpreted as affine models, after an appropriate expansion of the state vector.

[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 (2002) and Gourieroux, Monfort, and Polimenis (2002).

[4] See, for example, Rudebusch and Wu (2003), Hordahl, Tristani, and Vestin (2003), Wu (2000), and Bekaert, Cho, and Moreno (2005).

rules of Dai and Phillipon (2004) and Ang, Dong, and Piazzesi (2005) are easily extended to allow for stochastic volatility and richer formulations of the market prices of risk.

With regard to estimation of $DTSM$s, even when the state vector follows a continuous-time, 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.[5] By shifting to discrete time, we obtain exact representations of the likelihood functions of bond yields even for our most flexible nonlinear models. In particular, we have known likelihood functions for the (discrete-time counterparts to the) entire class of affine $DTSM$s classified by DS. Therefore, no approximations are necessary in estimation.

The construction 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, 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 $DA_M^{\mathbb{Q}}(N)$, 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– the $A_M^{\mathbb{Q}}(N)$ models– examined in Duffie and Kan (1996) and DS.[6]

Second, for each family $DA_M^{\mathbb{Q}}(N)$, we specify an associated family of state-price densities $(d\mathbb{P}/d\mathbb{Q})_{t+1}^D$ linking the $\mathbb{P}$ and $\mathbb{Q}$ distributions of X_{t+1} that has a natural interpretation as a discrete-time counterpart to the state-price density associated with affine diffusion-based, continuous-time $DTSM$s. Moreover, just as in a continuous-time model, we allow the modeler substantial flexibility in specifying the dependence of the market price of factor risks, Λ_t, on X_t. By roaming over admissible choices of Λ_t, we are effectively ranging across the entire

[5] 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).

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

family of admissible arbitrage-free *DTSM*s constructed under the assumption that, under $\mathbb{Q}$, X follows a discrete-time affine process residing in one of the families $A_M^{\mathbb{Q}}(N)$.

Importantly, a key difference between our discrete-time construction and the continuous-time counterpart is that each choice of $(d\mathbb{P}/d\mathbb{Q})_{t+1}^D$, when combined with a known $\mathbb{Q}$-affine distribution of the state X, 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. The only restrictions we impose on the choice of $\Lambda(X_t)$, beyond requiring that the model not admit arbitrage opportunities, is that it 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. 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 an *unknown* $\mathbb{P}$ 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 a nonlinear $(DA_1^{\mathbb{Q}}(3), \Lambda)$ model in which the $\mathbb{P}$-conditional mean of the volatility factor Z_{t+1} is nonlinear in Z_t. The properties of this model are compared to those of its nested linear counterpart along several dimensions, including their within and out-of sample forecasting powers for bond yields. Additionally, we also report results for a discrete-time counterpart to Duarte (2003)'s $SASR_1(3)$ model in which the square-root of the volatility factor appears in its own drift under $\mathbb{P}$. This is another interesting application of our framework in that the likelihood function for our discrete-time model is known in closed form, whereas Duarte had to resort to various approximations in his estimation strategy.

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 on 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 counterparts do not reside in one of the families $A_M^{\mathbb{Q}}(N)$). Moreover, Ang and Piazzesi (2002) 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.

In our concluding Section 7 we address the potentially important extension of our modeling framework to the case of multiple regimes. Numerous studies have presented descriptive evidence supporting multiple regimes in interest rates (e.g., Gray (1996) and Ang and Bekaert (2002)), and 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 Wei (2003), among others, have introduced regime switching into *DTSM*s. We briefly comment on how our modeling framework for nonlinear $(DA_M^{\mathbb{Q}}(N), \Lambda)$ models can be extended to the case of multiple regimes in the presence of stochastic volatility.

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:[7] 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

[7] See Duffie, Pan, and Singleton (2000) for a proof that the 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.

$\kappa^{\mathbb{Q}}$ and σ are $N \times N$ matrices and $\theta^{\mathbb{Q}}$ is a $N \times 1$ vector. Then in the limit $\Delta t \to 0$, 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 (2002), Dai, Singleton, and Wei (2003), 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 of 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:

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

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^i_{t+1} takes the form:

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

Using conditional independence, the distribution of a $DA^{\mathbb{Q}}_N(N)$ process Z_{t+1}, conditional on Z_t, is given by $f^{\mathbb{Q}}(Z_{t+1}|Z_t) = \prod_{i=1}^N f^{\mathbb{Q}}(Z^i_{t+1}|Z_t)$. 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

$$a(u) = -\sum_{i=1}^N \nu_i \log(1-u_i c_i), \quad b(u) = \sum_{i=1}^N \frac{u_i}{1-u_i c_i}\rho_i. \tag{5}$$

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^i_t, 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 $A^{\mathbb{Q}}_N(N)$. 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 ensure 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 ensures that we have a well-defined multivariate discrete-time process taking on strictly positive values. The conditional mean $E^{\mathbb{Q}}_t[Z_{t+1}]$ and conditional covariance matrix $V^{\mathbb{Q}}_t[Z_{t+1}]$ implied by the conditional moment-generating function (1) and (5) are

$$E^{\mathbb{Q}}_t[Z_{t+1}] = a^{(1)}(0) + \sum_{i=1}^n b^{(1)}_i(0)Z^i_t, \; V^{\mathbb{Q}}_t[Z_{t+1}] = a^{(2)}(0) + diag\left[\partial^2 b/\partial u^2_i(0)Z_t\right], \tag{6}$$

where $a^{(k)}(u)$ denotes the k^{th} derivative of $a(u)$ with respect to u, $b^{(1)}_i$ is the first derivative of b_i with respect to u_i, and $diag[\cdot]$ denotes the diagonal matrix generated by the elements in brackets. Specifically,

$$E^{\mathbb{Q}}_t[Z_{t+1}](i) = \nu_i c_i + \rho_i Z_t, \; V^{\mathbb{Q}}_t[Z_{t+1}](i,i) = \nu_i c^2_i + 2c_i\rho_i Z_t, \tag{7}$$

and the off-diagonal elements of $V^{\mathbb{Q}}_t[Z_{t+1}]$ 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 $A^{\mathbb{Q}}_N(N)$ process[8] can be seen by letting $\rho = I_{N\times N} - \kappa^{\mathbb{Q}}\Delta t$, $c_i = \frac{\sigma^2_i}{2}\Delta t$, and $\nu_i = \frac{2(\kappa^{\mathbb{Q}}\theta^{\mathbb{Q}})_i}{\sigma^2_i}$, 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 $DA^{\mathbb{Q}}_N(N)$ process converges to:

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

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

[8]Gourieroux and Jasiak (2001) attribute the insight that the $DA^{\mathbb{Q}}_1(1)$ 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}[9] and normally distributed with conditional mean and variance

$$\omega_{Yt}^{\mathbb{Q}} \equiv \mu_0 + \mu_Y X_t \text{ and } \Omega_{Yt} \equiv \Sigma_Y S_{Yt} \Sigma_Y', \tag{8}$$

where μ_0 is a $(N-M) \times 1$ vector, $\mu_Y \equiv \begin{pmatrix} \mu_Y^Z & \mu_Y^Y \end{pmatrix}$ 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

$$f^{\mathbb{Q}}(X_{t+1}|X_t) = f^{\mathbb{Q}}(Y_{t+1}|Y_t, Z_t) \times f^{\mathbb{Q}}(Z_{t+1}|Z_t), \tag{9}$$

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

$$a(u) = -\sum_{i=1}^M \nu_i \log(1 - u_{Z,i} c_i) + \mu_0' u_Y + \frac{1}{2} u_Y' h_0 u_Y, \tag{10}$$

$$b(u) = \left[\sum_{i=1}^M \frac{u_{Z,i}}{1 - u_{Z,i} c_i} \rho_i + \left(\frac{1}{2} u_Y' h_i u_Y \right)_{i=1,2,\ldots,M} + u_Y' \mu_Y^Z \quad u_Y' \mu_Y^Y \right], \tag{11}$$

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), (10), and (11).*

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:

[9] 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.*[10]

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

$$D_t^n = E_t^{\mathbb{Q}}\left[e^{-\sum_{i=0}^{n-1} r_{t+i}}\right] = e^{-r_t} E_t^{\mathbb{Q}}\left[D_{t+1}^{n-1}\right] = e^{-A_n - B_n X_t}, \tag{12}$$

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

$$A_n - A_{n-1} = \delta_0 + A_{n-1} - a(-B_{n-1}), \tag{13}$$

$$B_n = \delta_X - b(-B_{n-1}), \tag{14}$$

with the initial condition $A_0 = B_0 = 0$.[11]

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 (12) 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

$$(d\mathbb{Q}/d\mathbb{P})_{t,t+1}^C = \frac{e^{-\frac{1}{2}\int_t^{t+1}\eta(s)'\eta(s)ds - \int_t^{t+1}\eta(s)'dB^{\mathbb{P}}(s)}}{E_t^{\mathbb{P}}\left[e^{-\frac{1}{2}\int_t^{t+1}\eta(s)'\eta(s)ds - \int_t^{t+1}\eta(s)'dB^{\mathbb{P}}(s)}\right]} \tag{15}$$

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.

[10] 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 11 for further elaboration on the second point.

[11] 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

$$f^{\mathbb{P}}(X_{t+1}|X_t) = (d\mathbb{P}/d\mathbb{Q})^D(X_{t+1}; \Lambda_t) \times f^{\mathbb{Q}}(X_{t+1}|X_t), \tag{16}$$

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 (16) with*

$$\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)^D (X_{t+1}; \Lambda_t) = \frac{e^{\Lambda_t' X_{t+1}}}{\phi^{\mathbb{Q}}(\Lambda_t; X_t)}, \tag{17}$$

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 $1 \leq \forall i \leq M$, and $Prob\{\Lambda_t^i < \infty\} = 1$, for $M + 1 \leq i \leq N$.

This formulation of $(d\mathbb{P}/d\mathbb{Q})^D$ is a conditional version of the Esscher (1932) transform for the conditional $\mathbb{Q}$ distribution of X.[12] Under Assumption $N(\mathbb{P})$, the conditional $\mathbb{P}$-Laplace transform of X_t is given by

$$\phi^{\mathbb{P}}(u; X_t) \;=\; \frac{\phi^{\mathbb{Q}}(u + \Lambda_t; X_t)}{\phi^{\mathbb{Q}}(\Lambda_t; X_t)} = e^{\mathcal{A}(u;\Lambda_t) + \mathcal{B}(u;\Lambda_t) X_t}, \tag{18}$$

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

$$\mathcal{M}_{t,t+1} = e^{-r_t} \times \frac{e^{-\Lambda_t' X_{t+1}}}{\phi^{\mathbb{P}}(-\Lambda_t; X_t)}, \tag{19}$$

where we have used the fact that $\phi^{\mathbb{P}}(-\Lambda_t; X_t) = \left[\phi^{\mathbb{Q}}(\Lambda_t; X_t)\right]^{-1}$, which follows from (18) evaluated at $u = -\Lambda_t$.[13]

[12] Buhlmann, Delbaen, Embrechts, and Shiryaev (1996) formally develop the conditional Essher transform using martingale theory in the context of no-arbitrage pricing. A notable application of the Esscher transform (with constant Λ) to option pricing is Gerber and Shiu (1994) who demonstrate that many variants of the Black-Scholes option pricing model can be developed using the Esscher transform. For our purposes, the conditional transform is essential, because of our linkage (see below) of Λ_t to the market prices of risk.

[13] Note that, though $\phi^{\mathbb{P}}(u; X_t)$ has an exponential-affine form, $\mathcal{A}(u; \Lambda_t)$ and $\mathcal{B}(u; \Lambda_t)$ are functions of Λ_t which, in turn, may be a nonlinear function of X_t. Thus, in general X is not an affine process under $\mathbb{P}$.

To motivate this choice of Radon-Nykodym derivative– equivalently pricing kernel $\mathcal{M}$– consider again the continuous-time formulation in (15). For a small time interval Δ, and approximate affine state process $X_{t+\Delta} \approx \mu_X^{\mathbb{P}}(X_t)\Delta + \Sigma_X\sqrt{S_{Xt}}\epsilon_{t+\Delta}^{\mathbb{P}}$, with $\epsilon_{t+\Delta}|X_t \sim N(0, \Delta I)$,

$$\begin{aligned}(d\mathbb{Q}/d\mathbb{P})_{t,t+\Delta}^{C} &\approx \frac{e^{-\frac{1}{2}\eta_t'\eta_t\Delta - \eta_t'\epsilon_{t+\Delta}^{\mathbb{P}}}}{E_t^{\mathbb{P}}\left[e^{-\frac{1}{2}\eta_t'\eta_t\Delta - \eta_t'\epsilon_{t+\Delta}^{\mathbb{P}}}\right]} = \frac{e^{-\Lambda_t'\Sigma_X\sqrt{S_{Xt}}\epsilon_{t+\Delta}^{\mathbb{P}}}}{E_t^{\mathbb{P}}\left[e^{-\Lambda_t'\Sigma_X\sqrt{S_{Xt}}\epsilon_{t+\Delta}^{\mathbb{P}}}\right]} \\ &= \frac{e^{-\Lambda_t'X_{t+\Delta}}}{E_t^{\mathbb{P}}\left[e^{-\Lambda_t'X_{t+\Delta}}\right]} = \frac{e^{-\Lambda_t'X_{t+\Delta}}}{\phi^{\mathbb{P}}(-\Lambda_t; X_t)},\end{aligned} \tag{20}$$

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

$$\mathcal{M}_{t,t+\Delta} \equiv e^{-r_t} \times \frac{f^{\mathbb{Q}}(X_{t+\Delta}|X_t)}{f^{\mathbb{P}}(X_{t+\Delta}|X_t)} \approx e^{-r_t\Delta}\frac{e^{-\Lambda_t'X_{t+\Delta}}}{\phi^{\mathbb{P}}(-\Lambda_t; X_t)}. \tag{21}$$

This kernel takes exactly the same form as (19).

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 (19).[14] Subject to this "matching condition," no approximations are involved in deriving either $f^{\mathbb{P}}(X_{t+1}|X_t)$ in (16) or the associated pricing kernel $\mathcal{M}_{t,t+1}$ in (19).

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,

$$\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)_{t,t+\Delta}^{D} \approx \left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)_{t,t+\Delta}^{C}. \tag{22}$$

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 ensure 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 $u_ic_i < 1$. 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, $f^{\mathbb{P}}$

[14]The challenges that arise when the sampling and modeling intervals do not coincide are discussed in more depth in Section 5.

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.[15]

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 bond yields.[16] Furthermore, the relationship between the observed yields y_t and the state vector X_t are also known due to the pricing equation (12), 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.

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.

Notice first of all that[17]

$$\begin{aligned} E_t^{\mathbb{P}}[X_{t+1}] - E_t^{\mathbb{Q}}[X_{t+1}] &= \left[\mathcal{A}^{(1)}(0;\Lambda_t) - a^{(1)}(0)\right] + \left[\mathcal{B}^{(1)}(0;\Lambda_t) - b^{(1)}(0)\right] X_t \\ &= V_t^{\mathbb{P}}[X_{t+1}] \times \Lambda_t + o(\Lambda_t), \end{aligned} \tag{23}$$

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

$$\begin{aligned} E_t^{\mathbb{P}}\left[\log \frac{e^{-c'X_{t+1}}}{E_t^{\mathbb{Q}}\left[e^{-r_t} e^{-c'X_{t+1}}\right]}\right] - r_t &= -\left[a(-c) + c'a^{(1)}(\Lambda_t)\right] - \left[b(-c) + c'b^{(1)}(\Lambda_t)\right] X_t, \\ &= -c' V_t^{\mathbb{P}}[X_{t+1}] \times \Lambda_t + o(c) + o(\Lambda_t). \end{aligned} \tag{24}$$

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

[16] 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 (13) and (14) as (possibly nonlinear) cross-equation restrictions.

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

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_t 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

$$\Lambda_t = \left(\Sigma_X \sqrt{S_{Xt}}\right)'^{-1} \eta_t. \tag{25}$$

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 (25) 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 (25) into (17) to construct $f^{\mathbb{P}}(X_{t+1}|X_t)$ within the model $(DA_M^{\mathbb{Q}}(N), \Lambda)$, we ensure 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. Pursing our connection with continuous time, we define by

$$\Lambda_t \equiv (\Sigma_X S_X(t) \Sigma_X')^{-1} (\mu^{\mathbb{P}}(X_t) - \mu^{\mathbb{Q}}(X_t)), \tag{26}$$

where $\Sigma\sqrt{S(t)}$ is the diffusion term in an $A_M^{\mathbb{Q}}(N)$ affine diffusion model. From the first and second derivatives of the MGF (18) evaluated at $u = 0$,

$$E^{\mathbb{P}}[X_{t+\Delta}|X_t] = \frac{E^{\mathbb{Q}}[X_{t+\Delta} e^{\Lambda_t'(X_{t+\Delta} - E^{\mathbb{Q}}[X_{t+\Delta}|X_t])}]}{E^{\mathbb{Q}}[e^{\Lambda_t'(X_{t+\Delta t} - E^{\mathbb{Q}}[X_{t+1\Delta}|X_t])}]}, \tag{27}$$

$$Cov^{\mathbb{P}}[X_{t+\Delta}|X_t] = \frac{E^{\mathbb{Q}}[(X_{t+\Delta} - E^{\mathbb{P}}[X_{t+\Delta}|X_t]) X_{t+\Delta}' e^{\Lambda_t'(X_{t+\Delta} - E^{\mathbb{Q}}[X_{t+1\Delta}|X_t])}]}{E^{\mathbb{Q}}[e^{\Lambda_t'(X_{t+\Delta t} - E^{\mathbb{Q}}[X_{t+1\Delta}|X_t])}]}. \tag{28}$$

Expanding the numerators and denominators of (27) and (28) in Taylor series and focusing on terms of order less than or equal to Δ (viewed as a small time interval), we obtain

$$E^{\mathbb{P}}[X_{t+\Delta}|X_t] = X_t + \mu^{\mathbb{P}}(X_t)\Delta + o(\Delta) \tag{29}$$

$$Cov^{\mathbb{P}}[X_{t+\Delta}|X_t] = \Sigma S(t) \Sigma' \Delta + o(\Delta). \tag{30}$$

Thus, in the limit to continuous time (i.e., as $\Delta \to 0$) the $\mathbb{P}$ drift of X approaches $\mu^{\mathbb{P}}(X_t)$ and the diffusion term approaches the affine diffusion term $\Sigma\sqrt{S(t)}$. It follows that, starting with an affine specification of the $\mathbb{Q}$ drift $\mu^{\mathbb{Q}}(X_t)$, we can generate essentially any desired nonlinear X_t dependence of the $\mathbb{P}$ drift of X, $\mu^{\mathbb{P}}(X_t)$ by choosing Λ_t as in (26).

Of course with this choice of Λ_t, in discrete time, the conditional Esscher transform (17) in general induces nonlinear conditional $\mathbb{P}$ moments of all orders, not just a nonlinear

conditional mean. For example, Letting Λ_{Zt} and Λ_{Yt} form a conformal partition of Λ_t, the conditional $\mathbb{P}$-mean of the i^{th} member of the M-vector of volatility factors Z_{t+1} is

$$E_t^{\mathbb{P}}\left[Z_{t+1}^i\right] = \frac{\partial}{\partial u_{Zi}}\left[\mathcal{A}(u;\Lambda_t) + \mathcal{B}(u;\Lambda_t)X_t\right]\Big|_{u=0} = \frac{\nu_i c_i}{1-\Lambda_{Zt,i}c_i} + \frac{\rho_i}{(1-\Lambda_{Zt,i}c_i)^2}Z_t. \tag{31}$$

Similarly, the conditional variance of Z_{t+1}^i is given by

$$\text{Var}_t^{\mathbb{P}}[Z_{t+1}^i] = \frac{\nu_i c_i^2}{(1-\Lambda_{Zt,i}c_i)^2} + \frac{2c_i\rho_i Z_t}{(1-\Lambda_{Zt,i}c_i)^3},\ i=1,\ldots,M. \tag{32}$$

The nonlinearity of these moments, in contrast to their affine counterparts under $\mathbb{Q}$ (see (7)), 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 $(A_M^{\mathbb{Q}}(N),\eta)$ models. Within the canonical $A_M^{\mathbb{Q}}(N)$ model Σ_X is normalized to the identity matrix ($\Sigma_X = I$) so, from (26),

$$\Lambda_{Zt} = (diag[Z_t^i])^{-1}[(diag[Z_t^i])^{1/2}\eta_{Zt}]. \tag{33}$$

From (23) it follows that, to first order, the term in brackets, $[(diag[Z_t^i])^{1/2}\eta_{Zt}]$ is the adjustment to the drift of Z in the measure change from $\mathbb{Q}$ to $\mathbb{P}$. The special case of Dai and Singleton (2000)'s "completely affine" specification of η_Z has $(diag[Z_t^i])^{1/2}\eta_{Zt} = \lambda_{Z1}^D Z_t$, where λ_{Z1}^D is an $M\times M$ *diagonal* matrix. Therefore, under their *MPR*,

$$\Lambda_{Zt} = \left(diag[Z_t^i]\right)^{-1}\lambda_{Z1}^D Z_t = \lambda_{Z1}^D, \tag{34}$$

and (31) and (32) imply that the conditional moments of Z are affine under $\mathbb{P}$ as well as under $\mathbb{Q}$. In other words, completely affine $(A_M^{\mathbb{Q}}(N),\eta)$ and $(DA_M^{\mathbb{Q}}(N),\Lambda)$ 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 (2003), and Cheridito, Filipovic, and Kimmel (2003) has

$$\Lambda_{Zt} = \left(diag[Z_t^i]\right)^{-1}\left(\sqrt{diag[Z_t^i]}\lambda_d + (\lambda_{Z0} + \lambda_{Z1}Z_t) + \Upsilon_{Zt}\right), \tag{35}$$

where λ_d and λ_{Z0} are $M\times 1$ vectors and λ_{Z1} is a (not necessarily diagonal) $M\times M$ matrix. Setting $\lambda_d = 0$ and $\Upsilon_{Zt} = 0$, and imposing sufficient structure on λ_{Z0} and λ_{Z1} to ensure non-attainment by Z of the zero boundary under $\mathbb{P}$ and $\mathbb{Q}$, gives the model in Cheridito, et. al. The special case of $\lambda_{Z0} = 0$ and $\Upsilon_{Zt} = 0$ gives Duarte's model. Whenever Λ_{Zt} is state-dependent, the conditional $\mathbb{P}$-moments of Z_{t+1} show nonlinear dependence on Z_t. The term Υ_{Zt} is introduced to illustrate that the modeler is free to add essentially any nonlinear dependence of Λ_{Zt} on Z.[18] We investigate empirically a model with $\Upsilon_{Zt}\neq 0$ in Section 6.

[18] We restrict attention to cases where Υ_{Zt} depends only on Z to ensure that Z remains a non-negative process under $\mathbb{P}$.

Turning to the conditionally Gaussian components of X, and recalling the definitions in (8), the conditional mean of Y_{t+1} under $\mathbb{P}$ is

$$E_t^{\mathbb{P}}[Y_{t+1}] = \omega_{Yt}^{\mathbb{Q}} + \Omega_{Yt}\Lambda_{Yt}. \tag{36}$$

To interpret the consequences of alternative specifications of Λ for the functional form of $E_t^{\mathbb{P}}[Y_{t+1}]$, it is instructive to express the market price of risk for the entire state vector X in a $(DA_M^{\mathbb{Q}}(N), \Lambda)$ model (again under the normalization $\Sigma_X = I$) as

$$\Lambda_{Xt} = S_{Xt}^{-1}\left(\lambda_{X0} + \lambda_{X1}X_t + \sqrt{S_{Xt}}\lambda_{Xd} + \Upsilon_{Xt}\right), \tag{37}$$

where λ_{X0} and λ_{Xd} are $N \times 1$ vectors of constants, λ_{X1} 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 subvector of market prices of risk associated with Y is thus

$$\Lambda_{Yt} = S_{Yt}^{-1}\left(\lambda_{Y0} + \lambda_{Y1}X_t + \sqrt{S_{Yt}}\lambda_{Yd} + \Upsilon_{Yt}\right), \tag{38}$$

where λ_{Y1} is the $(N-M) \times N$ matrix containing the last $N-M$ rows of λ_{X1}. Substituting (38) into (36) gives

$$E^{\mathbb{P}}[Y_{t+1}|Y_t] = \omega_{Yt}^{\mathbb{Q}} + \lambda_{Y0} + \lambda_{Y1}X_t + \sqrt{S_{Yt}}\lambda_{Yd} + \Upsilon_{Yt}. \tag{39}$$

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– the square roots of affine functions of X– to the drift of Y. 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 (31) with (32)). Indeed, the variance has a very similar structure to the mean with each term of $Var_t^{\mathbb{P}}[Z_{t+1}^i]$ 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 ensure 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.[19] That X will not be a geometrically ergodic process for all specifications of Λ_t can be seen immediately from (31). 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 (36) 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 ensures 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 process (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. *$\Lambda_Z(z) \to \bar{\lambda} \leq 0$ as $z \to \infty$ 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 (35) 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 of geometric ergodicity on a case-by-case basis, as we do in our illustrations.

Finally we note that, for our particular choice of Radon-Nykodym derivative, there is also a computationally fast way to simulate directly from the conditional $\mathbb{P}$ distribution

[19] 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).

of X. Specifically, returning to the exponential-affine representations (1) and (18) for the conditional *MGF*s, upon making the dependence of the coefficients $a(\cdot)$ and $b(\cdot)$ of $\phi^{\mathbb{Q}}$ on the risk-neutral parameters explicit by writing

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

the coefficients $\mathcal{A}(u,v)$ and $\mathcal{B}(u,v)$ of $\phi^{\mathbb{P}}$ 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

$$\begin{aligned} c_i(v) &= \frac{c_i}{1 - v_{Z,i} c_i},\ \rho_i(v) = \frac{\rho_i}{(1 - v_{Z,i} c_i)^2}, \\ \mu_0(v) &= \mu_0 + h_0' v_Y,\ \mu_Y(v) = \left(\begin{array}{cc} \mu_Y^Z + \{h_i' v_Y\}_{i=1,2,\ldots,M} & \mu_Y^Y \end{array} \right). \end{aligned}$$

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 $\Theta^{\mathbb{P}}(\Lambda_t)$. 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. It follows that, given X_t, the value of the state at date $t+1$ can be simulated exactly using the $\mathbb{Q}$ density, with the parameters adjusted to reflect the state dependence induced by the measure change.

Now consider the problem of computing the conditional $\mathbb{P}$-expectation of a measurable function $g(X_{t+\tau})$, for any $\tau > 1$, by Monte Carlo methods. Such computations can be approached in either of two ways. First, defining the random variable

$$\pi_{t,t+\tau}^D = \prod_{j=1}^{\tau} \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \right)_{t+j-1,t+j}^D, \tag{40}$$

we can write

$$E^{\mathbb{P}}\left[g(X_{t+\tau}) | X_t \right] = E^{\mathbb{Q}}\left[g(X_{t+\tau}) \pi_{t,t+\tau}^D | X_t \right]. \tag{41}$$

The expectation on the right-hand-side of (41) can be computed, for a given value of X_t, by simulation under $\mathbb{Q}$ using the known density $f^{\mathbb{Q}}(X_{t+1}|X_t)$. Moreover, the nonlinearity in the $\mathbb{P}$ distribution– its non-affine structure– is captured through the random variable $\pi_{t,t+\tau}^D$ which is also known in closed form.

Alternatively, using the preceding short-cut to simulating from the $\mathbb{P}$ distribution of X directly, we can compute the left-hand side of (41) by Monte Carlo simulation without reference to the right-hand side. This second approach is used in our empirical illustrations in Section 6.

5 Sampling Intervals Longer than Modeling Intervals

Up to this point we have presumed that the sampling interval of the data coincides with the modeling interval for the risk factors. That is, the discrete time interval over which the distribution $f^{\mathbb{Q}}(X_{t+1}|X_t)$ is specified matches the sampling interval of the data. For the purpose of standard term structure modeling, with latent risk factors X, this assumption seems innocuous in that bond yield data can often be sampled a frequencies as short as one day. The decision the researcher is left with then is at what sampling (equals modeling) interval will the model-implied distributions under $\mathbb{P}$ and $\mathbb{Q}$ best match the historical and pricing distributions, respectively, of bond yields.

The inclusion of macroeconomic information, as in the growing literature on macro term structure models, often leads researchers to work with the monthly or quarterly sampling intervals. While sampling and decision intervals are frequently equated in situations like this in the macroeconomics literature, there may well be circumstances where one would like to allow the decision interval to be shorter than the sampling interval.

When working with continuous time models researchers fix, somewhat artificially, the modeling interval to be instantaneous, and then estimation is typically based on moments of the implied conditional distribution of X for the relevant sampling interval. The model-implied likelihood function is not used (at least in its exact form) because, for $A_M(N)$ models with $M < N$, it is not known in closed form. In extending our framework, we encounter precisely the same obstacles with regard to the conditional $\mathbb{Q}$ distribution of the state.

To fix notation, we let t index the modeling interval and τ index the sampling interval, with $\tau \geq 1$. Consider first the $\mathbb{Q}$ distribution $f^{\mathbb{Q}}(X_{t+\tau}|X_t)$. Repeated application of (1) and the law of iterated expectations gives

$$\phi_\tau^{\mathbb{Q}}(u; X_t) \equiv E^{\mathbb{Q}}\left[e^{u'X_{t+\tau}}\middle| X_t\right] = e^{\sum_{j=0}^{\tau-1} a(\tilde{b}^{j'}(u)) + \tilde{b}^{\tau-1}(u)X_t}, \tag{42}$$

where $\tilde{b}$ is defined recursively by $\tilde{b}^0(u) = b(u)$, $\tilde{b}^1(u) = b(\tilde{b}^{0'}(u))$, etc. Thus the conditional distribution of $X_{t+\tau}$ given X_t is that of an affine process. In general, the functional form of this distribution is unknown.[20] One can, in principle, compute $f^{\mathbb{Q}}(X_{t+\tau}|X_t)$ by Fourier inversion of the conditional characteristic function (see, e.g., Singleton (2001)). However, in multi-variate setting this computation can become burdensome. In Appendix B we discuss special cases– when the sampling and modeling intervals do not coincide– where the conditional density implied by (42) is known in closed from, as well as some accurate approximations for some cases where this density is not known.

6 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

[20] This is the discrete-time counterpart to the result in Duffie, Pan, and Singleton (2000) that the characteristic functions of general affine diffusions are exponential affine, again with the implied densities of these processes being generally not known in closed form.

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:

$$dX_t = \kappa^{\mathbb{Q}}(\theta^{\mathbb{Q}} - X_t)dt + \Sigma_X \sqrt{S(t)} dB_t^{\mathbb{Q}} \tag{43}$$

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 $r_t = \delta_0 + \delta_X' X_t$, where $X' = (Z, Y')$ with Z following a one-dimensional autoregressive-gamma process and Y following a two-dimensional Gaussian process conditional on X, and $\delta_X' = (\delta_Z, \delta_Y')'$. Consequently, the yield on an n-period bond y_t^n satisfies

$$y_t^n = \delta_0^n + \delta_X^{n\prime} X_t. \tag{44}$$

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

$$\kappa^{\mathbb{Q}} = \begin{bmatrix} \kappa^{ZZ} & 0 \\ \kappa_{2\times 1}^{YZ} & \kappa_{2\times 2}^{YY} \end{bmatrix}; \tag{45}$$

$\Sigma_X = I_3$; and

$$\theta^{\mathbb{Q}} = \begin{bmatrix} \theta_1^{\mathbb{Q}} \\ 0_{2\times 1} \end{bmatrix},\ \alpha = \begin{bmatrix} 0 \\ 1_{2\times 1} \end{bmatrix},\ \beta = \begin{bmatrix} 1 & 0_{1\times 2} \\ \beta_{2\times 1}^{YZ} & 0_{2\times 2} \end{bmatrix}. \tag{46}$$

For the autoregressive-gamma process Z we impose conditions to ensure non-attainment of the zero boundary under both $\mathbb{P}$ and $\mathbb{Q}$:

$$\nu_1^{\mathbb{P}} = \frac{2\kappa_{(1,1)}^{\mathbb{P}}\theta_{(1,1)}^{\mathbb{P}}}{\Sigma_{(1,1)}^2} > 1, \tag{47}$$

$$\nu_1^{\mathbb{Q}} = \frac{2\kappa_{(1,1)}^{\mathbb{Q}}\theta_{(1,1)}^{\mathbb{Q}}}{\Sigma_{(1,1)}^2} > 1. \tag{48}$$

Dai and Singleton (2000), Duarte (2003) and Duffee (2002) imposed the weaker requirement that $\nu_1^{\mathbb{P}} > 0$ and $\nu_1^{\mathbb{Q}} > 0$. However, in addition to having $\Upsilon_{Zt} \neq 0$, we extend their specifications of the market price of risk by letting $\lambda_{Z0} \neq 0$ in (35). As discussed in Cheridito, Filipovic, and Kimmel (2003) within a setting with $\Upsilon_{Zt} = 0$, imposing both (47) and (48)

is sufficient to rule out arbitrage opportunities within our richer model. So long as Υ_{Zt} is bounded, their analysis also ensures that our extended models of Z also rule out arbitrage opportunities. For consistency, we impose these non-attainment restrictions in all of our illustrative models.

Two different specifications of the market prices of risk for the volatility factor Z are examined:

Model NLDA: A $DA_1^{\mathbb{Q}}(3)$ model with nonlinear polynomial drift for the volatility factor Z. The market prices of risk are given by (37) with $\lambda_{Xd} = 0$, $\Upsilon_{Yt} = 0$, and Υ_{Zt} being an affine function of Z^2, Z^3 and Z^4.

Model SASR This is Duarte's model which starts from an $DA_1^{\mathbb{Q}}(3)$ model and introduces a square-root term in the $\mathbb{P}$-drift of Z through the market price of risk. Λ_t is given by (37) with $\Upsilon_{Xt} = 0$, $\lambda_{Z0} = 0$, and $\lambda'_{Xd} = (\lambda_{Zd}, 0, 0)$.

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

$$\begin{pmatrix} \kappa^{\mathbb{P}}_{1,1} & 0 & 0 \\ \kappa^{\mathbb{P}}_{2,1} & \kappa^{\mathbb{P}}_{2,2} & \kappa^{\mathbb{P}}_{2,3} \\ \kappa^{\mathbb{P}}_{3,1} & \kappa^{\mathbb{P}}_{3,2} & \kappa^{\mathbb{P}}_{3,3} \end{pmatrix} \left[\begin{pmatrix} \theta^{\mathbb{P}}_1 \\ \theta^{\mathbb{P}}_2 \\ \theta^{\mathbb{P}}_3 \end{pmatrix} - \begin{pmatrix} Z_t \\ Y_{1t} \\ Y_{2t} \end{pmatrix} \right] + \begin{pmatrix} \lambda_d \sqrt{Z_t} + \lambda_{Z2} Z_t^2 + \lambda_{Z3} Z_t^3 + \lambda_{Z4} Z_t^4 \\ 0 \\ 0 \end{pmatrix}. \quad (49)$$

The SASR model is the special case with $\lambda_{Z2} = \lambda_{Z3} = \lambda_{Z4} = 0$, and the polynomial models are the special cases with $\lambda_{Zd} = 0$. In subsequent discussions we will refer to the NLDA models as the quadratic, cubic and quartic models, corresponding to the maximal power in their polynomial drift specification.

Our focus on these nonlinear *DTSM*s 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 $M > 0$. 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 functions of all of our discrete-time nonlinear *DTSM*s are known in closed-form and these examples illustrate 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 in the $\mathbb{P}$ distributions of Z. In particular, Duffee (2002) found that his essentially affine model substantially outperformed completely affine models in matching 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 $DA_1(3)$ models with the introduction of our more general market price of volatility risk. 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.[21] We assumed that bonds with .5, 2 and 10

[21] This is the same data used in Backus, Foresi, and Zin (1998) and Dai and Singleton (2002).

years to maturity were priced without errors, while bonds with maturities of 1, 5 and 7 years were priced with serially independent Gaussian errors. The data for the period 1996-2000 were omitted from the estimation in order to examine out-of-sample fits of the models.

The "maximal" models (in the sense of Dai and Singleton (2000)) are highly parameterized and, therefore, it is desirable to explore more parsimonious models that set those parameters with large associated standard errors to zero. One might expect to arrive at different "preferred" parsimonious models depending on which maximal model one starts from (e.g., the linear affine model, the fully parameterized NLDA model, etc.). We have chosen to focus our analysis on two preferred parameterizations: PMI is simplified from the maximal cubic model[22] and PMII is simplified from the maximal linear and SASR models.[23]

Table 1 reports the likelihood values and degrees of freedom of the estimated models. Next to the likelihood values of the preferred models are the likelihood ratio (LR) statistics and associated p-values for the constraints imposed in simplifying to the models PMI and PMII. All preferred models show very little deterioration in fit relative to the their maximal counterparts.[24] Henceforth, we focus on models PMI and PMII.

	Maximal[df]	**PMI[df]**	**LR [pval]**	**PMII[df]**	**LR [pval]**
Linear	10475.6 [30]	10471.5 [23]	8.38 [0.30]	10473.0 [22]	5.31 [0.72]
Nonlinear					
Quadratic	10476.9 [31]	10471.5 [24]	10.65 [0.15]	10473.9 [23]	5.89 [0.66]
Cubic	10485.9 [32]	10481.7 [25]	8.39 [0.30]	10478.8 [24]	14.17 [0.08]
Quartic	10486.6 [33]	10482.3 [26]	8.60 [0.20]	10481.1 [25]	11.11 [0.20]
SASR	10475.9 [31]	10471.5 [24]	8.77 [0.27]	10473.2 [23]	5.42 [0.71]

Table 1: **Log-likelihood values and LR statistics:** Log-likelihood values and degrees of freedom (in brackets) of the maximal and preferred models. Next to each likelihood value for models PMI and PMII are the likelihood ratio statistic (LR) and p-value (in brackets) for the associated constraints.

As a first step in evaluating the contribution of nonlinear $\mathbb{P}$-drifts to the fit of our models we compute likelihood ratio tests of the restricted models indicated in the column headings of Table 2 against the alternatives indicated by the row headings. Both the linear and SASR models are rejected (at conventional significance levels) against the NLDA model with a cubic term in the drift of the volatility factor. Adding a quartic term in addition to the cubic term does not add significantly to the fit of the NLDA model PMI, but does provide some incremental improvement in fit within model PMII. One reason for this might be that

[22] We chose the cubic model as representative for the polynomial models since adding a term in the forth power of Z did not substantially improve the fit.

[23] The maximal linear and SASR models reduce to the same preferred model. See the subsequent discussion of the parameter estimates.

[24] The only exception at a 10% significance level is the PMII cubic model where the likelihood ratio test statistic is significant at 8% confidence level.

the simplification achieved started from the $\mathbb{P}$-affine PMII model (as contrasted with the $\mathbb{P}$-cubic model) over-simplifies the dynamic structure of the conditional Gaussian factors under $\mathbb{P}$ and this is compensated for by the quartic term under the alternative formulation. Squared and SASR terms by themselves only modestly improve the fits of the models and, in fact, the nested linear model is not rejected against the alternative SASR model.

	Linear		**SASR**		**Cubic**	
PMI						
Quadratic	0.16	[0.69]	-	-	-	-
Cubic	20.44	[0.00]	20.29	[0.00]	-	-
Quartic	21.77	[0.00]	21.61	[0.00]	1.32	[0.25]
SASR	0.06	[0.80]	-	-	-	-
PMI						
Quadratic	1.85	[0.17]	-	-	-	-
Cubic	11.59	[0.00]	9.74	[0.00]	-	-
Quartic	16.19	[0.00]	14.34	[0.00]	4.60	[0.03]
SASR	0.35	[0.55]	-	-	-	-

Table 2: Likelihood ratio tests, along with their associated p-values are displayed for various nested special cases of the models examined. Each row defines the alternative model, while the column defines the null (constrained) model. p-values are given in brackets.

The *ML* estimates of the parameters under $\mathbb{Q}$ of the linear, SASR and cubic NLDA models within the PMI and PMII families are displayed in Table 3. For each parameter, the first number in parentheses is its estimated asymptotic standard error and the second is the likelihood ratio statistic for the null hypothesis that the parameter is zero. In those cases where the latter null hypothesis places us on the boundary of the admissible parameter space (e.g., $\kappa^{\mathbb{Q}}(i,i)$, $i = 1,2,3$ and $\theta^{\mathbb{Q}}(1)$), we display a "*". Parameters that are set to zero in simplifying the maximal models to the preferred models are indicated by a dash. Within each family of preferred models (I or II), the corresponding estimates across the linear, SASR, and cubic models are quite similar. Estimates across the PMI and PMII families of models are also similar, with the primary exception being $\beta(2,1)$, the parameter governing the dependence of the volatility of the first conditionally Gaussian factor Y_1 on the volatility factor Z. The latter parameter is much larger in PMI than in PMII.

The near invariance of the $\mathbb{Q}$ parameters to alternative formulations of the $\mathbb{P}$ distributions of bond yields (through different choices of Λ_t) is striking. This finding implies that the pricing implications of these $DTSM$s are invariant to the assumed $\mathbb{P}$ distribution of bond yields. Put differently, the same pricing parameters are extracted from a variety of models, at least some of which are surely misspecified (under $\mathbb{P}$). Evidently the cross-maturity restrictions implied by the no-arbitrage pricing restrictions are easily satisfied, and recovered econometrically, within a variety of $DTSM$s.

Figure 1 displays the loadings on the state variables within the cubic NLDA models for zero-coupon bonds with maturities ranging from three months to ten years. The graphs for all three models– linear, SASR, and cubic– are virtually on top of each other because of the

A. Preferred Model I

	Linear		SASR		Cubic	
$\kappa^{\mathbb{Q}}(1,1)$	0.56	(0.02 , *)	0.55	(0.02 , *)	0.56	(0.02 , *)
$\kappa^{\mathbb{Q}}(2,1)$	-10.5	(36.6 , 70.7)	-10.7	(48.3 , 67.4)	-12.4	(8.03 , 86.2)
$\kappa^{\mathbb{Q}}(3,1)$	0.22	(0.09 , 13.6)	0.21	(0.09 , 13.2)	0.24	(0.09 , 15.2)
$\kappa^{\mathbb{Q}}(2,2)$	1.63	(0.18 , *)	1.66	(0.17 , *)	1.65	(0.17 , *)
$\kappa^{\mathbb{Q}}(3,2)$	-0.03	(0.12 , 14.5)	-0.03	(0.15 , 14.1)	-0.03	(0.03 , 16.1)
$\kappa^{\mathbb{Q}}(2,3)$	-9.89	(33.9 , 73.0)	-9.78	(43.9 , 72.4)	-11.4	(6.93 , 81.9)
$\kappa^{\mathbb{Q}}(3,3)$	0.21	(0.16 , *)	0.20	(0.15 , *)	0.23	(0.16 , *)
$\theta^{\mathbb{Q}}(1,1)$	5.85	(1.10 , *)	5.83	(1.20 , *)	5.44	(0.804 , *)
$\beta(2,1)$	38.4	(267 , 78.5)	38.2	(347 , 77.1)	51.4	(61.5 , 85.3)
$\beta(3,1)$	-	(NA)	-	(NA)	-	(NA)
$\delta_0(1,1)$	0.699	(0.516 , 109)	0.684	(0.508 , 105)	0.645	(0.433 , 125)
$\delta_Y(1,1)$	0.003	(0.000 , 33.1)	0.003	(0.001 , 33.1)	0.003	(0.001 , 34.0)
$\delta_Y(2,1)$	0.002	(0.005 , 124)	0.002	(0.007 , 125)	0.001	(0.001 , 145)
$\delta_Y(3,1)$	0.002	(0.003 , 0.26)	0.002	(0.003 , 0.31)	0.002	(0.003 , 4.59)

B. Preferred Model II

	Linear		SASR		Cubic	
$\kappa^{\mathbb{Q}}(1,1)$	0.56	(0.02 , *)	0.56	(0.02 , *)	0.56	(0.02 , *)
$\kappa^{\mathbb{Q}}(2,1)$	-4.78	(6.91 , 70.5)	-4.78	(6.63 , 70.1)	-4.65	(6.66 , 79.5)
$\kappa^{\mathbb{Q}}(3,1)$	0.29	(0.05 , 58.6)	0.29	(0.05 , 58.5)	0.28	(0.04 , 59.2)
$\kappa^{\mathbb{Q}}(2,2)$	1.55	(0.10 , *)	1.55	(0.10 , *)	1.53	(0.10 , *)
$\kappa^{\mathbb{Q}}(3,2)$	-0.10	(0.13 , 63.9)	-0.10	(0.13 , 63.9)	-0.09	(0.13 , 64.5)
$\kappa^{\mathbb{Q}}(2,3)$	-5.23	(7.51 , *)	-5.21	(7.17 , *)	-5.27	(7.46 , *)
$\kappa^{\mathbb{Q}}(3,3)$	0.32	(0.04 , *)	0.32	(0.04 , *)	0.32	(0.04 , *)
$\theta^{\mathbb{Q}}(1,1)$	5.66	(0.95 , *)	5.63	(0.95 , *)	5.89	(1.04 , *)
$\beta(2,1)$	7.65	(22.4 , 70.8)	7.7	(21.5 , 70.7)	7.67	(22.2 , 72.2)
$\beta(3,1)$	-	(NA)	-	(NA)	-	(NA)
$\delta_0(1,1)$	0.671	(0.464 , 94.4)	0.672	(0.467 , 93.9)	0.715	(0.555 , 99.4)
$\delta_Y(1,1)$	0.003	(0.000 , 29.5)	0.003	(0.000 , 29.4)	0.003	(0.000 , 29.5)
$\delta_Y(2,1)$	0.004	(0.005 , *)	0.004	(0.005 , *)	0.003	(0.005 , *)
$\delta_Y(3)$	-	(NA)	-	(NA)	-	(NA)

Table 3: **Parameters of the $\mathbb{Q}$ representation of r_t:** The first number in parentheses is the associated asymptotic standard error and the second is the likelihood ratio test statistic for the null hypothesis that the parameter is zero. For the LR statistic, we report a "*" in those cases where the null would place us on the boundary of the admissible parameter space. A "-" indicates that the parameter was set to zero in the preferred model.

similarity in the parameters of the $\mathbb{Q}$ representation of r in Table 3. For the same reason, the corresponding loadings across the preferred families have a similar shape. The loadings on the volatility factor Z capture curvature at the very short end the yield curve and have a "slope"-like shape for maturities beyond a year or so. The third factor appears to capture features of both the "level" and "slope" factors in standard principal components analyses of bond yields (e.g., Litterman, Scheinkman, and Weiss (1988)). Its level-like shape for longer maturities is reflected in its degree of mean reversion. $\kappa^{\mathbb{Q}}(3,3)$ is the smallest of the diagonal

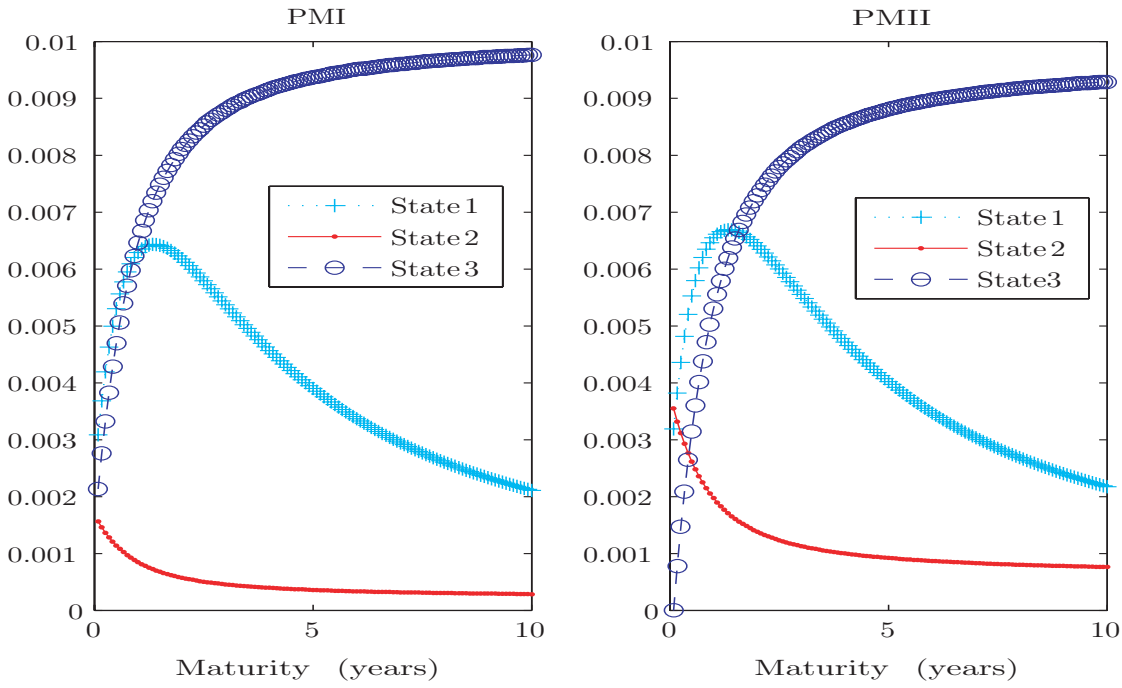

Figure 1: **Loadings on Risk Factors:** δ_X^n as n varies across maturities for cubic models within the families PMI (left) and PMII (right).

elements of $\kappa^{\mathbb{Q}}$, consistent with the typical finding that the "level" of the yield curve shows the least amount of mean reversion. The largest diagonal element of $\kappa^{\mathbb{Q}}$ is $\kappa^{\mathbb{Q}}(2,2)$.

Table 4 displays the parameters that determine the market price of risk Λ_t. Rows for the parameters $\lambda_0(3)$, $\lambda_1(2,1)$, $\lambda_1(3,1)$, and $\lambda_1(2,2)$ are omitted, because these parameters were set to zero in *both* the preferred families of models PMI and PMII. Focusing first on the parameter vector λ_0, we see that the test statistics suggest that the extension of square-root diffusion models suggested by Cheridito, Filipovic, and Kimmel (2003) is not playing a substantial role in our analysis of the PMII family of models (it is set to zero). There is some evidence for a role of their extension from the LR statistic within the cubic PMI model, though not from the asymptotic Wald test.

The parameter governing the "completely affine" market prices of risk for the volatility factor ($\lambda_1(1,1)$) is significantly different from zero in most of the PMI models (according to the LR statistics), but this parameter is set to zero in arriving at the parsimonious PMII models. Evidently, within the latter family, the nonlinear polynomial terms suffice to capture the market prices of risk for the volatility factor Z.

Within the NLDA models, the nonlinearity in the drift of Z is governed by the parameters λ_{Z2} (the quadratic term) and λ_{Z3} (the cubic term). For the cubic models, both the asymptotic t- and LR statistics indicate rejection of the null hypotheses that $\lambda_{Z2} = 0$ and $\lambda_{Z3} = 0$ at conventional significance levels. With regard to the SASR specification, λ_d is positive for the PMI models and negative for the PMII models, and in both cases it is

A. Preferred Model I

	Linear		SASR		Cubic	
$\lambda_0(1)$	1.69	(1.16 , 2.25)	-0.30	(2.95 , -0.08)	4.08	(4.50 , 8.05)
$\lambda_0(2)$	-259	(915 , 10.1)	-259	(1196 , 9.72)	-259	(288 , 11.4)
$\lambda_1(1,1)$	-0.39	(0.23 , 2.94)	-1.02	(1.28 , 0.49)	-6.04	(3.75 , 11.2)
$\lambda_1(3,2)$	-	(NA)	-	(NA)	-	(NA)
$\lambda_1(2,3)$	-4.59	(16.0 , 7.58)	-4.68	(21.1 , 7.53)	-5.04	(3.71 , 8.65)
$\lambda_1(3,3)$	-	(NA)	-	(NA)	-	(NA)
λ_d			2.31	(4.29 , 0.06)		
λ_{Z2}					2.06	(0.85 , 14.2)
λ_{Z3}					-0.20	(0.07 , 20.3)

B. Preferred Model II

	Linear		SASR		Cubic	
$\lambda_0(1)$	-	(NA)	-	(NA)	-	(NA)
$\lambda_0(2)$	-114	(195 , 9.05)	-114	(190 , 8.87)	-114	(202 , 8.44)
$\lambda_1(1,1)$	-	(NA)	-	(NA)	-	(NA)
$\lambda_1(3,2)$	0.08	(0.11 , 10.3)	0.08	(0.11 , 5.62)	0.08	(0.11 , 7.97)
$\lambda_1(2,3)$	-2.18	(3.23 , 8.48)	-2.18	(3.09 , 8.42)	-2.03	(2.98 , 7.93)
$\lambda_1(3,3)$	-0.27	(0.10 , 14.9)	-0.27	(0.10 , 14.9)	-0.27	(0.10 , 14.6)
λ_d			-0.12	(0.23 , 0.35)		
λ_{Z2}					0.22	(0.10 , 8.01)
λ_{Z3}					-0.04	(0.02 , 9.74)

Table 4: This table reports the parameters of the market prices of risk Λ_t. The first number in parentheses is the associated asymptotic standard error and the second is the likelihood ratio test statistic for the null hypothesis that the parameter is zero. A "-" indicates that the parameter was set to zero in the preferred model.

estimated with relatively little precision.

The estimated λ_{Z3} are negative in both the PMI and PMII models. This induces a similar shape to the conditional mean function as the one documented by Ait-Sahalia (1996) and Ang and Bekaert (2002). Ait-Sahalia (1996) noted that short-term interest rates tend to behave like a random walk when rates are in the vicinity of their average values, but tend to mean revert at increasing rates as the levels of rates become unusually high or low. As Z becomes large, the pull-to-the-mean from this (negative) cubic term is much stronger than that of the (negative) linear term. This is visually illustrated in Figure 2 which plots the drifts of Z for the cubic and SASR models. The drift in the PMI cubic model (on the left) 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 sharply for smaller values of Z than are depicted in this graph, though the likelihood of observing values in this region is very small. The drift of the PMII cubic model (on the right) also displays mean reversion around the mean, but to a much lesser extent. This may be due to the fact that the PMII family is a simplified

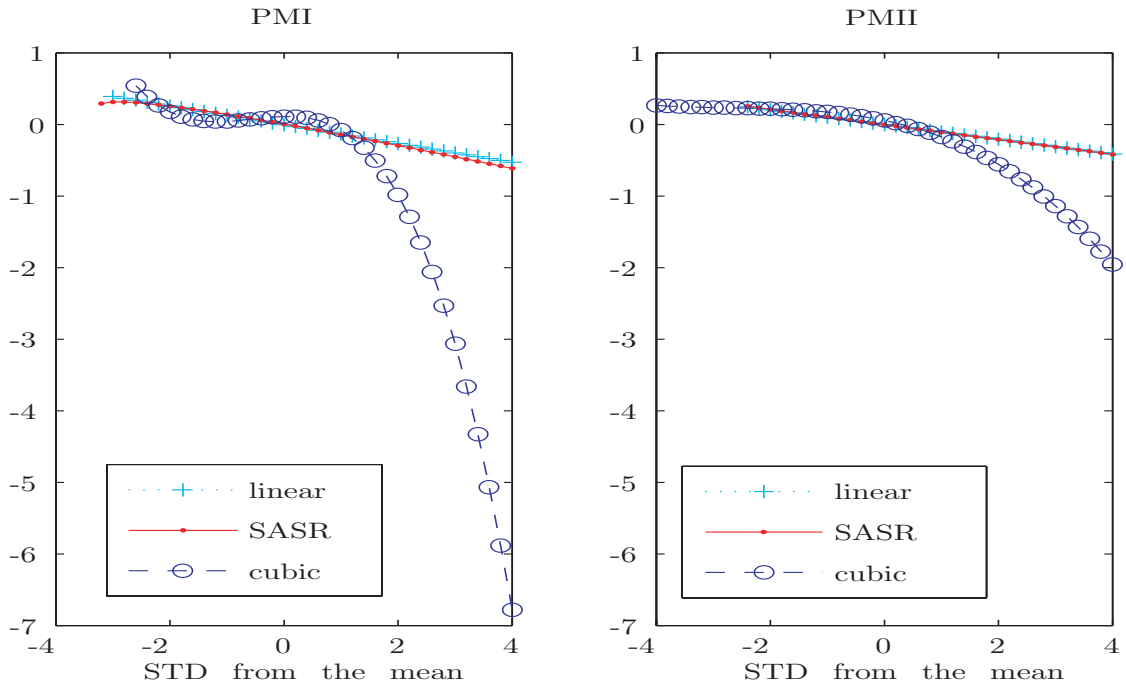

Figure 2: The drift of the volatility factor, plotted as a function of the level of volatility measured in units of standard deviation from from the mean of Z.

version of the maximal linear model.

Pursing these differences, the drift of Z in the cubic PMII model crosses zero at one point, close to the mean of its stationary distribution implied by this model. In this regard, over a wide range around its long-run mean, Z behaves much like a $\mathbb{P}$-affine model: the drift of Z is nearly linear and is zero at its long-run mean. In contrast, the drift of Z in the cubic PMI model crosses zero three times at the points -1.3, -0.95, and 0.40 standard deviations from its long-run mean. Though the precise crossing points will be sensitive to the value of $\kappa^{\mathbb{P}}(1,1)\theta^{\mathbb{P}}(1)$, the fitted pattern for the cubic PMI model highlights the sharp contrast between the essentially zero drift of Z over a wide portion of the support of its distribution and the rapid mean reversion in both tails. This pattern is characteristic of that generated by regime-switching models of interest rates (e.g., Ang and Bekaert (2002)).

In both the SASR model (due to the presence of $\sqrt{Z}$) and the quadratic NLDA model (due to the presence of Z^2), there is the possibility of higher mean reversion than in the linear model as Z increases relative to its unconditional mean. However, this state-dependent mean reversion is one-sided: it either shows higher mean reversion for large values of Z or higher mean reversion for small values of Z, but not both. Particularly when the critical point of the parabolic shape of the drift is near the unconditional mean of Z, we will see mean reversion in one direction and mean diversion in the other. When only one wing of the parabola is used in modelling the drift as in the SASR specification, there is no possibility of inducing the "S on its side" pattern that we saw in the cubic models. This difference is illustrated

clearly in Figure 2 where it is seen that the drift of the SASR model does not turn up for low values of Z.

	Linear		**SASR**		**Cubic**	
	PMI	**PMII**	**PMI**	**PMII**	**PMI**	**PMII**
$\kappa^{\mathbb{P}}(1,1)$	0.95	0.56	1.57	0.56	6.60	0.56
$\kappa^{\mathbb{P}}(2,1)$	-10.55	-4.78	-10.67	-4.78	-12.41	-4.65
$\kappa^{\mathbb{P}}(3,1)$	0.22	0.29	0.21	0.29	0.24	0.28
$\kappa^{\mathbb{P}}(2,2)$	1.63	1.55	1.66	1.55	1.65	1.53
$\kappa^{\mathbb{P}}(3,2)$	-0.03	-0.17	-0.03	-0.17	-0.03	-0.17
$\kappa^{\mathbb{P}}(2,3)$	-5.30	-3.04	-5.11	-3.03	-6.40	-3.23
$\kappa^{\mathbb{P}}(3,3)$	0.21	0.60	0.20	0.60	0.23	0.59
$\theta^{\mathbb{P}}(1)$	5.20	5.66	1.87	5.63	1.08	5.89
$\theta^{\mathbb{P}}(2)$	-342.00	-171.64	-349.13	-171.49	-385.25	-188.02
$\theta^{\mathbb{P}}(3)$	-55.29	-50.05	-54.40	-50.19	-50.43	-53.78

Table 5: This table reports the parameters of the $\mathbb{P}$-representation of r_t. For each model, the PMI results are presented on the left and the PMII results are presented on the right.

The estimated values of the parameters of the $\mathbb{P}$-distribution of r_t are displayed in Table 5. Most of the parameters governing the distribution of the conditional Gaussian Y are similar, both across models and the families PMI and PMII. Most notably, for the volatility factor Z, the parameter governing *linear* mean reversion ($\kappa^{\mathbb{P}}(1,1)$) is much larger for the nonlinear (SASR and cubic) models in family PMI than in the family PMII. At the same time $\theta^{\mathbb{P}}(1)$ is larger in the PMII family than in the PMI family. That is, the affine component of the drift of Z shows relatively fast mean reversion to a relatively small $\theta^{\mathbb{P}}(1)$ compared to the nonlinear models in the family PMII.

The implications of these differences for the *within* sample fits of the models is summarized graphically in Figure 3 where the root mean-squared errors (RMSE), computed over the sample period 1970:1 to 1995:12, are displayed. The forecast horizon is six months and the state variables used in forecasting are inverted from the models at the ML estimates.[25] One thousand six-month ahead simulations are computed at each point in time and the resulting yields are then averaged to compute the forecasts for future yields.[26] Pricing errors are differences between these forecasts and the actual yields realized at the forecast horizon. For ease of exposition, we only show graphs for the maturities of five years and above, noting that the pattern is very similar for other maturities.

We see that, within sample, the cubic models out perform both the linear and SASR models. Comparing across the preferred families PMI and PMII, the cubic model in family PMII (obtained by first eliminating insignificant parameters in the maximal linear model) produces the lower RMSEs. We conclude that, based on the metric of lower RMSEs, the cubic model with slow reversion to a high level of Z through the affine component of the drift performs the best.

Of equal, if not greater, interest is the out-of-sample forecasting performance of the

[25] We experimented with forecast horizons of 4, 8 and 10 months and obtained essentially the same pattern

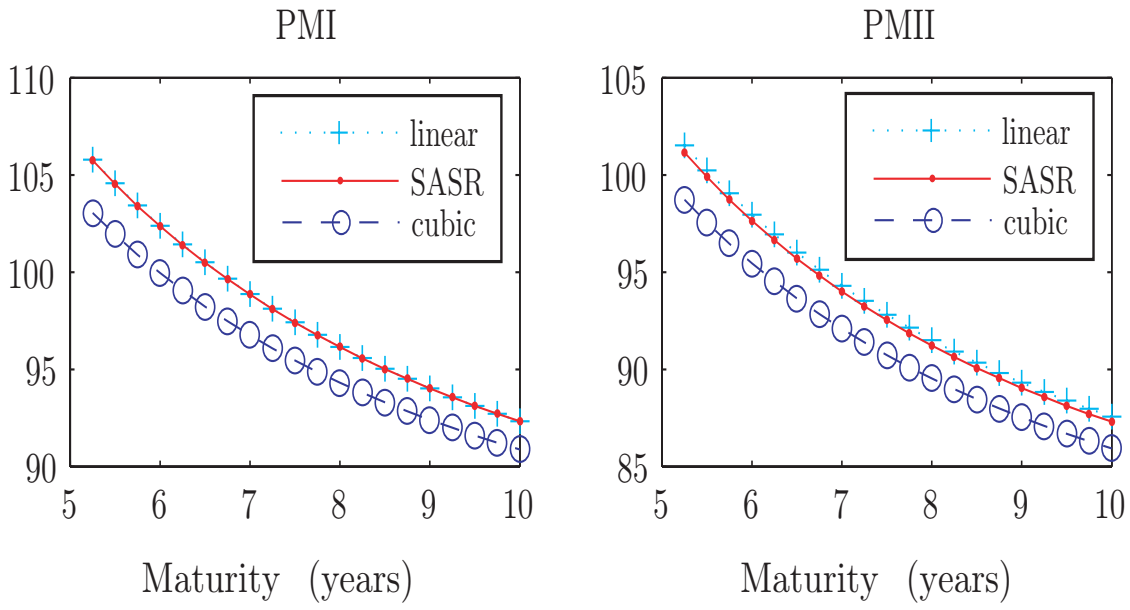

Figure 3: This figure displays the Root Mean Squared Errors (RMSE) computed within the sample period 1970:1 to 1995:12. The results for the family PMI are on the left and for PMII on the right.

various models examined. We explore this issue using the hold-out sample of the final five years of our data set. From the out-of-sample RMSEs displayed in Figure 4 for the family PMII, it is seen that the cubic models again perform the best (have the smallest RMSEs) for these six-month ahead forecasts. Most striking is the poor performance of the linear and SASR PMII models. In both cases, this relatively weak performance is overcome through the addition of a cubic term to the drift, as the RMSE of the cubic PMII model is comparable to that of the cubic PMI model.

Within the PMI family, the linear, SASR, and cubic models all perform similarly (in terms of out-of-sample RMSEs). Our interpretation of this contrasting performance of the PMI and PMII models is as follows. When fitting to our entire sample period, the inclusion of the cubic terms is important for accurately describing the joint distribution of bond yields. This is reflected in the highly significant estimates of λ_{Z3} documented in Table 4. At the same time, our hold-out sample, from the late 1990's, is a relatively quiet period by historical standards in terms of interest rate behavior. Therefore the cubic term in the market price of risk is not playing a key role and, within the PMI family, the linear, SASR, and cubic models all have comparable RMSEs out of sample.

However, if the benchmark model for the entire sample is obtained by fitting the data without the cubic term in Λ_t– as in the PMII family of models– then the other parameters

of results.

[26]We experimented with a larger number of simulations and the results were qualitatively the same.

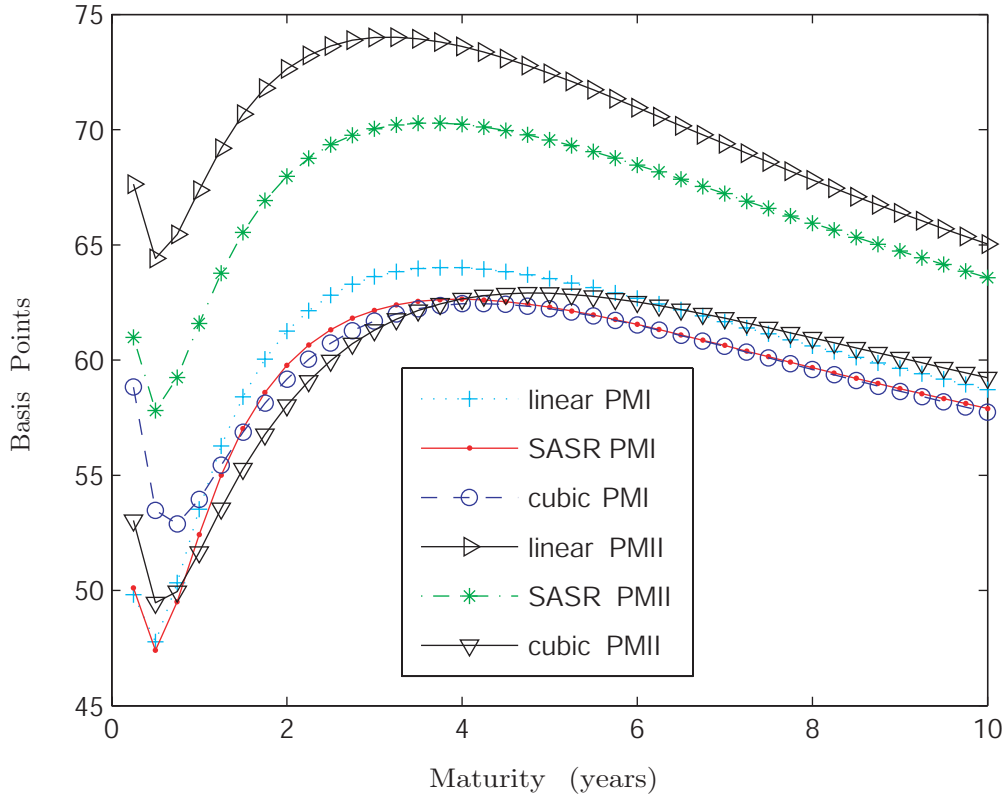

Figure 4: **Out of sample RMSE of pricing errors:** For each model, state variables are implied from the ML estimates of parameters and bond yield data. Pricing errors are the differences between forecasts and actual yields. RMSEs are computed for the hold out period from 1996:1 to 2000:12.

of Λ_t (and hence of the $\mathbb{P}$ distribution of the state X) must compensate for this omission. This model misspecification shows up in our out-of-sample analysis in the form of relatively large RMSEs for the linear and SASR models within the PMII family. Only when we put the cubic term in Λ_t back into the PMII model do we recover the same low level of RMSEs.

In sum, this out-of-sample analysis reinforces the view that the cubic term in the market price of risk of Z is important for accurately describing the historical behavior of U.S. Treasury bond yields. Returning to the shapes of the drifts of Z in Figure 2, these observations also suggest that the "S on its side" shape to the drift of Z in family PMI is a contributing factor to the better performance of this family on an out-of-sample basis.

The components of the RMSEs in Figure 4– the means and standard deviations of the pricing errors– are displayed in Figures 5 and 6. Though all models seem to overestimate yields on average, it can be seen that the PMI models provide better forecasts of the bond yields than the PMII models. Among the PMII models, there is a clear hierarchy of performance: the cubic fits best, followed by the SASR and then the linear models. Interestingly, the cubic PMI model does not fit the yields (on average) as well as the SASR or linear PMI

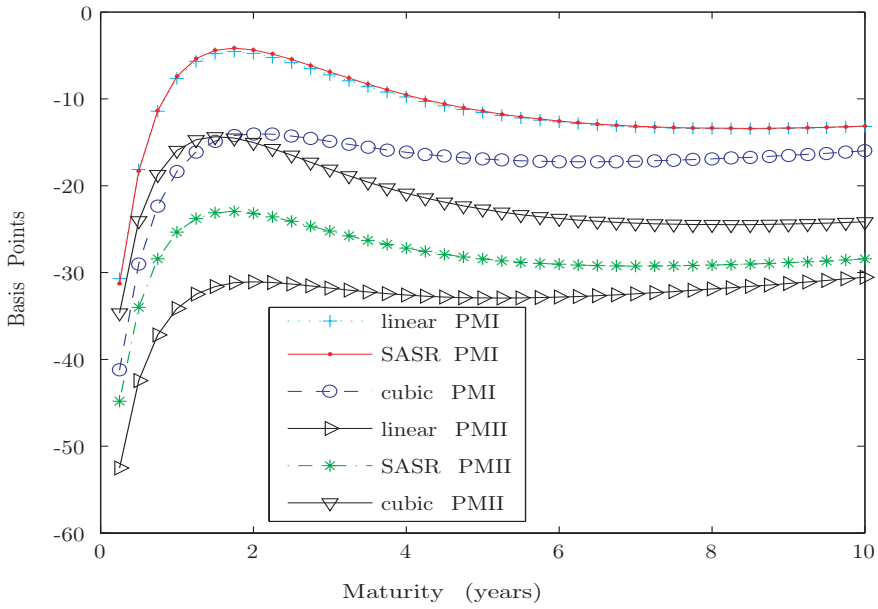

Figure 5: **Out of sample average pricing errors:** Pricing errors are differences between forecasts and actual yields. Average pricing errors are taken for the hold out period from 1996:1 to 2000:12.

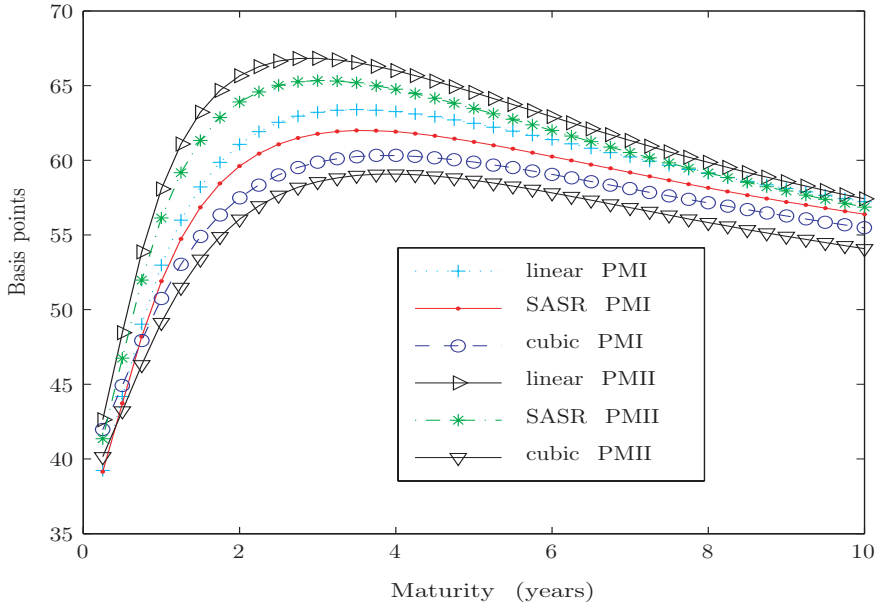

Figure 6: **Out of sample standard deviations of pricing errors:** Pricing errors are differences between forecasts and actual yields. Standard deviations of pricing errors are taken for the hold out period from 1996:1 to 2000:12.

models.

The two cubic models (from the PMI and PMII families) give rise to the smallest standard deviations of the out-of-sample pricing errors. The linear and SASR PMII models perform the worst by having the largest volatilities of pricing errors. In the middle are the linear and SASR PMI models. Once again this is consistent with our view that inclusion of the cubic terms in deriving the benchmark PMI family improves the out-of-sample performance of the models that exclude the cubic term. Put differently, allowing for the "S on its side" shape in the drift of Z appears to matter on an out-of-sample basis even during relatively quiet periods for interest rates.

7 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 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. An interesting topic for future research is a systematic analysis of the nature of nonlinearity in the $\mathbb{P}$ distributions of bond yields in multi-factor models.

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. 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. Ang and Bekaert (2003) and Dai, Singleton, and Wei (2003) 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. Monfort and Pegoraro (2006) propose several families of regime-switching, affine models based on Gaussian and autoregressive gamma models.

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*

$$Z_{t+1} = H(Z_t, \epsilon_{t+1}, \theta), \tag{50}$$

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*

$$\|H(z, \epsilon_1, \theta)\|_q < \delta_\theta \|z\|, \quad \|z\| > K. \tag{51}$$

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

$$\frac{\|H(z, \epsilon_1, \theta)\|_2}{\|z\|} \leq \frac{\|a^{(1)}(\lambda(z))\|}{\|z\|} + \frac{\|b^{(1)}(\lambda(z))z\|}{\|z\|} + \frac{\|\sqrt{\Omega(z)}\epsilon_1\|_2}{\|z\|}. \tag{52}$$

The first term on the right-hand-side of (52) satisfies

$$\frac{\|a^{(1)}(\lambda(z))\|}{\|z\|} = \frac{\left\|\text{vec}\left[\frac{\nu_i c_i}{1-\lambda_i(z)c_i}\right]\right\|}{\|z\|} \leq \frac{\|\text{vec}\left[\nu_i c_i\right]\|}{\|z\|} \to 0, \ \|z\| \to \infty, \tag{53}$$

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 (52) 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 (52) 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 $\sqrt{z}$.[27]

The only term on the right-hand side of (52) 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.

B Sampling Intervals Longer than Decision Intervals

Fortunately, within the context of *DTSM*s, there is an important subclass of discrete-time affine processes for which the joint density implied by (42) can be expressed analytically. Specifically, suppose that $f^{\mathbb{Q}}(Z_{t+1}|Z_t)$ is the density of a $DA^{\mathbb{Q}}_M(M)$ process for the M autoregressive gamma processes are mutually independent (the discrete-time counterpart to the case of M independent square-root diffusions). As shown by Gourieroux and Jasiak (2001), the conditional densities $f^{\mathbb{Q}}(Z^i_{t+\tau}|Z^i_t)$ are known in closed form and, indeed, are also those of scalar autoregressive gamma processes with

$$a_i(u) = -\nu_i \log\left(1 - u_i \tilde{c}(\tau, i)\right), \quad b_i(u) = \frac{u_i}{1 - u_i \tilde{c}(\tau, i)} \rho^\tau_{ii}, \tag{54}$$

where $\tilde{c}(\tau, i) \equiv c_i(1 - \rho^\tau_{ii})/(1 - \rho_{ii})$ and ρ_{ii} is the i^{th} diagonal element of the matrix ρ. Thus the joint conditional density function of Z can be constructed using the marginal densities

$$f^{\mathbb{Q}}(Z^i_{t+\tau}|Z^i_t) = \frac{1}{\tilde{c}(\tau, i)} \sum_{k=0}^{\infty} \left[\frac{\left(\frac{\rho^\tau_{ii} Z^i_t}{\tilde{c}(\tau,i)}\right)^k}{k!} e^{-\frac{\rho^\tau_{ii} Z^i_t}{\tilde{c}(\tau,i)}} \times \frac{\left(\frac{Z^i_{t+1}}{\tilde{c}(\tau,i)}\right)^{\nu_i + k - 1} e^{-\frac{Z^i_{t+1}}{\tilde{c}(\tau,i)}}}{\Gamma(\nu_i + k)} \right]. \tag{55}$$

To construct the conditional density $f^{\mathbb{Q}}(X_{t+\tau}|X_t)$ of an $DA^{\mathbb{Q}}_M(N)$ process, with $0 < M < N$, it remains to address the distribution of Y, the $N - M$ state variables whose conditional variance depends on the autoregressive gamma process Z. Just as with multi-period forecasting in *GARCH*-style models (Bollerslev (1986)), the conditional distribution of $Y_{t+\tau}$ given X_t does not remain Gaussian for $\tau > 1$. Following Duffie, Pedersen, and Singleton (2003), we can construct a very accurate approximation to the density $f^{\mathbb{Q}}(Y_{t+\tau}|X_t)$ by conditioning on both X_t and the sample path $(Z_t, Z_{t+1}, \ldots, Z_{t+\tau})$. Recalling the notation introduced below equation (8), we can write

$$Y_{t+\tau} = \left(\sum_{j=0}^{\tau-1} (\mu^Y_Y)^j \right) \mu_0 + \sum_{j=0}^{\tau-1} (\mu^Y_Y)^j \mu^Z_Y Z_{t+\tau-j} + \sum_{j=0}^{\tau-1} (\mu^Y_Y)^j \sqrt{S_{Y,t+j}} \epsilon_{t+\tau+1}, \tag{56}$$

where $(\mu^Y_Y)^0 = I_{N-M}$, $\epsilon_t \sim N(0, I_{N-M})$ and, consistent with our construction of canonical $DA^{\mathbb{Q}}_M(N)$ models, we have normalized $\Sigma_Y = I_{N-M}$. It follows that $f^{\mathbb{Q}}(Y_{t+\tau}|Z_{t+\tau}, \ldots, Z_{t+1}, X_t)$

[27] 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.

is the density of a normal random vector with conditional mean and variance that are easily derived from (56).

The insight of Duffie, et. al. is that, to a high degree of accuracy,

$$f^{\mathbb{Q}}(Y_{t+\tau}|Z_{t+\tau},\ldots,Z_{t+1},X_t) \approx \tilde{f}^{\mathbb{Q}}(Y_{t+\tau}|Z_{t+\tau},X_t), \tag{57}$$

where the right-hand density is obtained by replacing the Z_s $(t+1 \leq s < t+\tau)$ by linearly interpolated values between Z_t and $Z_{t+\tau}$. Accordingly, we can take as our approximation to $f^{\mathbb{Q}}(X_{t+1}|X_t)$ the density $f^{\mathbb{Q}}(Z_{t+\tau}|Z_t)\times\tilde{f}^{\mathbb{Q}}(Y_{t+\tau}|Z_{t+\tau},X_t)$. We stress that the first term in this construction, $f^{\mathbb{Q}}(Z_{t+\tau}|Z_t)$, is known exactly in closed form, so that the only approximation is with regard to the density $f^{\mathbb{Q}}(Y_{t+\tau}|Z_{t+\tau},\ldots,Z_{t+1},X_t)$. This approximation is very accurate because, for moderate τ, the non-normality induced by the time-varying volatility from Z builds up quite slowly with increasing τ (see Das and Sundaram (1999) and Singleton (2006)).[28]

With the conditional $\mathbb{Q}$ density $f^{\mathbb{Q}}(X_{t+\tau}|X_t)$ in hand, we turn next to the derivation of the $\mathbb{P}$ density $f^{\mathbb{P}}(X_{t+\tau}|X_t)$. If Λ_t is such that the $\mathbb{P}$ distribution of X_{t+1} given X_t is that of an affine process in the family $DA_M^{\mathbb{P}}(N)$, then the preceding discussion yields an accurate approximation to the conditional density of the risk factors X and, hence, of the likelihood function of bond yields. That is, for $\mathbb{P}$-affine processes, (approximate) *ML* estimation when the sampling interval exceeds the modeling interval is straightforward. Though this discussion preserves the assumption of linear (affine) models under both $\mathbb{P}$ and $\mathbb{Q}$, it is nevertheless of some practical importance because of its applicability to macro-term structure models. Using this approach, researchers can explore the sensitivity of integrated discrete-time term structure and structural macroeconomic models to the specification of the decision interval of agents relative to the sampling interval of the data.

Once nonlinearity is introduced through the choice of Λ_t, so that the $\mathbb{P}$ distribution of X is no longer affine, then we cannot in general derive a simple approximation to $f^{\mathbb{P}}(X_{t+\tau}|X_t)$ when $\tau > 1$.

[28] Duffie, Pedersen, and Singleton (2003) discuss ways of improving the accuracy of their approximation, but it involves some additional computational complexity due to the use of simulations.

References

Ahn, D.-H., Dittmar, R. F., Gallant, A. R., 2001. Quadratic term structure models: Theory and evidence. forthcoming, Review of Financial Studies.

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. Working Paper, Columbia University.

Ang, A., Piazzesi, M., 2002. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. forthcoming, *Journal of Monetary Economics*.

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 pp. 69–83.

Bekaert, G., Cho, S., Moreno, A., 2005. New-keynesian macroeconomics and the term structure. Unpublished working paper. Columbia University.

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

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

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.

Buhlmann, H., Delbaen, F., Embrechts, P., Shiryaev, A., 1996. No-arbitrage, change of measure and conditional esscher transforms. CWI Quarterly 9, 291–317.

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.

Cheng, P., Scaillet, O., 2002. Linear-quadratic jump-diffusion modelling with application to stochastic volatility. .

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 Working Paper 10756.

Cox, J., Ingersoll, J., Ross, S., 1985. A theory of the term structure of interest rates. Econometrica 53(2), 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 LV, 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. J., Wei, Y., 2003. Is regime shift risk priced in the u.s. treasury market?. Working paper, New York University and Stanford University.

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

Das, S., Sundaram, R., 1999. Of smiles and smirks: A term structure perspective. Journal of Financial and Quantitative Analysis 34, 211–239.

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

Duarte, J., 2003. Evaluating alternative risk preferences in affine term structure models. Review of Financial Studies, forthcoming.

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.

Esscher, F., 1932. On the probability function in the collective theory of risk. Skandinavisk Aktuarietidskrift 15, 175–195.

Evans, M. D., 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.

Gerber, H., Shiu, E., 1994. Option pricing by esscher transforms. Transactions of Society of Actuaries 46, 99–140.

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

Gourieroux, C., Monfort, A., Polimenis, V., 2002. Affine term structure models. .

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., 1988. Volatility and the yield curve. Unpublished working paper. Goldman Sachs.

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

Monfort, A., Pegoraro, F., 2006. Switching varma term structure models. CREST.

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.

——— , 2006. Empirical Asset Pricing Models: Specification and Empirical Assessment. Princeton University Press, .

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.

Wu, T., 2000. Macro factors and the affine term structure of interest rates. Working Paper, Yale University.