

# Discrete-time Affine<sup>Q</sup> Term Structure Models with Generalized Market Prices of Risk

Anh Le, Kenneth J. Singleton, and Qiang Dai <sup>1</sup>

This draft: February 8, 2009

<sup>1</sup>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. Dai is with Capula Investment Management LLP.

## Abstract

This paper develops a rich class of discrete-time, *nonlinear* dynamic term structure models (*DTSMs*). Under the risk-neutral measure  $\mathbb{Q}$ , 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  $\Lambda_t$ , linking the risk-neutral and historical distributions of  $X$ , to depend generally on the state  $X_t$ . The conditional likelihood functions for zero-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 develop an equilibrium, nonlinear term structure model in which agents exhibit habit formation. Though nonlinear, by design this model shares many of the features of habit-based models in the literature. Moreover, zero-coupon bond prices and the conditional likelihood function of bond yields, consumption growth, and inflation are known in closed form. When evaluated at the maximum likelihood estimates of the parameters, our habit-based model is not able to match key features of the conditional distribution of bond yields.

# 1 Introduction

This paper develops a rich class of discrete-time, *nonlinear* dynamic term structure models (*DTSMs*) in which zero-coupon bond yields and their conditional densities are known exactly in closed form. Under the risk-neutral measure  $\mathbb{Q}$ , the distribution of the state vector  $X_t$  resides within a family of discrete-time affine <sup>$\mathbb{Q}$</sup>  processes<sup>1</sup> 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).<sup>2</sup> Moreover, we allow the market price of risk  $\Lambda_t$ , linking the  $\mathbb{Q}$  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  $\Lambda_t$  leads to a family of *DTSMs* in which the conditional  $\mathbb{P}$ -distributions of  $X_{t+1}$  and bond yields can show very rich *nonlinear* dependence on  $X_t$ .

While this leads immediately to a much richer family of arbitrage-free, affine <sup>$\mathbb{Q}$</sup>  *DTSMs* than has heretofore been implemented econometrically,<sup>3</sup> the primary motivation for this paper derives from the growing literature on equilibrium macro-finance models of the term structure. In particular, the literature on integrating *DTSMs* with linearized neo-Keynesian (“IS-LM” style) macroeconomic models (e.g., Rudebusch and Wu (2008), Hordahl, Tristani, and Vestin (2007), Wu (2005), and Bekaert, Cho, and Moreno (2006)) has focused exclusively on discrete-time Gaussian *DTSMs*.<sup>4</sup> Arbitrage-free *DTSMs* are overlaid onto log-linear macro models with Gaussian, homoskedastic shocks.

Concurrently, there is a growing literature exploring the ability of preference-based, equilibrium *DTSMs* to resolve various empirical asset pricing puzzles. Campbell and Cochrane (1999) and Wachter (2005), for instance, develop *DTSMs* in which agents’ preferences exhibit external habit formation. Alternatively, Bansal and Shaliastovich (2007) and Wu (2008) examine the properties of *DTSMs* in which agents exhibit preferences for the early resolution of uncertainty and face “long-run” risks in their consumption streams. To date, most of these models have been evaluated using calibrated parameters rather than at estimates from the model-implied likelihood functions.

The focus on Gaussian models in the macro-finance literature appears to be driven largely by the absence of tractable discrete-time, multi-factor *DTSMs* with flexible market prices

---

<sup>1</sup> We use the notation *affine* <sup>$\mathbb{Q}$</sup>  to denote processes that are affine under the risk neutral measure  $\mathbb{Q}$ .

<sup>2</sup>Our analysis extends immediately to the case of quadratic-Gaussian models discussed in Beaglehole and Tenney (1991), Ahn, Dittmar, and Gallant (2002) 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.

<sup>3</sup> With few exceptions, econometric specifications under  $\mathbb{P}$  of continuous-time, affine <sup>$\mathbb{Q}$</sup>  *DTSMs* have chosen market prices of risk that preserve the affine structure under  $\mathbb{P}$  (see, e.g., Dai and Singleton (2000), Duffee (2002), and Cheridito, Filipovic, and Kimmel (2005)). In discrete-time, most of the empirical literature has focused on the even more restrictive case of  $\mathbb{P}$  and  $\mathbb{Q}$  Gaussian models. Ang and Piazzesi (2003) and Ang, Dong, and Piazzesi (2007) are examples of studies focusing on monetary policy, while Dai and Philippon (2005) examine fiscal policy within a Gaussian *DTSM*.

<sup>4</sup>Often, agent’s preferences do not appear explicitly in these models, but they are implicit in the specification of the aggregate demand or “IS” function.

of risk and stochastic volatility. The use of calibration methods rather than likelihood-based estimators in the preference-based literature has been influenced, no doubt in part, by the computational burden associated with the absence of close-form solutions for bond prices. Our proposed framework explicitly addresses both of these issues.

Moreover, we overcome many of the challenges with estimation in the literature on continuous-time diffusions. 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.<sup>5</sup> 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 *DTSMs* classified by *DS*. Therefore, no approximations are necessary in estimation.

To illustrate our modeling strategy we develop a habit-based model of the term structure of interest rates, starting from the pricing kernel examined in Campbell and Cochrane (1999) (hereafter *CC*), Wachter (2005), and Verdelhan (2008). These authors posit affine representations of the state under  $\mathbb{P}$  which, when combined with the habit-based pricing kernel, lead to nonlinear expressions for bond prices that must be solved numerically. Moreover, likelihood functions for the data are not known in closed form. Instead, we assume that  $X_t$  follows an affine<sup>Q</sup> process of a form that embeds all of the key features of extant models with habit formation, including time-varying volatility of the surplus consumption ratio  $S_t$ , nonzero correlation between this ratio and inflation  $\pi_t$ , and an implied persistence in consumption growth.

The market prices of risk associated with our habit-based *DTSM* and state process give rise to a nonlinear (non-affine) representation of bond yields under the historical distribution. Nevertheless, we show that, by appropriate choice of the consumption growth process, an *equilibrium* implication of our model is that the short rate is an affine function of the state. Consequently, zero-coupon bond yields are affine functions of the state. Moreover, the likelihood function of the data is known in closed form. *CC*, Wachter, and Verdelhan calibrated their model to selected sets of parameters. Others have used *GMM* methods to estimate equilibrium models off Euler equations; see, for example, Fuhrer (2000) and Engsted and Moller (2008) for habit-based models, and Bansal, Kiku, and Yaron (2007) and Constantinides and Ghosh (2008) for models with long-run risks in consumption growth. Our framework renders full-information maximum likelihood feasible for these (and other) equilibrium asset pricing models.

We proceed to compute *ML* estimates of our habit-based model using historical data on consumption growth, inflation, and U.S. Treasury bond yields. We compare our estimates,

---

<sup>5</sup>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).

and the model-implied properties of the conditional distribution of bond yields, to those implied by parameters chosen according to several sensible calibration schemes. The results highlight some of the limitations of the habit-based models that have been examined to date.

In what is perhaps the closest precursor to our construction of arbitrage-free pricing models, Gouriéroux, Monfort, and Polimenis (2002) developed *DTSMs* 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  $DA_M^{\mathbb{Q}}(N)$ ,  $M = 0, 1, \dots, N$ . Furthermore, Gouriéroux, et. al. assumed that the market price of risk  $\Lambda$  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 *DTSMs* that allow general dependence of  $\Lambda_t$  on  $X_t$ . Moreover, we illustrate this flexibility by computing *ML* estimates of an equilibrium asset pricing model using both macroeconomic and bond market data.

The remainder of this paper is organized as follows. We start in Section 2 with a more in depth motivation for our modeling framework using the habit-based asset pricing model introduced by Campbell and Cochrane (1999). We then proceed to develop both the theoretical properties of our modeling approach and their application to a habit-based *DTSM* in parallel. Section 3 presents the canonical families of affine<sup>Q</sup> processes  $DA_M^{\mathbb{Q}}(N)$ ,  $0 \leq M \leq N$ . The specific formulations of the state process in the habit-based model are set forth in Section 4, and closed-form expressions for equilibrium bond prices are derived. In the process, we also demonstrate that, as an equilibrium implication of our formulation, the short rate is an affine function of the surplus consumption ratio and inflation rate.

The distribution of bond yields under the physical measure is taken up in Section 5. 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 *DTSMs*. 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,  $\Lambda_t$ , on  $X_t$ . By roaming over admissible choices of  $\Lambda_t$ , we are effectively ranging across the entire family of admissible arbitrage-free *DTSMs* constructed under the assumption that, under  $\mathbb{Q}$ ,  $X$  follows a process residing in one of the families  $DA_M^{\mathbb{Q}}(N)$ . 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 affine<sup>Q</sup> distribution of the state  $X$ , leads to a known parametric representation of the  $\mathbb{P}$ -distribution of bond yields.

The properties of the market prices of risk underlying our choice of  $(d\mathbb{P}/d\mathbb{Q})_{t+1}^D$  are elaborated on in Section 6. Details of the  $\mathbb{P}$  distribution of the state and the associated market prices of risk in our habit-based, illustrative *DTSM* are presented in Section 7. Finally, the empirical examples are presented in Section 8.

## 2 An Illustrative Model with Habit Formation

Following *CC* and Wachter (2005), we assume that agents maximize the utility function:

$$E \sum_{t=0}^{\infty} \delta^t \frac{(C_t - H_t)^{1-\gamma} - 1}{1-\gamma}, \quad (1)$$

where  $H_t$  is the level of habit,  $\delta$  is the subjective discount factor and  $\gamma$  is the utility curvature parameter. The *consumption surplus ratio* is defined as  $S_t = (C_t - H_t)/C_t$ . For any asset with nominal (total) return  $R_{t+1}$ , this leads to the Euler equation

$$E_t \left[ \delta \left( \frac{S_{t+1}}{S_t} \right)^{-\gamma} \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma} \frac{P_t}{P_{t+1}} R_{t+1} \right] = 1, \quad (2)$$

where  $P_t$  is the price level in this economy at date  $t$ .

The nominal pricing kernel, in natural logarithm, can be written as:

$$m_{t,t+1} = \log \delta - \gamma(s_{t+1} - s_t) - \gamma g_{t+1} - \pi_{t+1} \quad (3)$$

where  $s_t = \log(S_t)$ ,  $g_{t+1} = \log(C_{t+1}/C_t)$  and  $\pi_{t+1} = \log(P_{t+1}/P_t)$ . As in *CC* and Wachter (2005), we assume that the state vector  $X_t$  is comprised of the current consumption surplus ratio,  $s_t$ , and current inflation rate,  $\pi_t$ . Consumption growth,  $g_t$ , is assumed to be conditionally perfectly correlated with the consumption surplus ratio. Finally, the upper bound of  $s_t$  is captured by a free parameter,  $s_{max}$ .<sup>6</sup> We let  $z_t = s_{max} - s_t$  denote the inverse consumption surplus ratio.

Several practical issues arise when flushing out an implementable version of this model with habit formation. First,  $z$  is a strictly positive process and, therefore, innovations in consumption growth (equivalently,  $z_t$ ) cannot literally be Gaussian as assumed by Wachter. We circumvent this inconsistency by directly positing a strictly positive, discrete-time stochastic process for the inverse consumption ratio. As in *CC* and Wachter, our representation of  $z_t$  also exhibits conditional heteroskedasticity.

Additionally, in illustrative models of habit formation, researchers have typically assumed that consumption growth follows an affine process under  $\mathbb{P}$ ; *CC* and Wachter, for example, assume that  $g_t$  is an *i.i.d.* process (consumption follows a random walk with drift). This, together with their model for  $s_t$ , implies that bond prices must be determined numerically from the representative agent's Euler equation. Instead, we formulate our model so that  $X_t$  follows an affine<sup>Q</sup> process and the one-period short-rate  $r_t$  is an affine function of  $X_t$ , and this leads immediately to closed-form solutions for zero-coupon bond prices (Duffie and Kan (1996)). Then market prices of risk are chosen so that the  $\mathbb{P}$  distribution of consumption growth shares many of the features of previous specifications. In fact, in the continuous-time

---

<sup>6</sup>Note that  $s_t$  is always negative therefore a natural (trivial) upper bound for  $s_t$  is  $s_{max} = 0$ .  $s_t = 0$  implies a zero habit level:  $H_t = 0$ . Therefore a non-zero upper bound of  $s_t$  essentially imposes a minimum level of habit,  $H_t$ , as a fraction of current consumption,  $C_t$ .

limit of our discrete-time model,  $g_t$  is *i.i.d.* and conditionally homoskedastic, just as in *CC* and Wachter.

In this equilibrium setting, the functional dependence of  $r_t$  on  $X_t$  depends on the structure of preferences and the specifications of the  $\mathbb{P}$  and  $\mathbb{Q}$  distributions of the state. As part of the development of our pricing model with habit formation, we demonstrate that the affine dependence of  $r_t$  on  $X_t$  is an equilibrium *implication* of the model. This is achieved by judicious choice of the drift of  $g_t$  under  $\mathbb{Q}$ .

We expand on these and related issues subsequent to presenting the affine <sup>$\mathbb{Q}$</sup>  family of models used in our empirical illustrations.

### 3 Canonical Discrete-Time Affine <sup>$\mathbb{Q}$</sup> Processes

Following Duffie, Filipovic, and Schachermayer (2003), we will refer to a Markov process  $X$  as *affine* <sup>$\mathbb{Q}$</sup>  if the conditional Laplace transforms of  $X_{t+1}$  given  $X_t$  is an exponential-affine function of  $X_t$ :<sup>7</sup> under a probability measure  $\mathbb{Q}$ , for an  $N \times 1$  state vector  $X$ ,

$$\phi^{\mathbb{Q}}(u; X_t) = E^{\mathbb{Q}} \left[ e^{u'X_{t+1}} \mid X_t \right] = e^{a(u)+b(u)X_t}. \quad (4)$$

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  $A_M^{\mathbb{Q}}(N)$ .<sup>8</sup> 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 (4) 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  $DA_M^{\mathbb{Q}}(N)$ ,  $M = 0, \dots, N$ .

#### 3.1 $DA_0^{\mathbb{Q}}(N)$

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 (4) with

$$a(u) = \mu_0' u + \frac{1}{2} u' V u, \quad b(u) = u' \mu_Y. \quad (5)$$

---

<sup>7</sup>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 Gouriéroux and Jasiak (2006) and Darolles, Gouriéroux, and Jasiak (2006) for discussions of discrete-time affine processes related to those examined in this paper.

<sup>8</sup>These are not the only well-defined discrete-time affine *DTSMs*. Gouriéroux, 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  $A_M^{\mathbb{Q}}(N)$ ).

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

### 3.2 $DA_N^{\mathbb{Q}}(N)$

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 (2006) and Darolles, Gourieroux, and Jasiak (2006). 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  $\varrho$  be an  $N \times N$  matrix with elements satisfying

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

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

$$\frac{Z_{t+1}^i}{c_i} | (\mathcal{P}, Z_t) \sim \text{gamma}(\nu_i + \mathcal{P}), \quad \text{where } \mathcal{P} | Z_t \sim \text{Poisson}(\rho_i Z_t / c_i). \quad (6)$$

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

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

$$f^{\mathbb{Q}}(Z_{t+1}^i | Z_t) = \frac{1}{c_i} \sum_{k=0}^{\infty} \left[ \frac{\left(\frac{\rho_i Z_t}{c_i}\right)^k}{k!} e^{-\frac{\rho_i Z_t}{c_i}} \times \frac{\left(\frac{Z_{t+1}^i}{c_i}\right)^{\nu_i+k-1} e^{-\frac{Z_{t+1}^i}{c_i}}}{\Gamma(\nu_i + k)} \right]. \quad (7)$$

Using conditional independence, the distribution of a  $DA_N^{\mathbb{Q}}(N)$  process  $Z_{t+1}$ , conditional on  $Z_t$ , is given by  $f^{\mathbb{Q}}(Z_{t+1} | Z_t) = \prod_{i=1}^N f^{\mathbb{Q}}(Z_{t+1}^i | Z_t)$ . Finally, it is straight-forward to show that for any  $u$ , such that  $u_i < \frac{1}{c_i}$ , the conditional Laplace transform of  $Z_{t+1}$  is given by (4) with

$$a(u) = - \sum_{i=1}^N \nu_i \log(1 - u_i c_i), \quad b(u) = \sum_{i=1}^N \frac{u_i}{1 - u_i c_i} \rho_i. \quad (8)$$

When the off-diagonal elements of the  $N \times N$  matrix  $\rho$  are non-zero, the autoregressive gamma processes  $\{Z^i\}$  are (unconditionally) correlated. Thus, even in the case of correlated  $Z_t^i$ , the conditional density of  $Z_{t+1}$  is known in closed form. This is not the case for correlated  $Z$  in the continuous-time family  $A_N^{\mathbb{Q}}(N)$ . The nature of the correlation between  $Z^i$  and  $Z^j$  ( $i \neq j$ ) is constrained by our requirement that  $\rho_{ij} \leq 0$ . 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  $E_t^{\mathbb{Q}}[Z_{t+1}]$  and conditional covariance matrix  $V_t^{\mathbb{Q}}[Z_{t+1}]$  implied by the conditional moment-generating function (4) and (8) are

$$E_t^{\mathbb{Q}}[Z_{t+1}](i) = \nu_i c_i + \rho_i Z_t, \quad V_t^{\mathbb{Q}}[Z_{t+1}](i, i) = \nu_i c_i^2 + 2c_i \rho_i Z_t, \quad (9)$$

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

That this process converges to the multi-factor correlated  $A_N^{\mathbb{Q}}(N)$  process<sup>9</sup> can be seen by letting  $\rho = I_{N \times N} - \kappa^{\mathbb{Q}} \Delta t$ ,  $c_i = \frac{\sigma_i^2}{2} \Delta t$ , and  $\nu_i = \frac{2(\kappa^{\mathbb{Q}} \theta^{\mathbb{Q}})_i}{\sigma_i^2}$ , where  $\kappa^{\mathbb{Q}}$  is a  $N \times N$  matrix and  $\theta^{\mathbb{Q}}$  is a  $N \times 1$  vector. In the limit as  $\Delta t \rightarrow 0$ , the  $DA_N^{\mathbb{Q}}(N)$  process converges to:

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

where  $\sigma$  is a  $N \times N$  diagonal matrix with  $i^{\text{th}}$  diagonal element given by  $\sigma_i$ .

### 3.2.1 $DA_M^{\mathbb{Q}}(N)$ 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) the Laplace transform of

$$f^{\mathbb{Q}}(X_{t+1}|X_t) = f^{\mathbb{Q}}(Y_{t+1}|Z_{t+1}, Y_t, Z_t) \times f^{\mathbb{Q}}(Z_{t+1}|Z_t), \quad (10)$$

is an exponential-affine function of  $X_t$ .

This will be the case if  $Y_{t+1}$  is exponentially affine with respect to  $(Z_{t+1}, Y_t, Z_t)$ .<sup>10</sup> For example, if  $f^{\mathbb{Q}}(Y_{t+1}|Z_{t+1}, Y_t, Z_t)$  is the density of a Gaussian process with conditional mean and variance  $\omega_{Y_t}^{\mathbb{Q}} \equiv \mu_0 + \mu_Z Z_{t+1} + \mu_Y Y_t$  and  $\Omega_{Y_t} \equiv \Sigma_Y S_{Y_t} \Sigma_Y'$ , where  $\Sigma_Y$  is an  $(N - M) \times (N - M)$  matrix, and  $S_{Y_t}$  is a  $(N - M) \times (N - M)$  diagonal matrix with the  $i^{\text{th}}$  diagonal

<sup>9</sup>Gourieroux and Jasiak (2006) attribute the insight that the  $DA_1^{\mathbb{Q}}(1)$  process is a discrete-time counterpart to the square-root diffusion to Lambertson and Lapeyre (1992).

<sup>10</sup>To see this, consider:  $E[e^{u_Y Y_{t+1} + u_Z Z_{t+1}} | Y_t, Z_t] = E[E[e^{u_Y Y_{t+1}} | Z_{t+1}, Y_t, Z_t] e^{u_Z Z_{t+1}} | Y_t, Z_t]$ . If  $Y_{t+1}$  is exponentially affine with respect to  $(Z_{t+1}, Y_t, Z_t)$  then  $E[e^{u_Y Y_{t+1}} | Z_{t+1}, Y_t, Z_t] = e^{a_Z Z_{t+1} + b_Y Y_t + c_Z Z_t}$  which implies  $E[e^{u_Y Y_{t+1} + u_Z Z_{t+1}} | Y_t, Z_t] = E[e^{(a_Z + u_Z) Z_{t+1}} | Z_t] e^{b_Y Y_t + c_Z Z_t}$  which is exponential-affine in  $X_t$ .

element given by  $\alpha_i + \beta'_i Z_t$ ,  $1 \leq i \leq N - M$ .<sup>11</sup> The proposed formulation of a habit-based *DTSM* follows this structure. We will assume that the inverse surplus consumption ratio  $z_{t+1}$  follows a  $DA_1^{\mathbb{Q}}(1)$  process and inflation  $\pi_{t+1}$  is Gaussian conditional on  $(z_{t+1}, X_t)$ , and this will be shown to imply that  $X_t$  follows an affine<sup>Q</sup> process.

### 3.3 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 an affine function of the state:  $r_t = \delta_0 + \delta_X X_t$ , where  $\delta_X > 0$  is a  $1 \times N$  vector.<sup>12</sup>

With this additional assumption, the time- $t$  zero-coupon bond price with maturity of  $n$  periods is given by

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

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

$$\begin{aligned} A_n - A_{n-1} &= \delta_0 + A_{n-1} - a(-B_{n-1}), \\ B_n &= \delta_X - b(-B_{n-1}), \end{aligned} \quad (12)$$

with the initial condition  $A_0 = B_0 = 0$ .<sup>13</sup>

## 4 Pricing in the Habit-Based *DTSM*

In this section we apply the framework just presented to the pricing of nominal zero-coupon bonds in the habit-based *DTSM*. We proceed in three steps: first we present the risk-neutral, affine<sup>Q</sup> representation of the state; then we show that, by appropriate choice of the drift of consumption growth, in equilibrium the short-rate  $r_t$  is affine in  $X_t$ ; and finally we combine these results to derive close-form expressions for zero-coupon bond prices.

### 4.1 Risk-Neutral Representation of the State

*The inverse consumption surplus ratio:*

<sup>11</sup>For continuous-time formulations, Collin-Dufresne, Goldstein, and Jones (2008) and Joslin (2007) show that, when  $N \geq 4$  and  $2 \leq M \leq N - 2$ , then this formulation of the conditional variance is not the maximal canonical  $A_M^{\mathbb{Q}}(N)$  model. Our framework accommodates the discrete-time counterpart to their maximal models by appropriate choice of  $\Omega_{Y_t}$ .

<sup>12</sup>If  $X_t$  is a  $DA_M^{\mathbb{Q}}(N)$  process, then setting  $\delta_{X_i} > 0$  for  $i > M$  is a normalization, but setting  $\delta_{X_i} > 0$  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 13 for further elaboration on the second point.

<sup>13</sup>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.

Since the inverse surplus consumption ratio  $z_t$  is strictly positive, it is natural to model  $z_t$  as a  $DA_1^{\mathbb{Q}}(1)$  process. That is:

$$\frac{z_{t+1}}{c_z} | (\mathcal{P}, z_t) \sim \text{gamma}(\nu_z + \mathcal{P}), \quad \text{and} \quad \mathcal{P} | z_t \sim \text{Poisson}\left(\frac{\rho_z z_t}{c_z}\right). \quad (14)$$

The first two conditional moments of  $z_t$  are:

$$E_t^{\mathbb{Q}}[z_{t+1}] = \rho_z z_t + \nu_z c_z \quad (15)$$

$$\sigma_t^{\mathbb{Q}}[z_{t+1}]^2 = 2\rho_z c_z z_t + \nu_z c_z^2 \quad (16)$$

*The consumption growth:*

We assume that under  $\mathbb{Q}$  consumption growth follows the process

$$g_{t+1} = f(z_t) - \sigma_g \frac{z_{t+1} - E_t^{\mathbb{Q}}[z_{t+1}]}{\sigma_t^{\mathbb{Q}}[z_{t+1}]}. \quad (17)$$

The innovation in  $g_{t+1}$ ,  $z_{t+1} - E_t^{\mathbb{Q}}[z_{t+1}]$ , is the shock to  $s_{t+1}$ ; that is,  $s_{t+1}$  and  $z_{t+1}$  are perfectly correlated conditional on date  $t$  information. The scaling by  $\sigma_t^{\mathbb{Q}}[z_{t+1}]$  renders  $g_{t+1}$  approximately conditionally homoskedastic, an assumption maintained in both *CC* and Wachter.<sup>14</sup>

The conditional mean of consumption growth,  $f(z_t)$ , will be chosen subsequently to ensure that, in equilibrium, the short rate  $r_t$  is an affine function of the state.

*The inflation process:*

Following Wachter we assume that inflation has no impact on the real side of the economy. However, we do allow both for nonzero correlation between the innovations in  $\pi_{t+1}$  and  $z_{t+1}$ , as well as feedback from the real side of the economy to inflation ( $z_t$  affects  $\pi_{t+1}$ ):

$$\pi_{t+1} = \bar{\pi} + \rho_{\pi}(\pi_t - \bar{\pi}) + \rho_{\pi,z}(z_t - E^{\mathbb{Q}}[z_t]) - \sigma_{\pi,g}(z_{t+1} - E_t^{\mathbb{Q}}[z_{t+1}]) + \sigma_{\pi}\epsilon_{\pi,t+1}^{\mathbb{Q}}, \quad (18)$$

where  $\epsilon_{\pi,t+1}^{\mathbb{Q}} \sim N(0, 1)$  and the risk-neutral long run  $\mathbb{Q}$ -mean of  $z_t$  is  $\nu_z c_z / (1 - \rho_z)$ . The parameters  $\rho_{\pi}$  and  $\sigma_{\pi}$  govern the autoregressive nature of inflation and idiosyncratic inflation shocks, respectively. The parameters  $\rho_{\pi,z}$  and  $\sigma_{\pi,g}$  modulate the unconditional and conditional correlation between consumption growth and inflation.

*Risk-neutral density of states:*

In the notation of the last section,  $X_t$  follows a  $DA_1^{\mathbb{Q}}(2)$  with one-period ahead density

$$f^{\mathbb{Q}}(z_{t+1}, \pi_{t+1} | z_t, \pi_t) = f^{\mathbb{Q}}(z_{t+1} | z_t) \times f^{\mathbb{Q}}(\pi_{t+1} | z_{t+1}, z_t, \pi_t) \quad (19)$$

where  $f^{\mathbb{Q}}(z_{t+1} | z_t)$  is given by equation (7) and  $f^{\mathbb{Q}}(\pi_{t+1} | z_{t+1}, z_t, \pi_t)$  is a Gaussian density with

$$E_t^{\mathbb{Q}}[\pi_{t+1} | z_{t+1}, z_t, \pi_t] = \bar{\pi} + \rho_{\pi}(\pi_t - \bar{\pi}) + \rho_{\pi,z}(z_t - E^{\mathbb{Q}}[z_t]) - \sigma_{\pi,g}(z_{t+1} - E_t^{\mathbb{Q}}[z_{t+1}]) \quad (20)$$

$$\sigma_t^{\mathbb{Q}}[\pi_{t+1} | z_{t+1}, z_t, \pi_t] = \sigma_{\pi}. \quad (21)$$

---

<sup>14</sup> $g_{t+1}$  is exactly conditionally homoskedastic if  $\sigma_t^{\mathbb{Q}}[s_{t+1}] = \sigma_t^{\mathbb{P}}[s_{t+1}]$ . However, it can be shown that the difference between these two quantities is small for typical sampling intervals  $\Delta$ , on the order of  $(\Delta)^2$ .

Given this structure, it follows immediately that  $X$  is an affine<sup>Q</sup> process with Laplace transform

$$\phi^{\mathbb{Q}}(u; [z_t, \pi_t]) = E_t^{\mathbb{Q}}[e^{u_z z_{t+1} + u_\pi \pi_{t+1}}] = e^{a(u) + b_z(u)z_t + b_\pi(u)\pi_t}, \quad (22)$$

where:

$$a(u) = u_\pi \left( \bar{\pi}(1 - \rho_\pi) - \rho_{\pi,z} \frac{\nu_z c_z}{1 - \rho_z} + \sigma_{\pi,g} \nu_z c_z \right) + \frac{1}{2} \sigma_\pi^2 u_\pi^2 - \nu_z \log(1 - (u_z - u_\pi \sigma_{\pi,g})c_z) \quad (23)$$

and

$$b_z(u) = u_\pi (\rho_{\pi,z} + \sigma_{\pi,g} \rho_z) + \frac{\rho_z (u_z - u_\pi \sigma_{\pi,g})}{1 - (u_z - u_\pi \sigma_{\pi,g})c_z} \quad (24)$$

$$b_\pi(u) = u_\pi \rho_\pi. \quad (25)$$

## 4.2 Bond Prices in the Habit-Based DTSM

Key to obtaining closed-form representations of bond prices are the conditions that  $X_t$  follows an affine<sup>Q</sup> process and  $r_t$  is an affine function of  $X_t$ . The former property of the model is introduced by assumption on the exogenous variables in the model. We turn next to a sufficient set of restrictions on the risk-neutral expectation of consumption growth to ensure that the model-implied, equilibrium short rate is an affine function of  $X_t$ .

**Proposition 1** *If the conditional expectation of  $g_{t+1}$  under  $\mathbb{Q}$  is given by:*

$$f(z_t) = \mathcal{C} - (\gamma + \sigma_{\pi,g}) \sigma_g \sigma_t^{\mathbb{Q}}[z_{t+1}] - \frac{1}{\gamma} \log \left( \frac{E_t^{\mathbb{Q}}[e^{u_\Lambda z_{t+1}}]}{E_t^{\mathbb{Q}^G}[e^{u_\Lambda z_{t+1}}]} \right) \quad (26)$$

where

- $\mathcal{C}$  is a constant
- $u_\Lambda = -\gamma \left( 1 + \frac{\sigma_g}{\sigma_t^{\mathbb{Q}}[z_{t+1}]} \right) - \sigma_{\pi,g}$
- $\mathbb{Q}^G$  denotes a Gaussian measure with the same conditional mean and variance implied by the measure  $\mathbb{Q}$ ,

then the nominal interest rate per unit of time interval is affine in the state:

$$r_t = \delta_0 + \delta_z z_t + \delta_\pi \pi_t \quad (27)$$

where  $\delta_\pi = \rho_\pi$ ,<sup>15</sup>

$$\begin{aligned} \delta_0 &= -\log \delta + (1 - \rho_\pi) \bar{\pi} - \rho_{\pi,z} \frac{\nu_z c_z}{1 - \rho_z} - \gamma \nu_z c_z + \gamma \mathcal{C} \\ &\quad + \frac{1}{2} \gamma^2 \sigma_g^2 + \frac{1}{2} (\gamma + \sigma_{\pi,g})^2 \nu_z c_z^2 + \frac{1}{2} \sigma_\pi^2, \end{aligned} \quad (28)$$

$$\delta_z = \gamma(1 - \rho_z) + \rho_{\pi,z} + (\gamma + \sigma_{\pi,z})^2 \rho_z c_z. \quad (29)$$

---

<sup>15</sup>At first glance, the fact that  $r_t$  increases in  $\sigma_g^2$  and  $\sigma_\pi^2$  might seem contrary to investors' pre-cautionary savings motive. However, this is a consequence of representing  $\delta_0$  and  $\delta_z$  in terms of parameters of the risk-neutral distribution. If the risk-neutral mean is replaced by its equivalent expression in terms of the physical mean and market prices of risk, then we recover the usual negative coefficients on volatilities.

**Proof:** See Appendix D.

We defer further interpretation of the nonlinear conditional  $\mathbb{Q}$ -mean  $f(z_t)$  of consumption growth until after we have specified the market prices of risk. This will allow direct comparisons between the model-implied  $\mathbb{P}$  and  $\mathbb{Q}$  distributions of surplus consumption and consumption growth.

From Proposition 1 and our assumption that the states follow a  $DA_1^{\mathbb{Q}}(2)$  process, it follows that nominal zero-coupon bond prices of any maturity are exponentially affine in the state.

## 5 Physical Distribution of Bond Yields

A standard means of constructing an affine *DTSM* in continuous time is to start with a representation of  $X$  in one of the families  $A_M^{\mathbb{Q}}(N)$  and then to specify a market price of risk  $\eta_t$  that defines the change of measure from  $\mathbb{Q}$  to  $\mathbb{P}$  for  $X$ . In principle, starting with an affine<sup>Q</sup> 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  $\eta$ , up to the weak requirement that  $\eta$  not admit arbitrage opportunities. 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  $\eta$  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 *DTSMs* 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

$$f^{\mathbb{P}}(X_{t+1}|X_t) = (d\mathbb{P}/d\mathbb{Q})^D(X_{t+1}; \Lambda_t) \times f^{\mathbb{Q}}(X_{t+1}|X_t), \quad (30)$$

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 3); **(P2)**  $\Lambda_t$  is naturally interpreted as the market price of risk of  $X_{t+1}$ ; and **(P3)** rich nonlinear dependence of  $\Lambda_t$  on  $X_t$  is accommodated. In principle, any choice of  $(d\mathbb{P}/d\mathbb{Q})^D$  that is a known function of  $(X_{t+1}, \Lambda_t)$  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  $(d\mathbb{P}/d\mathbb{Q})^D$ :

$$\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)^D(X_{t+1}; \Lambda_t) = \frac{e^{\Lambda_t' X_{t+1}}}{\phi^{\mathbb{Q}}(\Lambda_t; X_t)}, \quad (31)$$

where  $\phi^{\mathbb{Q}}$  is the conditional Laplace transform of  $X$  under  $\mathbb{Q}$ ,  $\Lambda_t$  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  $X$ .<sup>16</sup> With this choice of  $(d\mathbb{P}/d\mathbb{Q})^D$ ,

---

<sup>16</sup>Buhlmann, Delbaen, Embrechts, and Shiryaev (1996) formally develop the conditional Esscher transform using martingale theory in the context of no-arbitrage pricing. A notable application of the Esscher transform (with constant  $\Lambda$ ) 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  $\Lambda_t$  to the market prices of risk.

the conditional  $\mathbb{P}$ -Laplace transform of  $X_t$  is given by

$$\phi^{\mathbb{P}}(u; X_t) = \frac{\phi^{\mathbb{Q}}(u + \Lambda_t; X_t)}{\phi^{\mathbb{Q}}(\Lambda_t; X_t)} = e^{\mathcal{A}(u; \Lambda_t) + \mathcal{B}(u; \Lambda_t)X_t}, \quad (32)$$

where  $\mathcal{A}(u; v) \equiv a(u + v) - a(v)$  and  $\mathcal{B}(u; v) \equiv b(u + v) - b(v)$ . 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  $\Lambda_t$  which, in turn, may be a nonlinear function of  $X_t$ . Thus, in general  $X$  is not an affine process under  $\mathbb{P}$ . We elaborate on the nature of the non-affine nature of this distribution below.

With this choice of  $(d\mathbb{P}/d\mathbb{Q})^D$ , the pricing kernel for pricing one-period ahead payoffs in our discrete-time model is

$$\mathcal{M}_{t,t+1} \equiv e^{-r_t} \times \frac{f^{\mathbb{Q}}(X_{t+1}|X_t)}{f^{\mathbb{P}}(X_{t+1}|X_t)} = e^{-r_t} \times \frac{e^{-\Lambda'_t X_{t+1}}}{\phi^{\mathbb{P}}(-\Lambda_t; X_t)}, \quad (33)$$

where we have used the fact that  $\phi^{\mathbb{P}}(-\Lambda_t; X_t) = [\phi^{\mathbb{Q}}(\Lambda_t; X_t)]^{-1}$ , which follows from (32) evaluated at  $u = -\Lambda_t$ . This choice of Radon-Nykodym derivative—equivalently pricing kernel  $\mathcal{M}$ —is natural in that, for small time interval  $\Delta$ , its counterpart in affine<sup>Q</sup> diffusion models  $(d\mathbb{P}/d\mathbb{Q})^C$  is approximately equal to  $(d\mathbb{P}/d\mathbb{Q})^D(t, t + \Delta)$ .<sup>17</sup> 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  $\Lambda$ , 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 *DTSMs* derived under the assumption that the  $\mathbb{Q}$ -representation of  $X$  resides in one of the families  $A_M^{\mathbb{Q}}(N)$ .

The restrictions 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  $u_i c_i < 1$ . Unless  $\Lambda_{it}c_i < 1$  almost surely, for  $i = 1, \dots, M$ ,  $\phi^{\mathbb{Q}}(\Lambda_t; X_t)$  is infinite with positive probability. In this case,  $f^{\mathbb{P}}$  would not integrate to unity for a set of  $X_t$  that has positive measure, and  $\mathbb{P}$  and  $\mathbb{Q}$  would not be equivalent. Examining these restrictions more closely, and using our mapping to the parameters of the related *CIR* process, we see that we are effectively requiring that  $2/(\sigma_i^2 \Delta t) > \Lambda_{it}$ ,  $i = 1, \dots, M$ . Typically  $\sigma_i^2$  is small and, depending on the application,  $\Delta t$  may also be small. Therefore,

<sup>17</sup> That is, for a small time interval  $\Delta$ , 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  $\epsilon_{t+\Delta}|X_t \sim N(0, \Delta I)$ ,

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

where  $\Lambda_t \equiv (\Sigma_X \sqrt{S_{Xt}})^{\prime -1} \eta_t$  is a transformation of the market price of risk  $\eta_t$ .

these bounds are typically weak and in the applications we have encountered so far they are far from binding. As  $\Delta t$  approaches zero (continuous time), the only requirement is that the  $\Lambda_{it}$  be finite almost surely.

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  $\mathbb{P}$ . We effectively know  $f^{\mathbb{Q}}(X_{t+1}|X_t)$  from the cross-sectional behavior of bond yields.<sup>18</sup> Furthermore, the relationship between the observed yields  $y_t$  and the state vector  $X_t$  are also known due to the pricing equation (11), which depends only on the risk-neutral distribution  $f^{\mathbb{Q}}(X_{t+1}|X_t)$ . Thus, the unknown function  $(d\mathbb{P}/d\mathbb{Q})^D(X_{t+1}; \Lambda_t)$  can be estimated from the time-series observations of bond yields,  $y_t$ .

## 6 The Market Prices of Risk

An immediate implication of (31) 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,  $\Lambda_t$  is not literally the market price of  $X$  risk (*MPR*), but rather the *MPR* is a nonlinear (deterministic) function of  $\Lambda_t$ . However, in a sense that we now make precise,  $\Lambda_t$  is the dominant term in the *MPR*. Accordingly, we will refer to  $\Lambda_t$  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<sup>19</sup>

$$\begin{aligned} E_t^{\mathbb{P}}[X_{t+1}] - E_t^{\mathbb{Q}}[X_{t+1}] &= [\mathcal{A}^{(1)}(0; \Lambda_t) - a^{(1)}(0)] + [\mathcal{B}^{(1)}(0; \Lambda_t) - b^{(1)}(0)] X_t \\ &= V_t^{\mathbb{P}}[X_{t+1}] \times \Lambda_t + o(\Lambda_t), \end{aligned} \quad (34)$$

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:  $\Lambda_t$  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

$$\begin{aligned} E_t^{\mathbb{P}} \left[ \log \frac{e^{-c'X_{t+1}}}{E_t^{\mathbb{Q}}[e^{-r_t} e^{-c'X_{t+1}}]} \right] - r_t &= -[a(-c) + c'a^{(1)}(\Lambda_t)] - [b(-c) + c'b^{(1)}(\Lambda_t)] X_t, \\ &= -c'V_t^{\mathbb{P}}[X_{t+1}] \times \Lambda_t + o(c) + o(\Lambda_t). \end{aligned} \quad (35)$$

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  $\Lambda_t$  is the dominant term in the true market price of risk underlying expected excess returns.

<sup>18</sup>Intuitively, taking the leading principal components as the state vector, we can estimate  $\delta_0$ ,  $\delta_X$ ,  $A_n$ , and  $B_n$  by regressing bond yields on this state vector. The parameters that characterize  $f^{\mathbb{Q}}(X_{t+1}|X_t)$  can then be estimated by treating the recursions (12) and (13) as (possibly nonlinear) cross-equation restrictions.

<sup>19</sup>The terms  $\mathcal{A}^{(1)}(0; \Lambda_t)$  and  $\mathcal{B}^{(1)}(0; \Lambda_t)$  are the first derivatives of  $\mathcal{A}$  and  $\mathcal{B}$  with respect to their first arguments, and  $a^{(1)}(u)$  and  $b^{(1)}(u)$  are the first derivatives of  $a(u)$  and  $b(u)$ .

A notable difference between  $\Lambda_t$  and the market price of risk  $\eta_t$  that appears in continuous-time  $(A_M^{\mathbb{Q}}(N), \eta)$  models is that  $\Lambda_t$  measures the price of risk per per unit of variance, whereas  $\eta$  measures risk in units of standard deviation. From the heuristic mapping between our choice of  $(d\mathbb{P}/d\mathbb{Q})^D$  and its continuous-time counterpart (see footnote 17) it is seen that this difference is simply a consequence of our (implicit) convention that

$$\Lambda_t = \left( \Sigma_X \sqrt{S_{Xt}} \right)^{\prime -1} \eta_t. \quad (36)$$

Researchers who want to replicate features of a continuous-time affine *DTSM*, can do so by setting  $\Lambda_t$  as in (36). For instance, choosing  $\eta_t$  as in Duffee (2002), Duarte (2004), or Cheridito, Filipovic, and Kimmel (2005) would lead to a discrete-time *DTSM* that locally (for small time interval  $\Delta$ ) would replicate the  $\mathbb{P}$  moments of their models.

More generally, it is evident from (34) that, starting from an affine  $E_t^{\mathbb{Q}}[X_{t+1}]$ , essentially any functional form for  $E_t^{\mathbb{P}}[X_{t+1}]$  is achievable by an appropriate choice of  $\Lambda_t$ . In particular, if one sets

$$\Lambda_t \equiv (\Sigma_X S_X(t) \Sigma_X')^{-1} (\mu^{\mathbb{P}}(X_t) - \mu^{\mathbb{Q}}(X_t)), \quad (37)$$

where  $\Sigma \sqrt{S(t)}$  is the diffusion term in an  $A_M^{\mathbb{Q}}(N)$  affine diffusion model and  $\mu^{\mathbb{P}}(X_t)$  is the desired  $\mathbb{P}$ -drift of a diffusion model for  $X$ , then locally one would obtain

$$E^{\mathbb{P}}[X_{t+\Delta}|X_t] = X_t + \mu^{\mathbb{P}}(X_t)\Delta + o(\Delta) \quad (38)$$

$$Cov^{\mathbb{P}}[X_{t+\Delta}|X_t] = \Sigma S(t) \Sigma' \Delta + o(\Delta). \quad (39)$$

That is, 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  $\Lambda_t$  as in (37).

Of course with  $\Lambda_t$  set to induce a nonlinear  $E_t^{\mathbb{P}}[X_{t+1}]$ , the conditional Esscher transform (31) in general induces nonlinear conditional  $\mathbb{P}$  moments of all orders, not just a nonlinear conditional mean. For example, letting  $\Lambda_{Z_t}$  and  $\Lambda_{Y_t}$  form a conformal partition of  $\Lambda_t$ , the conditional  $\mathbb{P}$ -mean of the  $i^{\text{th}}$  member of the  $M$ -vector of volatility factors  $Z_{t+1}$  is

$$E_t^{\mathbb{P}}[Z_{t+1}^i] = \frac{\partial}{\partial u_{Z_i}} [\mathcal{A}(u; \Lambda_t) + \mathcal{B}(u; \Lambda_t) X_t] \Big|_{u=0} = \frac{\nu_i c_i}{1 - \Lambda_{Z_t, i} c_i} + \frac{\rho_i}{(1 - \Lambda_{Z_t, i} c_i)^2} Z_t. \quad (40)$$

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

$$\text{Var}_t^{\mathbb{P}}[Z_{t+1}^i] = \frac{\nu_i c_i^2}{(1 - \Lambda_{Z_t, i} c_i)^2} + \frac{2c_i \rho_i Z_t}{(1 - \Lambda_{Z_t, i} c_i)^3}, \quad i = 1, \dots, M. \quad (41)$$

The nonlinearity of these moments, in contrast to their affine counterparts under  $\mathbb{Q}$  (see (9)), is induced by the state-dependence of  $\Lambda_{Z_t, i}$  through the terms  $1/(1 - \Lambda_{Z_t, i} c_i)$ .

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 (40) and (41) where each term of  $\text{Var}_t^{\mathbb{P}}[Z_{t+1}^i]$  is divided by one higher power of  $(1 - \Lambda_{Z_t, i} c_i)$ . Thus, the nonlinear

dependence in the mean achieved by one's choice of  $\Lambda_{Z_t}$  effectively determines the structure of the nonlinearity of the conditional second moments. This specialized structure is the discrete-time counterpart to the similarly special structure of 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  $\Lambda_t$  on  $X_t$ , it is desirable to impose sufficient structure on  $\Lambda_t$  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.<sup>20</sup> That  $X$  will not be a geometrically ergodic process for all specifications of  $\Lambda_t$  can be seen immediately from (40). If  $\Lambda_{Z_t, i}$  approaches  $\frac{1}{c_i}$  as  $Z_t^i$  increases, then the second term eventually dominates and the state variable is 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 2 (G.E.(Z))** *Suppose that the market price of risk  $\Lambda_Z(Z_t)$  is a continuous function of  $Z_t$ , and the eigenvalues of the matrix  $\rho$ ,  $\psi_i$  ( $i = 1, 2, \dots, M$ ), satisfy  $\max_i |\psi_i| < 1$ . If, in addition,*

1.  $\Lambda_Z(z) \leq 0$  for  $\forall z \geq 0$ , or
2.  $\Lambda_Z(z) \rightarrow \bar{\lambda} \leq 0$  as  $z \rightarrow \infty$  and  $\rho_{ij} = 0$  for  $0 \leq i \neq j \leq M$ ,

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

Establishing geometric ergodicity for the entire state vector  $X_t$  is more challenging, because of the range of possible specifications of  $\Lambda_{Y_t}$ , many of which 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 of  $X$ . Specifically, returning to the exponential-affine representations (4) and (32) for the conditional *MGFs*, upon making the dependence of the coefficients  $a(\cdot)$  and  $b(\cdot)$  of  $\phi^{\mathbb{Q}}$  on the risk-neutral parameters explicit by writing

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

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

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

---

<sup>20</sup>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).

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

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

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

$$\pi_{t,t+\tau}^D = \prod_{j=1}^{\tau} \left( \frac{d\mathbb{P}}{d\mathbb{Q}} \right)_{t+j-1,t+j}^D, \quad (42)$$

we can write

$$E^{\mathbb{P}} [g(X_{t+\tau}) | X_t] = E^{\mathbb{Q}} [g(X_{t+\tau}) \pi_{t,t+\tau}^D | X_t]. \quad (43)$$

The expectation on the right-hand-side of (43) 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 (43) by Monte Carlo simulation without reference to the right-hand side. This second approach is used in our empirical illustrations in Section 8.

## 7 The $\mathbb{P}$ Distribution in the Habit-based *DTSM*

To complete the specification of our habit-based *DTSM*, it remains to specify the market prices of risk and derive the physical distribution of bond yields. We take up these issues in this section, along with discussions of steady-state conditions and the continuous-time limit of our discrete-time model. The latter facilitates comparison with the habit-based models studied by *CC* and Wachter.

## 7.1 The Market Price of Risk in the Habit-based *DTSM*

Substituting (17) into (3) leads to

$$\begin{aligned}
 -m_{t,t+1} &= -\gamma \left( 1 + \frac{\sigma_g}{\sigma_t^{\mathbb{Q}}[z_{t+1}]} \right) z_{t+1} + \pi_{t+1} \\
 &\quad -\log \delta + \gamma z_t + \gamma f(z_t) + \gamma \sigma_g \frac{E_t^{\mathbb{Q}}[z_{t+1}]}{\sigma_t^{\mathbb{Q}}[z_{t+1}]}.
 \end{aligned} \tag{44}$$

Since the market price of risk<sup>21</sup>  $\Lambda_t$  is, by definition, the loading on  $X_t$  in  $m_{t+1}$ ,

$$\Lambda_t = \begin{bmatrix} -\gamma \left( 1 + \frac{\sigma_g}{\sigma_t^{\mathbb{Q}}[z_{t+1}]} \right) \\ 1 \end{bmatrix}. \tag{45}$$

It follows that the market price of inflation risk is constant at 1 and the market price of inverse surplus consumption risk is time-varying and (potentially highly) nonlinear in  $z_t$ . The corresponding physical density of  $X_{t+1}$  is given by:

$$f^{\mathbb{P}}(z_{t+1}, \pi_{t+1} | z_t, \pi_t) = f^{\mathbb{Q}}(z_{t+1}, \pi_{t+1} | z_t, \pi_t) \times \frac{e^{\Lambda_t' [z_{t+1}, \pi_{t+1}]'}}{\phi^{\mathbb{Q}}(\Lambda_t; [z_t, \pi_t])}. \tag{46}$$

In implementing the *ML* estimator using this physical density, we constrain the parameters of our model to rule out non-stationarity and an absorbing boundary for surplus consumption. The following proposition gives conditions under which the state variables are geometrically ergodic and  $z_t$  is non-absorbing at zero.

**Proposition 3** *If*

$$\sigma_{\pi,g} > \frac{\sqrt{\rho_z} - 1}{c_z} - \gamma, \quad \rho_{\pi} \in (0, 1), \quad \text{and } v_z \geq 1, \tag{47}$$

*then the state variables  $z_t$  and  $\pi_t$  are geometrically ergodic and non-absorbing at zero.*

**Proof:** See Appendix B.

It should be noted that Proposition 3 only gives sufficient conditions. Owing to the nonlinear dynamics under the physical measure, we have not discovered a set of necessary conditions for ergodicity. Simulations at various parameters values, however, suggest that the conditions of Proposition 3 are close to being necessary. Even slight violations of these constraints will often result in explosive behavior of the state variables.

---

<sup>21</sup>Consistent with the earlier sections, the concept of the market price of risk used here refers to the price per unit of variance of the state variables.

## 7.2 Steady State Conditions

Following  $CC$  and Wachter, we require that:

$$\frac{\partial \log H_{t+1}}{\partial c_{t+1}} = 0 \Bigg|_{z_t = \bar{z}}, \quad (48)$$

$$\frac{\partial (\partial \log H_{t+1} / \partial c_{t+1})}{\partial z_t} = 0 \Bigg|_{z_t = \bar{z}}. \quad (49)$$

As explained by  $CC$ , the first condition guarantees that the (log) habit level  $\log H_t$  is a deterministic function of past consumption around the steady state ( $\bar{z}$ ). The second condition ensures that this deterministic function is locally increasing in past consumption.

As shown in Appendix C, these conditions impose the following constraints on the model parameters:

$$\bar{z} = \frac{A\mathcal{B}\rho_z}{1 + 2\mathcal{B}\rho_z}, \quad (50)$$

$$\nu_z = \frac{(\mathcal{A} - 2\bar{z})\rho_z}{c_z}, \quad (51)$$

$$s_{max} = \bar{z} + \log(1 - \mathcal{A}), \quad (52)$$

where

$$\mathcal{A} = 1 + \frac{\sigma_g^2}{2c_z\rho_z} - \sqrt{\frac{\sigma_g^2}{c_z\rho_z} + \frac{\sigma_g^4}{4c_z^2\rho_z^2}} \quad \text{and} \quad \mathcal{B} = \frac{1 + \left(\frac{\gamma}{\mathcal{A}} + \sigma_{\pi,g}\right) c_z}{\left(1 + \left(\frac{\gamma}{\mathcal{A}} + \sigma_{\pi,g}\right) c_z\right)^2 - \rho_z}. \quad (53)$$

## 7.3 The Continuous Time Limit

To put the model parameters in connection with the time interval  $\Delta$ , let:

$$\begin{aligned} \rho_z &= 1 - \kappa_z\Delta; \quad c_z = \frac{1}{2}\sigma_z^2\Delta; \quad \nu_z = \frac{2\kappa_z\theta_z}{\sigma_z^2} \\ \rho_\pi &= 1 - \kappa_\pi\Delta; \quad \rho_{\pi,z} = -\kappa_{\pi,z}\Delta; \quad \sigma_\pi = \sigma_\pi^c\sqrt{\Delta} \\ \sigma_g &= \sigma_g^c\sqrt{\Delta}; \quad \mathcal{C} = \mathcal{C}^c\Delta. \end{aligned} \quad (54)$$

**Proposition 4** *In the continuous time limit,  $X'_t = (z_t, \pi_t)$  follows the risk-neutral process*

$$dX_t = (\kappa\theta_X^{\mathbb{Q}} - \kappa_X^{\mathbb{Q}}X_t)dt + \Sigma S_{z,t}dB_{X,t}^{\mathbb{Q}}, \quad (55)$$

where

$$\kappa\theta_X^{\mathbb{Q}} = \begin{bmatrix} \kappa_z\theta_z \\ \kappa_\pi\bar{\pi} + \kappa_{\pi,z}\theta_z \end{bmatrix} \quad \text{and} \quad \kappa_X^{\mathbb{Q}} = \begin{bmatrix} \kappa_z & 0 \\ \kappa_{\pi,z} & \kappa_\pi \end{bmatrix},$$

$$\Sigma = \begin{bmatrix} \sigma_z & 0 \\ -\sigma_{\pi,g}\sigma_z & \sigma_\pi^c \end{bmatrix} \text{ and } S_{z,t} = \begin{bmatrix} \sqrt{z_t} & 0 \\ 0 & 1 \end{bmatrix}.$$

Under the  $\mathbb{P}$  measure,  $X_t$  follows the process

$$dX_t = (\kappa\theta_X^{\mathbb{P}} - \kappa_X^{\mathbb{P}}X_t - \phi\sqrt{z_t})dt + \Sigma S_{z,t}dB_{X,t}^{\mathbb{P}}, \quad (56)$$

where

$$\kappa\theta_X^{\mathbb{P}} = \kappa\theta_X^{\mathbb{Q}} + \begin{bmatrix} 0 \\ \sigma_\pi^{c2} \end{bmatrix}, \quad \kappa_X^{\mathbb{P}} = \kappa_X^{\mathbb{Q}} + \sigma_z^2(\gamma + \sigma_{\pi,g}) \begin{bmatrix} 1 & 0 \\ -\sigma_{\pi,g} & 0 \end{bmatrix}, \quad \text{and } \phi = \gamma\sigma_g^c\sigma_z \begin{bmatrix} 1 \\ -\sigma_{\pi,g} \end{bmatrix}.$$

The consumption growth process approaches the diffusion

$$g_t = [\mathcal{C}^c - (\gamma + \sigma_{\pi,g})\sigma_g^c\sigma_z\sqrt{z_t}] dt - \sigma_g^c dB_{z,t}^{\mathbb{Q}} \quad (57)$$

under  $\mathbb{Q}$ , and the process

$$g_t = (\mathcal{C}^c + \gamma\sigma_g^{c2})dt - \sigma_g^c dB_{z,t}^{\mathbb{P}} \quad (58)$$

under the historical distribution.

**Proof:** See Appendix E.

Proposition 4 confirms that the states processes under our formulation are exponentially affine under  $\mathbb{Q}$ . Moreover, from equation (56), it is seen that the nonlinearity in the drift of the physical state processes takes a particularly simple form: it depends on the square-root of the inverse consumption surplus  $z_t$ . This form of nonlinearity bears close resemblance to that considered by Duarte (2004). Whereas Duarte (2004) studies a reduced-form model with the coefficient on the nonlinear term being a free parameter, our structural setup defines this coefficient in terms of the underlying parameters of the model. For example, the coefficient of  $\sqrt{z_t}$  for the  $\mathbb{P}$ -process of  $z_t$  is the product of  $\gamma$ ,  $\sigma_g^c$  and  $\sigma_z$ . Since the  $\mathbb{P}$ -nonlinearity of  $z_t$  arises from the nonlinear risk premiums implied by habit formation, it is intuitive that the nonlinear component in the  $\mathbb{P}$ -drift of  $z_t$  is a function of the utility curvature  $\gamma$  (which modulates the price of risk) and  $\sigma_g^c$  and  $\sigma_z$  (which modulate the quantities of risks).

Note also from equation (58) that consumption growth  $g_t$  approaches a homoskedastic process with a constant mean under  $\mathbb{P}$ . This justifies the choice we made earlier in modeling the risk-neutral conditional expectation of  $g_{t+1}$ ,  $f(z_t)$ . Specifically, to align our model to those of *CC* and Wachter,  $f(z_t)$  is chosen so that its nonlinearity is exactly netted out by the nonlinearity generated by the habit-based market prices of risk, giving a homoskedastic  $\mathbb{P}$ -process with a constant mean. More generally, allowing for some degree of predictability of consumption growth under  $\mathbb{P}$  is feasible within our modeling framework. For example, with a slight modification to  $f(z_t)$ , the  $\mathbb{P}$ -drift of  $g_{t+1}$  could be driven by  $z_t$  in a manner very similar to the long run risk model of Bansal and Yaron (2004).

## 8 Empirical Illustrations

Previous studies of habit-based models of asset prices have typically focused on parameters chosen by matching model-implied moments to a selected set of sample moments of the data.<sup>22</sup> As we document subsequently, the degree to which habit-based models resolve puzzles in the bond pricing literature depend on which of seemingly equally sensible sets of moments are used in calibration. This sensitivity motivates our interest in examining the properties of our model evaluated at the maximum likelihood (*ML*) estimates of the model. The likelihood function implicitly uses all of the moments of the distributions of the variables in the model, weighted by the precision with which they are estimated. *ML* estimation is relatively challenging in Wachter (2005)'s formulation of the habit-based model, owing to the nonlinear dependence of bond yields on the state. Within our framework, joint *ML* estimation of all model parameters is feasible since both analytical bond prices and likelihood function are available.

Summarizing the estimation problem, the  $\mathbb{Q}$  distribution of the inverse surplus consumption ratio  $z_t$ , a CIR-like process, is governed by three parameters: the persistence parameter  $\rho_z$ , the volatility parameter  $c_z$ , and risk-neutral long-run mean of  $z_t$ ,  $\nu_z$ . Whereas  $\rho_z$  and  $c_z$  are free parameters,  $\nu_z$  is determined as function of other parameters of the model (to satisfy the steady-state conditions described in Appendix C). Similarly,  $\rho_\pi$ ,  $\sigma_\pi$  and  $\theta_\pi$  govern the persistence, volatility and long run mean of the risk-neutral inflation process. In addition, the contemporaneous correlation and feedback effect between  $z_t$  and  $\pi_t$  are captured through  $\sigma_{\pi g}$  and  $\rho_{\pi g}$ , respectively.  $\delta$  is the subjective discount factor.  $\delta_0$  is the constant term in the short rate equation. Finally,  $\gamma$  determines the curvature of the habit utility function.

### 8.1 Data

We follow Piazzesi and Schneider (2007) and construct our quarterly measures of inflation and real consumption from the NIPA price and quantity indexes.<sup>23</sup> Compared to the CPI index which covers a wide basket of goods, our inflation measure maps precisely to the measure of aggregate consumption used in the analysis. Only consumption of non-durable goods and services is included. Total real consumption is divided by the corresponding population series, obtained from the Census Bureau. To reduce the level of measurement noise in the inflation series, we follow the suggestion of Kim (2008) and process our inflation series through an ARMA(1,1) filter:

$$\pi_t = (1 - 0.924)0.010 + 0.924\pi_{t-1} + \epsilon_t - 0.346.\epsilon_{t-1} \quad (59)$$

---

<sup>22</sup> Wachter (2005), for example, calibrates her model through a two-step process: (1) parameters governing the physical dynamics of consumption growth and inflation are estimated using limited-information *ML* methods (the constraints imposed by the pricing model are not enforced); (2) given these estimated parameters, other parameters of the model are calibrated to match certain moments of the data. This two-step procedure, also adopted by Boudoukh (1993), reduces the size of the parameter space in calibration, thereby alleviating the computational burden in numerically computing bond prices.

<sup>23</sup>NIPA tables 2.3.3, 2.3.4 and 2.3.5.

We then use an exponentially smoothed measure of observed inflation:

$$(0.924 - 0.346) \sum_{j=0}^{\infty} 0.346^j (\pi_{t-j} - 0.010) + 0.010 \quad (60)$$

as the true inflation series purged of measurement noise.<sup>24</sup>

The interest rate data are downloaded from the Federal Reserve's web page accompanying Gurkaynak, Sack, and Wright (2006).<sup>25</sup> Available maturities are in whole numbers of years, ranging from one to seven years. Our analysis is performed using quarterly data over the sample period 1961 through 2007.

## 8.2 Calibration

As an informative first-step towards the analysis of our habit-based *DTSM* we calibrate the model to various sample moments in the data in order to explore the sensitivity of the model's properties to alternative choices of parameter values. We choose  $\sigma_g$  to match the standard deviation of consumption growth in the data. Then, for each set of  $\{\gamma, \rho_z, c_z, \rho_\pi, \rho_{\pi,z}, \sigma_{\pi g}, \sigma_\pi\}$ , we compute  $\nu_z$  from the steady state conditions described in Appendix C. Given  $\rho_z, c_z, \nu_z$ , we simulate a long series of  $z_t$ , and choose  $\theta_\pi$  to match the sample mean of inflation,  $E_T[\pi_t]$ .<sup>26</sup>  $\delta_0$  is chosen to match the observed level of the yield curve, defined as the midpoint between the mean of the 1-year zero yields and the 7-year zero yields. Next, we choose  $\delta$  to match the sample mean of consumption growth,  $E_T[g_t]$ .<sup>27</sup>

Finally, we choose  $\{\gamma, \rho_z, c_z, \rho_\pi, \rho_{\pi,z}, \sigma_{\pi g}, \sigma_\pi\}$  to match sample moments from the data according to one of the following two schemes. In the first scheme (*CS*), we place positive weights on the sample means and volatilities of interest rates, inflation volatility, inflation persistence, the unconditional correlation between consumption growth and inflation, and

<sup>24</sup>We also re-ran our analysis using the raw inflation series and found no significant qualitative changes. Motivated by similar considerations, Wachter (2005) also uses an ARMA filter to process her inflation and consumption data.

<sup>25</sup> <http://www.federalreserve.gov/Pubs/feds/2006/200628/200628abs.html>

<sup>26</sup>The simulation size is 50,000, after a burn-in sample of 5,000. It can be shown that:

$$\theta_\pi = E_T[\pi_t] - \frac{1}{1 - \rho_\pi} \left( \rho_{\pi z} \left( E_S[z_t] - \frac{\nu_z c_z}{1 - \rho_z} \right) - E_S[z_t] \sigma_{\pi g} (1 - \rho_z) + (\sigma_\pi^2 + \sigma_{\pi g} \nu_z c_z) \right),$$

where  $E_S[\cdot]$  denotes averaging over simulated values.

<sup>27</sup>Precisely, we match

$$\log(\delta) = \gamma E_T[g_t] - \left( \delta_0 + \gamma \nu_z c_z - \theta_\pi (1 - \rho_\pi) + \rho_{\pi z} \frac{\nu_z c_z}{1 - \rho_z} - \frac{1}{2} \sigma_\pi^2 + \frac{1}{2} \gamma^2 \sigma_g^2 - \frac{1}{2} \nu_z c_z^2 (\gamma + \sigma_{\pi g})^2 \right).$$

This equation matches the drift of the continuous-time consumption growth process to the discrete-time sample mean. This is convenient since a closed-form expression for average consumption growth is only available in the continuous time limit. As shown in Table 2 the errors from not using the model-implied mean in discrete time are small.

the Campbell and Shiller (1987) (CS) regression coefficients.<sup>28</sup> In the second scheme (VO), we adopt the same weighing scheme except that we place zero weights on the CS regression coefficients (the slope coefficient in the regression of changes in long-term bond yields on the slope of the yield curve). The calibrated values of the parameters are displayed in Table 1.

	Calibration		ML Estimation		
	Scheme CS	Scheme VO	Estimates	Asymptotic s.e.	Small-sample s.e.
$\gamma$	2.1977	5.0000	2.4005	0.1230	0.1369
$\delta$	0.9904	1.0134	0.9697	0.0027	0.0042
$\sigma_g$	0.0044	0.0044	0.0048	0.0002	0.0008
$\rho_z$	1.0162	1.0012	1.0273	0.0015	0.0038
$c_z$	0.0070	0.0006	0.0120	0.0002	0.0015
$\rho_\pi$	0.8941	0.8957	0.9467	0.0065	0.0091
$\theta_\pi$	0.0000	-0.0348	0.0131	0.0016	0.0019
$\rho_{\pi z}$	0.0015	0.0021	-0.0001	0.00003	0.00005
$\sigma_{\pi g}$	-0.0223	-0.1019	-0.0042	0.0023	0.0050
$\sigma_\pi$	0.0001	0.0005	0.0021	0.0001	0.0007
$\delta_0$	0.0053	0.0017	0.0043	0.0003	0.0008
$\nu_z$	1.0819	4.0318	1.2869	-	-
$\bar{z}$	0.471	0.416	0.471	-	-
$s_{max}$	-2.51	-1.38	-2.69	-	-

Table 1: Parameters values from calibration and from full-information ML estimation of the habit-based model. Small-sample standard errors are computed using ML estimates from 100 simulated samples with a length of 185 quarters.

The moments of the consumption growth and inflation process corresponding to two calibrated parameter sets are reported in Table 2. Both calibrated models do a good job of capturing the moments of inflation and consumption growth in the data.<sup>29</sup>

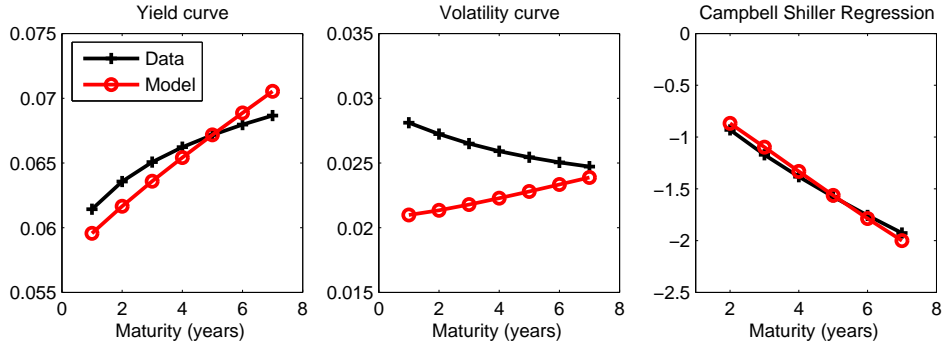
	Data	Scheme CS	Scheme VO
$E[g_t]$	0.0053	0.0054	0.0053
$\sigma(\pi_t)$	0.0059	0.0058	0.0062
$corr(g_t, \pi_t)$	-0.3382	-0.2670	-0.3392
$corr(\pi_t, \pi_{t+1})$	0.9324	0.9320	0.9320

Table 2: Sample and Model-Implied Moments

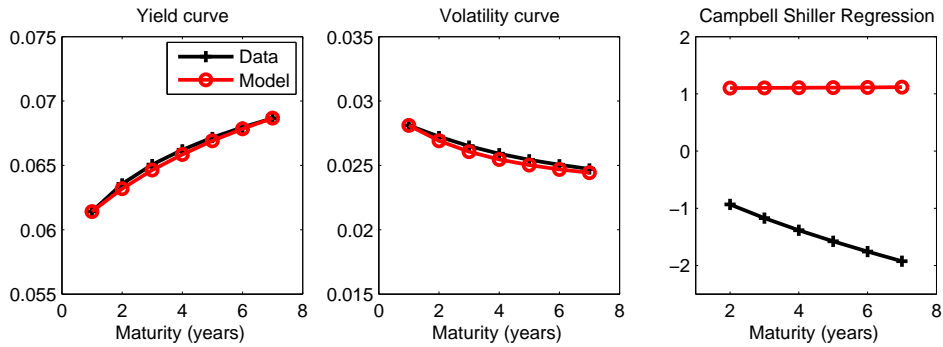
However, notable differences between the models emerge when we examine the model-implied moments of bond yields. For each calibration scheme, the three graphs in Figure 1 display, from left to right, the sample and model-implied average yield curve, term structure

<sup>28</sup>We minimize the sum of squared differences between the model-implied and sample moments. To ensure that the yield curve is matched on average, we multiply the difference in means of interest rates by a factor of 10 before computing the sum of squared errors. All other moments receive a weight of 1.

<sup>29</sup>The volatility of consumption growth and the long run mean of inflation are not reported since they are perfectly matched as part of our calibration process.



(a) Scheme CS



(b) Scheme VO

Figure 1: Moment Matching from Calibration Schemes: from left to right, average yields, volatility of yields, and Campbell-Shiller regression coefficients.

of volatility, and CS regression coefficients. Focusing first on scheme *CS* (Figure 1(a)) the model-implied CS regression coefficients match strikingly well with those in our sample. This near perfect match does not come without compromising the fit to other moments. In particular, scheme *CS* produces an upward sloping volatility curve which is contrary to what is seen in the data: yield volatilities decay with maturity. Under scheme *VO* (Figure 1(b)), the sample average yield curve and volatility curve are matched perfectly. However now, the model completely fails to match the CS coefficients— as if the expectations theory holds in this model.

In the light of the sensitivity of fit to the choice of calibration scheme documented in Figure 1, we turn next to an exploration of the fit based on full-information *ML* estimates of the parameters.

### 8.3 ML Estimation

For each quarter in our sample, we compute the inverse consumption surplus ratio,  $z_{t+1}$ , from equation (17), based on our observation of  $g_{t+1}$  and the previously implied value of  $z_t$ :

$$z_{t+1} = E_t^{\mathbb{Q}}[z_{t+1}] - \frac{\sigma_t^{\mathbb{Q}}[z_{t+1}]}{\sigma_g}(g_{t+1} - f(z_t)). \quad (61)$$

The physical density  $f^{\mathbb{P}}(z_{t+1}\pi_{t+1}|z_t, \pi_t)$  is then computed using equation (46). In addition, we assume that bonds with one, four and seven years to maturity are priced with normally distributed i.i.d. errors with mean zero and constant variances. This distributional assumption for the pricing errors introduces minimal additional flexibility in fitting yields, beyond that inherent in the habit-based *DTSM*. Combining these observations, and letting  $R_t$  denote the continuously compounded yields on these three bonds, the likelihood of the observed time series  $\{g_t, \pi_t, R_t'\}$  is

$$\mathcal{L}(\{g_t, \pi_t, R_t\}_{t=2}^T) = \prod_{t=2}^T \frac{\sigma_t^{\mathbb{Q}}[z_{t+1}]}{\sigma_g} f^{\mathbb{P}}(z_{t+1}, \pi_{t+1}|z_t, \pi_t) f^{\mathbb{P}}(R_{t+1}^1, R_{t+1}^4, R_{t+1}^7|z_{t+1}, \pi_{t+1}), \quad (62)$$

where the first term on the right-hand side of (62) is the Jacobian of the transformation between  $g_t$  and  $z_t$ .<sup>30</sup> The resulting estimates and their associated standard errors are reported in the last three columns of Table 1.

All of the parameters are estimated with considerable precision. The point estimate of the utility curvature parameter ( $\gamma$ ) is 2.4 - a value quite close to what is adopted in studies of the equity premium and Wachter's choice of 2. Likewise, the steady-state value of  $z_t$  ( $\bar{z}$ ) and the upper boundary of  $s_t$  ( $s_{max}$ ) are very close in magnitude to those used by *CC* and Wachter. Those parameters associated with risk-neutral distribution of the state are not directly comparable to the values in previous studies.

Moreover, the model-implied fitted values of surplus consumption and habit both seem plausible. From Figure 2(a) it is seen that  $s_t$  co-moves strongly with the business cycle, with four noticeable troughs corresponding to recessions in 1975, 1982, 1991 and 2002. The time-series behavior of  $H_t$  (Figure 2(b)) is very much in line with our expectations: it is smooth, persistent and increasing with the level of consumption.

Having established that our model fits many features of the macro variables well, we turn next to an exploration of the fit to moments on bond yields. Figure 3 displays model-implied population term structures of the means and volatilities of bond yields ("Long run"),

---

<sup>30</sup>In maximizing this likelihood function we address the possibility of negative fitted values of  $z_t$  by assuming that any such negative values of  $z_t$  are the manifestation of an exponentially distributed error. In this manner errors in  $z_t$  that lead to negative fitted values are continuously penalized and, in the presence of such errors, the likelihood function remains smooth. In our sample there were a small number of negative fitted  $z$ 's, all of which occurred prior to 1974.

Also, to ensure that our estimates are global optima we implement the optimization in two steps. First, we randomly generate thousands of starting points, quickly improve them within a short time window and then rank them in the order of likelihood value. Second, we use the best 500 parameter sets as starting points and numerically maximize the likelihood function (62) until convergence. Out of these 500 local optima we select the parameter set that yields the highest likelihood value.

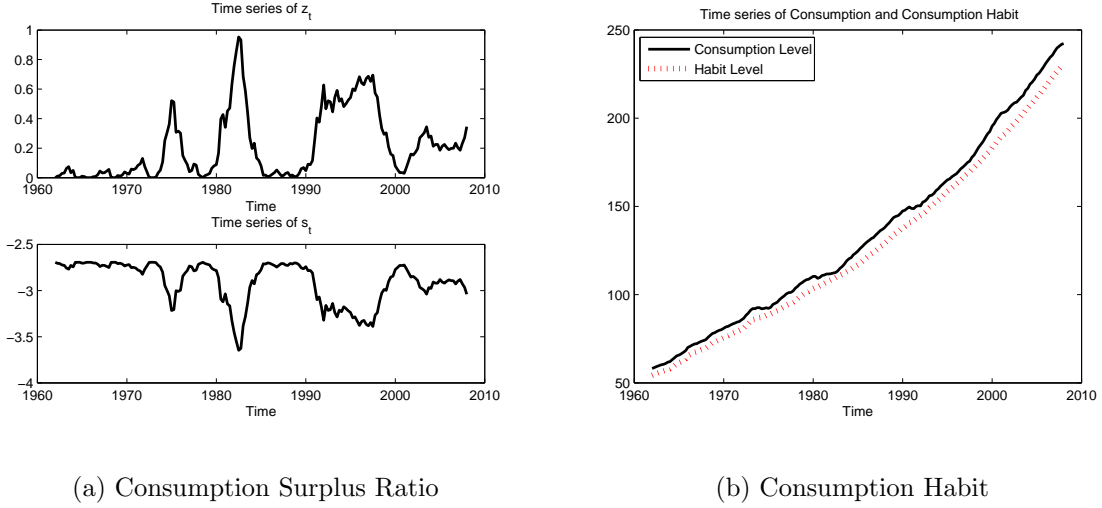


Figure 2: Time Series of Consumption Surplus Ratio ( $s_t$ ) and Consumption Habit ( $H_t$ )

their sample counterparts (“Data”), and fifth and ninety-fifth percentiles of the small-sample distributions of these statistics. The latter are computed by simulating 5000 sample paths of length 185 quarters (the size of our sample of bond yields) and, for each sample path, computing the moments of bond yields. The means of the small-sample distributions of these moments are very similar to their population counterparts, and so they are omitted to avoid congestion in these figures.

The level and the slope of the yield curve, as well as the term structure of volatilities of yields, are reasonably well matched. Although the long end of the population mean yield curve is higher than its sample counterpart, the latter is bracketed by the the 5<sup>th</sup> and 95<sup>th</sup> percentiles of the small-sample distribution of the sample means. The population term structure of volatility (Figure 3(b)) lies below our sample counterpart, perhaps owing in part to the fact that bond yields are priced with error in our setup. Nevertheless, the model clearly captures the pronounced downward slope in the volatility curve. Additionally, the percentiles of the model-implied small-sample distribution of volatilities come close to bracketing the sample estimates, even without adding on the volatilities of the pricing errors.

We are less successful at replicating the failure of the expectations hypothesis. From Figure 4 it can be seen that the population *CS* regression coefficients (“Long-run mean”) lie below one and exhibit a decreasing pattern. However, this line is substantially above the historically estimated coefficients (marked “Data”), and even the 5<sup>th</sup> percentile values of the small-sample distribution lie well above the sample coefficients.

Why does the habit-based *DTSM* fail to resolve the expectations puzzle? Letting  $\xi_t^n$  denote the expected excess return from holding the  $n$ -period bond for 1 period, the *CS* regression coefficients are

$$\phi_n = 1 - n \frac{\text{cov}(R_t^n - R_t^1, \xi_t^n)}{\text{var}(R_t^n - R_t^1)}. \quad (63)$$

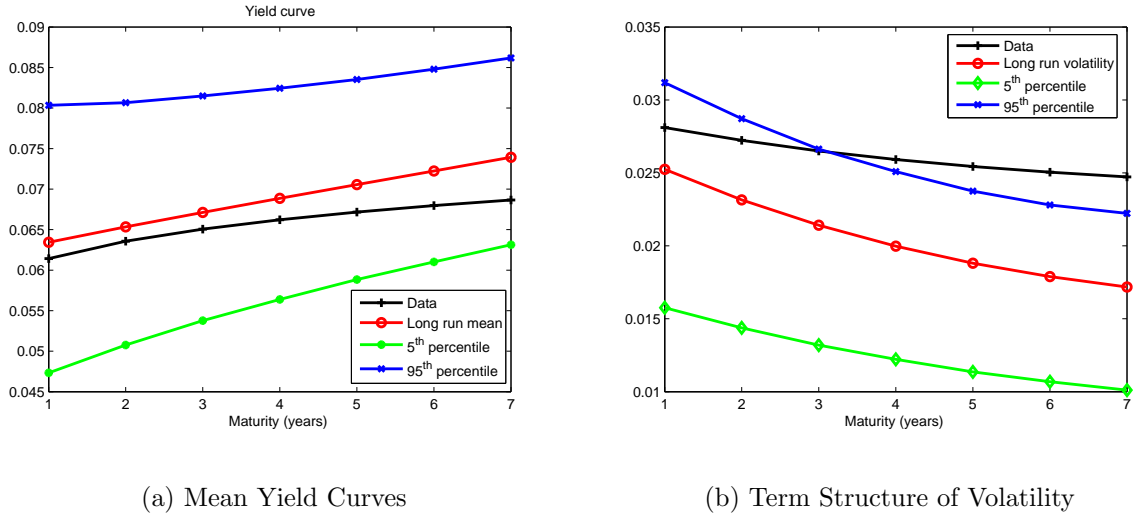


Figure 3: Sample, population, and small sample distributions of means and volatilities of bond yields.

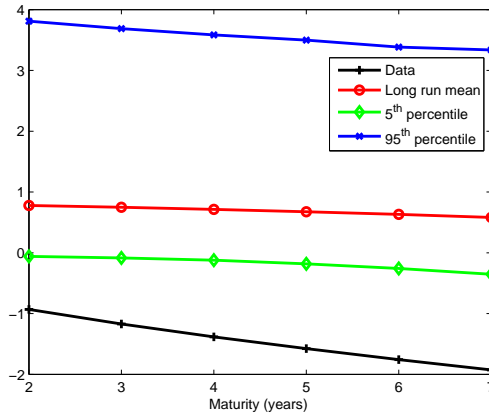


Figure 4: Campbell-Shiller Regressions

To generate negative  $\phi_n$  as required by the data a model needs to produce a positive correlation between the slope of the yield curve and  $\xi_t^n$ . Now for n-period bond with the corresponding zero yield  $R_T^n = a_n + b_n z_t + c_n \pi_t$ , its one-period expected excess return is approximately (see Section 6):

$$\xi_t^n \approx \begin{bmatrix} -b_n & -c_n \end{bmatrix} \begin{bmatrix} \sigma_z^2 z_t & -\sigma_{\pi g} \sigma_z^2 z_t \\ -\sigma_{\pi g} \sigma_z^2 z_t & \sigma_{\pi g}^2 \sigma_z^2 z_t + \sigma_{\pi}^2 \end{bmatrix} \begin{bmatrix} -\gamma \left(1 + \frac{\sigma_g}{\sigma_z \sqrt{z_t}}\right) \\ 1 \end{bmatrix} \quad (64)$$

$$= (b_n - c_n \sigma_{\pi g}) [(\gamma + \sigma_{\pi g}) \sigma_z^2 z_t + \gamma \sigma_g \sigma_z \sqrt{z_t}] - c_n \sigma_p^2, \quad (65)$$

where  $\sigma_z = \sqrt{2c_z}$ . For the yield curve to be upward sloping on average,  $\xi_t^n$  must be positive

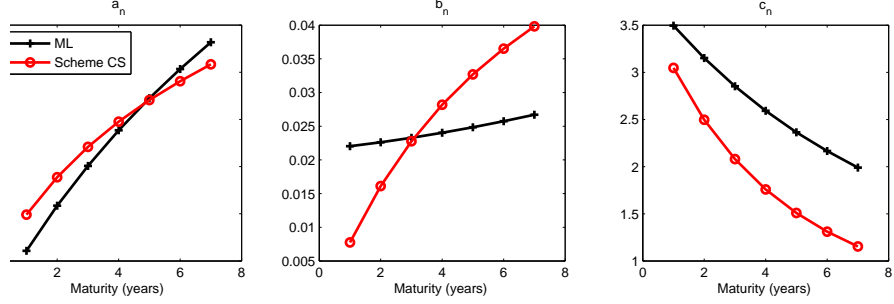


Figure 5: Loadings:  $R_t^n = a_n + b_n z_t + c_n \pi_t$

on average, which requires  $(b_n - c_n \sigma_{\pi g}) > 0$ .<sup>31</sup> This, in turn, makes  $\xi_t^n$  an increasing function of  $z_t$ . As a result, for the slope of the yield curve to be positively correlated with  $\xi_t^n$ , it must be positively correlated with  $z_t$ .

Ignoring the constant term, the slope of the yield curve is  $(b_n - b_1)z_t + (c_n - c_1)\pi_t$ . For the first term to contribute positively to  $\text{corr}(\xi_t^n, R_t^n - R_t^1)$ ,  $b_n - b_1$  must be positive and this, in turn, calls for a risk-neutral mean reversion parameter of  $z_t$  ( $\rho_z$ ) greater than 1. From Table 1,  $\rho_z$  is calibrated at 1.0162 under the *CS* scheme and estimated at 1.0273, thus both generating an increasing pattern of  $b_n$  (see Figure 5). Turning to the second term, since  $c_n - c_1 < 0$ , it will contribute positively to resolving the expectations puzzle if  $\text{corr}(z_t, \pi_t) < 0$ . However,  $\text{corr}(\pi_t, g_t) < 0$  and  $z_t$  is conditionally perfectly negatively correlated with  $g_t$ , so inducing  $\text{cov}(z_t, \pi_t) < 0$  within this habit model would be quite challenging. From simulations,  $\text{corr}(z_t, \pi_t)$  is indeed positive (0.0024) under the *CS* scheme. With the *ML* estimates,  $\text{corr}(z_t, \pi_t)$  is just negative (-0.0001).

Comparing the calibrated parameters for Scheme *CS* to corresponding the *ML* estimates, it is striking how closely many of them match up. Yet, as we have seen, these two parameter sets have very different implications for the moments of bond yields. Key to understanding this difference is by comparing the patterns of loadings across the different parameter sets. The fact that  $\text{corr}(z_t, \pi_t) > 0$  under the *CS* scheme means that  $b_n$  has to increase quite fast in  $n$  to, first, offset the negative correlation generated from  $(c_n - c_1)\pi_t$  and, second, create a positive correlation between  $\xi_t^n$  and  $z_t$ . However, this very effort to generate an increasing pattern in  $b_n$  contributes to an increasing pattern in bond yield volatility  $\text{var}(R_t^n) = b_n^2 \text{var}(z_t) + c_n^2 \text{var}(\pi_t) + 2b_n c_n \text{cov}(z_t, \pi_t)$ . Omitting the first (increasing) component ( $b_n^2 \text{var}(z_t)$ ), we confirm through simulations that the other two components ( $c_n^2 \text{var}(\pi_t) + 2b_n c_n \text{cov}(z_t, \pi_t)$ ) decrease in maturity  $n$  under the *CS* scheme.

Likewise, in order to match the downward sloping pattern of volatilities, the *ML* estimates generate a slowly increasing pattern of  $b_n$ . However, this proves to be too slow to induce sufficient positive correlation between  $\xi_t^n$  and  $z_t$  to match the data. To see this graphically we plot the implied  $b_n - b_1$  and  $\frac{b_n}{b_1}$  for the two calibration schemes and the *ML* estimates in Figure 6. The steep slopes of  $b_n - b_1$  as well as  $\frac{b_n}{b_1}$  distinctly set the *CS* scheme apart from

<sup>31</sup>Strictly speaking, to have  $(b_n - c_n \sigma_{\pi g}) > 0$ , we also need  $\gamma + \sigma_{\pi g} > 0$ . This is guaranteed, from the ergodicity condition and the requirement that  $\rho_z > 1$  discussed in the text.

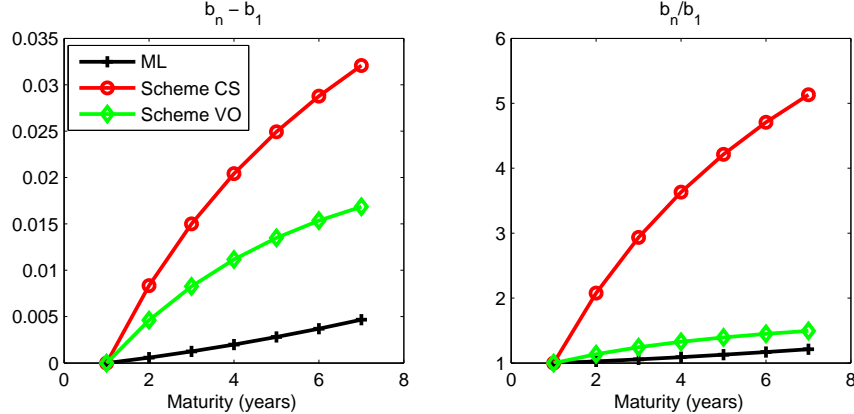


Figure 6: Comparison of  $b_n$

the *VO* scheme and the *ML* estimates.

This pattern of the loadings  $b_n$  helps the *CS* calibration scheme match the *CS* regression coefficients better, but at the expense of not matching the volatility curve. From Table 4 in Wachter (2005) it is seen that, at her calibrated value of the parameters, her habit-based model also produces a counter-factual upward sloping term structure of volatility. While we note that there are differences between our setup and that of Wachter (2005), given the similarity in how risk-premiums are generated under the two economies, it is quite possible that the same tension that we find here extends to Wachter (2005)'s setup as well.

Though the literature on evaluating the ability of equilibrium *DTSMs* to resolve the expectations puzzle has focused largely on the *CS* coefficients, Dai and Singleton (2002) show that a successful model should also replicate the risk-premium adjusted regressions coefficients. That is, in the regressions

$$R_{t+1}^{n-1} - R_t^n + \frac{1}{n-1} E_t[\xi_{t+1}^n] = \text{constant} + \phi_n \frac{R_t^n - R_t^1}{n-1} + \text{residual}, \quad (66)$$

the coefficients  $\phi_n$  should be one for all maturities. Figure 7 displays these adjusted coefficients for model-implied yields evaluated at our *ML* estimates as well as at the calibrated parameter values. The *ML* premium-adjusted coefficients come closest to having  $\phi_n = 1$ , but all of the estimates fall well below their theoretical value.

In summary, though our illustrative habit-based *DTSM* produces time varying expected excess returns, the magnitude and nature of this predictability is not sufficient to resolve the expectations puzzles in the literature. Central to this failure is the inherent tension between matching the slope of the volatility curve and the time series behavior of expected excess returns.<sup>32</sup> There is room for improvement. In its current form, the surplus consumption ratio is the lone driver of expected excess return. As such, the distribution of  $z_t$  is largely responsible for matching both the the volatility structure, the slope of the yield curve as well

<sup>32</sup>This finding is reminiscent of the result in Dai and Singleton (2002) that an  $A_1(3)$  reduced-form *DTSM* was unable to resolve the expectations puzzles.

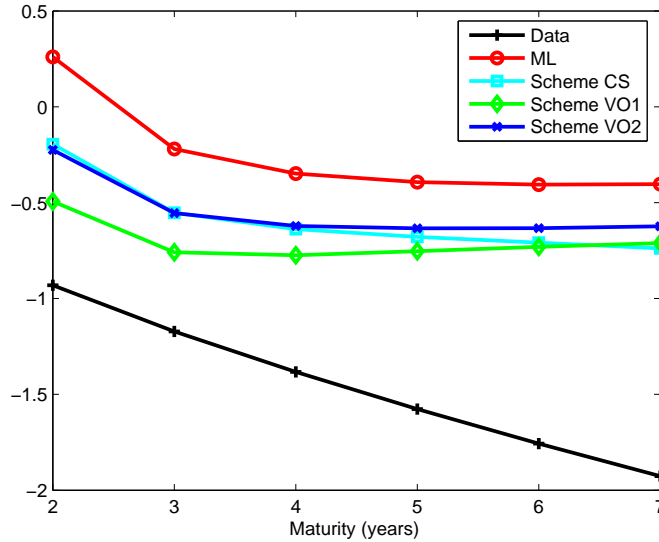


Figure 7: Risk-Premium Adjusted Regression Coefficients

as the CS regression coefficients. Meanwhile, inflation contributes very little to the overall risk-premium dynamics since the price of inflation risk is constant at 1 and the volatility of inflation-specific shocks is constant. Introducing more flexibility along either or both of these dimensions would likely assist the habit-based model in matching the distribution of bond yields.<sup>33</sup>

## 9 Concluding Remarks

In this paper we have argued that, along important dimensions, researchers can gain flexibility and tractability in analyzing *DTSMs* by switching from continuous to discrete time. We have developed a family of nonlinear *DTSMs* 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  $\Lambda_t$  of the risk factors, and (iii) for any admissible specification of  $\Lambda_t$ , the likelihood function of the bond yield data is known in closed form.

This modeling framework was illustrated by estimating a nonlinear (non-affine under  $\mathbb{P}$ ), equilibrium *DTSM* in which agents' preferences exhibit habit formation. A novel feature of our formulation is that we posit an affine <sup>$\mathbb{Q}$</sup>  representation of the state  $X_t$ , and choose the consumption process under the historical measure so that the one-period bond yield is an affine function of  $X_t$ . As such, an *equilibrium* implication of our model is that bond yields are

<sup>33</sup>Gallmeyer, Hollifield, Palomino, and Zin (2007), for example, consider an inflation process that is endogenously determined to satisfy the Taylor rule.

known in closed form, even though preferences are nonlinear and the state exhibits stochastic volatility. The market prices of risk associated with our habit-based preferences imply that the surplus consumption ratio follows a nonlinear (non-affine) process under the historical measure. Nevertheless, the likelihood function of the data is known in closed form.

The tractability of likelihood-based estimation means that our approach offers an attractive alternative estimation strategy to the calibration methods most often applied in the study of equilibrium asset pricing models. As is illustrated in our empirical analysis, calibration can easily lead to parameters that render models equally effective at matching salient features of the macroeconomic series while having fundamentally different implications for asset prices. Focus on the likelihood function provides one, systematic means of incorporating full information about the conditional joint distribution of the macroeconomic variables and asset returns.

Our framework is applicable more generally to other equilibrium *DTSMs* and also offers a means of exploring richer no-arbitrage, reduced form models.<sup>34</sup> Key to this applicability is the presumption that the state variables follow an affine process under  $\mathbb{Q}$ . Many of the current generation of macro-finance models of the term structure either presume an affine <sup>$\mathbb{Q}$</sup>  state process or they are easily reformulated to have this structure (seemingly) without changing their essential properties. We note in particular that this assumption is explicit in many of the macro-finance models of the term structure being developed at central banks (e.g., Rudebusch and Wu (2008) and Hordahl, Tristani, and Vestin (2006)), as well as in models with long-run risks based on the framework in Bansal and Yaron (2004). Our framework provides a means of enriching the data-generating processes in these and related studies.

---

<sup>34</sup> Furthermore, 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 (2005) and Dai, Singleton, and Yang (2005) study *DTSMs* 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*

$$Z_{t+1} = H(Z_t, \epsilon_{t+1}, \theta), \quad (67)$$

where  $\epsilon_t$  is an i.i.d. process. Fix  $\theta$  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

$$\|H(z, \epsilon_1, \theta)\|_q < \delta_\theta \|z\|, \quad \|z\| > K. \quad (68)$$

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

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

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

where  $\epsilon_1$  has a zero mean and unit variance, and  $\Omega(z) = a^{(2)}(\lambda(z)) + b^{(2)}(\lambda(z))z$ . Take  $q = 2$ , we have

$$\frac{\|H(z, \epsilon_1, \theta)\|_2}{\|z\|} \leq \frac{\|a^{(1)}(\lambda(z))\|}{\|z\|} + \frac{\|b^{(1)}(\lambda(z))z\|}{\|z\|} + \frac{\|\sqrt{\Omega(z)}\epsilon_1\|_2}{\|z\|}. \quad (69)$$

The first term on the right-hand-side of (69) satisfies

$$\frac{\|a^{(1)}(\lambda(z))\|}{\|z\|} = \frac{\|\text{vec} \left[ \frac{\nu_i c_i}{1 - \lambda_i(z) c_i} \right]\|}{\|z\|} \leq \frac{\|\text{vec} [\nu_i c_i]\|}{\|z\|} \rightarrow 0, \quad \|z\| \rightarrow \infty, \quad (70)$$

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

Since all elements of  $\rho$  are non-negative, if  $1 - \lambda_i(z) c_i \geq 1$  for all  $z$  and  $i$ , then the second term in (69) 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

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

Finally, the last term in (69) 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  $\sqrt{z}$ .<sup>35</sup>

The only term on the right-hand side of (69) 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  $\delta_\theta$  to satisfy  $\max_i |\psi_i| < \delta_\theta < 1$  so that Lemma 1 is satisfied.

## B Proof of Proposition 3

From equation (46), we have:

$$\begin{aligned}
f^{\mathbb{P}}(z_{t+1}, \pi_{t+1} | z_t, \pi_t) &= f^{\mathbb{Q}}(z_{t+1}, \pi_{t+1} | z_t, \pi_t) \times \frac{e^{\Lambda'_t[z_{t+1}, \pi_{t+1}]'}}{\phi^{\mathbb{Q}}(\Lambda_t; [z_t, \pi_t])} \\
&= f^{\mathbb{Q}}(z_{t+1} | z_t) \times \frac{e^{\Lambda_{z,t} z_{t+1}} E^{\mathbb{Q}}[e^{\pi_{t+1}} | z_{t+1}, z_t, \pi_t]}{\phi^{\mathbb{Q}}(\Lambda_t; [z_t, \pi_t])} \\
&\quad \times f^{\mathbb{Q}}(\pi_{t+1} | z_{t+1}, z_t, \pi_t) \times \frac{e^{\pi_{t+1}}}{E^{\mathbb{Q}}[e^{\pi_{t+1}} | z_{t+1}, z_t, \pi_t]} \\
&= f^{\mathbb{Q}}(z_{t+1} | z_t) \times \frac{e^{(\Lambda_{z,t} - \sigma_{\pi,g}) z_{t+1}}}{E^{\mathbb{Q}}[e^{(\Lambda_{z,t} - \sigma_{\pi,g}) z_{t+1}} | z_t]} \\
&\quad \times f^{\mathbb{Q}}(\pi_{t+1} | z_{t+1}, z_t, \pi_t) \times \frac{e^{\pi_{t+1}}}{E^{\mathbb{Q}}[e^{\pi_{t+1}} | z_{t+1}, z_t, \pi_t]} \tag{71}
\end{aligned}$$

As such, we have:

$$f^{\mathbb{P}}(z_{t+1} | z_t) = f^{\mathbb{Q}}(z_{t+1} | z_t) \times \frac{e^{(\Lambda_{z,t} - \sigma_{\pi,g}) z_{t+1}}}{E^{\mathbb{Q}}[e^{(\Lambda_{z,t} - \sigma_{\pi,g}) z_{t+1}} | z_t]} \tag{72}$$

$$f^{\mathbb{P}}(\pi_{t+1} | z_{t+1}, z_t, \pi_t) = f^{\mathbb{Q}}(\pi_{t+1} | z_{t+1}, z_t, \pi_t) \times \frac{e^{\pi_{t+1}}}{E^{\mathbb{Q}}[e^{\pi_{t+1}} | z_{t+1}, z_t, \pi_t]} \tag{73}$$

### B.1 Regularity of $z_t$

From equation (72),  $z_{t+1}$  follows an autonomous process under  $\mathbb{P}$  with an adjusted market prices of risk of  $\Lambda_{z,t} - \sigma_{\pi,g}$ . For this density to be well-defined, we need to make sure that:<sup>36</sup>

$$1 - (\Lambda_{z,t} - \sigma_{\pi,g}) c_z > 0 \text{ for all } z_t > 0. \tag{74}$$

Substitute  $\Lambda_{z,t} = -\gamma \left(1 + \frac{\sigma_g}{\sigma_t^{\mathbb{Q}}[z_{t+1}]}\right)$ , we have:

$$1 + \left( \gamma \left(1 + \frac{\sigma_g}{\sigma_t^{\mathbb{Q}}[z_{t+1}]}\right) + \sigma_{\pi,g} \right) c_z > 0, \text{ for all } z_t > 0, \tag{75}$$

<sup>35</sup>See Duffie and Singleton (1993) for a discussion of the geometric ergodicity of models in which volatility depends on terms of the form  $x^\gamma$ , 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.

<sup>36</sup>This expression goes under the logarithm operator in the density.

which requires

$$1 + (\gamma + \sigma_{\pi,g})c_z > 0 \quad (76)$$

Applying (40) and (41), we can write down the first two moments of  $z_t$  as follows:

$$\begin{aligned} E^{\mathbb{P}}[z_{t+1}|z_t] &= \frac{v_z c_z}{1 - (\Lambda_{z,t} - \sigma_{\pi,g})c_z} + \frac{\rho_z}{(1 - (\Lambda_{z,t} - \sigma_{\pi,g})c_z)^2} z_t \\ \sigma^{\mathbb{P}}[z_{t+1}|z_t]^2 &= \frac{v_z c_z^2}{(1 - (\Lambda_{z,t} - \sigma_{\pi,g})c_z)^2} + \frac{2c_z \rho_z}{(1 - (\Lambda_{z,t} - \sigma_{\pi,g})c_z)^3} z_t \end{aligned} \quad (77)$$

$z_t$  would be geometrically ergodic, according to Proposition G.E.(Z), if we have the limit of:

$$\frac{\left| \frac{v_z c_z}{1 - (\Lambda_{z,t} - \sigma_{\pi,g})c_z} \right| + \left| \frac{\rho_z}{(1 - (\Lambda_{z,t} - \sigma_{\pi,g})c_z)^2} \right| |z| + \sigma^{\mathbb{P}}[z_{t+1}|z_t]}{|z|} \quad (78)$$

strictly less than 1 as  $z \rightarrow \infty$ .

Since  $\frac{1}{|1 - (\Lambda_{z,t} - \sigma_{\pi,g})c_z|}$  is bounded from condition (76), the first and third terms go to zero in the limit. Condition (78) therefore reduces to

$$\text{Lim}_{z_t \rightarrow \infty} \left| \frac{\rho_z}{(1 - (\Lambda_{z,t} - \sigma_{\pi,g})c_z)^2} \right| < 1 \quad (79)$$

which, in turn, is equivalent to

$$1 + (\gamma + \sigma_{\pi,g})c_z > \sqrt{\rho_z}, \quad (80)$$

an even stronger condition than (76). With little modification, we have

$$\sigma_{\pi,g} > \frac{\sqrt{\rho_z} - 1}{c_z} - \gamma, \quad (81)$$

which is precisely the first equation in Proposition 3.

In addition, to prevent  $z_t$  from being absorbed at zero, under  $\mathbb{Q}$ , a standard requirement is that  $\nu_z \geq 1$ .<sup>37</sup> Intuitively,  $\nu_z$  controls the relative strength of the mean reverting drift that pulls  $z_t$  away from its zero boundary and the diffusive force that could possibly absorb  $z_t$  at zero. To have non-absorbing behavior under  $\mathbb{Q}$ , the former has to be stronger than the latter, which requires  $\nu_z \geq 1$ . Since  $\mathbb{P}$  and  $\mathbb{Q}$  are equivalent measures,  $\nu_z \geq 1$  also guarantees that  $z_t$  is non-absorbing under  $\mathbb{P}$ . Another way of seeing why this is the case is by noting that  $\nu_z$  are the same under both  $\mathbb{P}$  and  $\mathbb{Q}$  as discussed in section 6. Therefore  $\nu_z$  modulates the relative strength of the mean reversion and diffusion forces under both measures.

---

<sup>37</sup>This requirement is very similar in a continuous time setup. For a continuous CIR process characterized by the reversion parameter  $\kappa$ , the long run mean parameter  $\theta$  and the volatility parameter  $\sigma$  to be non-absorbing at zero, the usual constraint is  $\frac{2\kappa\theta}{\sigma^2} \geq 1$ .

## B.2 Regularity of $\pi_t$

From equation (73), it follows that  $\pi_{t+1}$  is Gaussian, conditional on  $z_{t+1}$ ,  $z_t$  and  $\pi_t$  with the first two moments given by:

$$\begin{aligned} E_t^{\mathbb{P}}[\pi_{t+1}|z_{t+1}, z_t, \pi_t] &= \bar{\pi} + \rho_{\pi}(\pi_t - \bar{\pi}) + \rho_{\pi,z}(z_t - E^{\mathbb{Q}}[z_t]) - \sigma_{\pi,g}(z_{t+1} - E_t^{\mathbb{Q}}[z_{t+1}]) + \sigma_{\pi}^2 \\ \sigma_t^{\mathbb{P}}[\pi_{t+1}|z_{t+1}, z_t, \pi_t] &= \sigma_{\pi} \end{aligned} \quad (82)$$

which implies that the auto-regressive coefficients ( $\rho_{\pi}$ ) are the same under both  $\mathbb{P}$  and  $\mathbb{Q}$ . If  $z_t$  is ergodic, therefore, all we need is  $0 < \rho_{\pi} < 1$ .

## C Derivation of Steady State Conditions

### C.1 $\frac{\partial x_{t+1}}{\partial c_{t+1}} = 0 \Big|_{z_t = \bar{z}}$

From the definition of surplus consumption ratio, the following identity must hold:

$$x_{t+1} = c_{t+1} + \log(1 - e^{s_{t+1}}). \quad (83)$$

It follows that

$$\frac{\partial x_{t+1}}{\partial c_{t+1}} = 1 + \frac{\partial s_{t+1}/\partial c_{t+1}}{1 - e^{-s_{t+1}}} = 1 - \frac{\partial z_{t+1}/\partial c_{t+1}}{1 - e^{z_{t+1}-s_{max}}}. \quad (84)$$

Since  $\frac{\partial z_{t+1}}{\partial c_{t+1}} = -\frac{\sigma_t^{\mathbb{Q}}[z_{t+1}]}{\sigma_g}$ , we have:

$$\begin{aligned} \frac{\partial x_{t+1}}{\partial c_{t+1}} &= 1 + \frac{\sigma_t^{\mathbb{Q}}[z_{t+1}]}{\sigma_g(1 - e^{z_{t+1}-s_{max}})} \\ &\approx 1 + \frac{\sigma_t^{\mathbb{Q}}[z_{t+1}]}{\sigma_g(1 - e^{z_t-s_{max}})}. \end{aligned} \quad (85)$$

The approximate relation arises since we exploit the fact that  $z_{t+1} \approx z_t$  around the steady state.

In order to have  $\frac{\partial x_{t+1}}{\partial c_{t+1}} = 0 \Big|_{z_t = \bar{z}}$ , therefore, we need:

$$\sigma_t^{\mathbb{Q}}[z_{t+1}|z_t = \bar{z}] = \sigma_g(e^{\bar{z}-s_{max}} - 1). \quad (86)$$

### C.2 $\frac{\partial(\partial x_{t+1}/\partial c_{t+1})}{\partial z_t} = 0 \Big|_{z_t = \bar{z}}$

First, taking the first order derivative of both sides of equation (85), we have:

$$\frac{\partial(\partial x_{t+1}/\partial c_{t+1})}{\partial z_t} = \frac{\partial \sigma_t^{\mathbb{Q}}[z_{t+1}]}{\partial z_t} \times \sigma_g(1 - e^{z_t-s_{max}}) + \sigma_t^{\mathbb{Q}}[z_{t+1}] (\sigma_g e^{z_t-s_{max}}). \quad (87)$$

Substituting (86) into the above equation, evaluated at the steady state value  $\bar{z}$ , it follows that in order to have  $\frac{\partial(\partial x_{t+1}/\partial c_{t+1})}{\partial z_t} = 0 \Big|_{z_t=\bar{z}}$  we need:

$$\frac{\partial \sigma_t^Q[z_{t+1}]}{\partial z_t} \Big|_{z_t=\bar{z}} = \sigma_g e^{\bar{z}-s_{max}}. \quad (88)$$

### C.3 Steady State Conditions

Together, equations (86) and (88) impose the following constraints:

$$\begin{aligned} \sqrt{\nu_z c_z^2 + 2c_z \rho_z \bar{z}} &= \sigma_g (e^{\bar{z}-s_{max}} - 1) \\ \frac{c_z \rho_z}{\sqrt{\nu_z c_z^2 + 2c_z \rho_z \bar{z}}} &= \sigma_g e^{\bar{z}-s_{max}}. \end{aligned} \quad (89)$$

Denoting  $\mathcal{A} = \frac{\nu_z c_z}{\rho_z} + 2\bar{z}$ , it can be shown that the above system is equivalent to:

$$\mathcal{A} = 1 + \frac{\sigma_g^2}{2c_z \rho_z} - \sqrt{\frac{\sigma_g^2}{c_z \rho_z} + \frac{\sigma_g^4}{4c_z^2 \rho_z^2}} \quad (90)$$

$$s_{max} = \bar{z} + \log(1 - \mathcal{A}). \quad (91)$$

Finally, we define the steady state value of  $z_t$  as a value  $\bar{z}$  such that:

$$E_t^{\mathbb{P}}[z_{t+1} | z_t = \bar{z}] = \bar{z}. \quad (92)$$

Since  $E^{\mathbb{P}}[z_{t+1} | z_t] = \frac{\nu_z c_z}{1 - (\Lambda_{z,t} - \sigma_{\pi,g})c_z} + \frac{\rho_z}{(1 - (\Lambda_{z,t} - \sigma_{\pi,g})c_z)^2} z_t$ , we have:

$$\frac{\nu_z c_z}{1 - (\Lambda_{\bar{z}} - \sigma_{\pi,g})c_z} + \frac{\rho_z}{(1 - (\Lambda_{\bar{z}} - \sigma_{\pi,g})c_z)^2} \bar{z} = \bar{z}, \quad (93)$$

where

$$\Lambda_{\bar{z}} = -\gamma \left( 1 + \frac{\sigma_g}{\sigma_t^Q[z_{t+1}]} \right) \Big|_{z_t=\bar{z}} \quad (94)$$

$$= -\gamma \left( 1 + \frac{1}{e^{\bar{z}-s_{max}} - 1} \right) \quad (95)$$

$$= -\frac{\gamma}{\frac{\nu_z c_z}{\rho_z} + 2\bar{z}} = -\frac{\gamma}{\mathcal{A}}. \quad (96)$$

From  $\mathcal{A} = \frac{\nu_z c_z}{\rho_z} + 2\bar{z}$ , we have:

$$\nu_z = \frac{(\mathcal{A} - 2\bar{z})\rho_z}{c_z}. \quad (97)$$

Substitute  $\nu_z$  in (93) and solve for  $\bar{z}$ , we have

$$\bar{z} = \frac{\mathcal{A}\mathcal{B}\rho_z}{1 + 2\mathcal{B}\rho_z}, \quad (98)$$

where

$$\mathcal{B} = \frac{1 + \left(\frac{\gamma}{\mathcal{A}} + \sigma_{\pi,g}\right) c_z}{\left(1 + \left(\frac{\gamma}{\mathcal{A}} + \sigma_{\pi,g}\right) c_z\right)^2 - \rho_z}. \quad (99)$$

## D Linearity of the Nominal Short Rate

First, according to the Euler equation, the nominal (unannualized) interest rate per unit of time interval is the  $r_t$  such that:

$$e^{-r_t} = E_t^{\mathbb{P}}[e^{m_{t,t+1}}] = E_t^{\mathbb{Q}}[e^{-m_{t,t+1}}]^{-1} \quad (100)$$

The second part of the identity follows from our construction of the risk-neutral densities and how they connect to their physical counter-parts through the market prices of risks. We have:

$$r_t = \log \left( E_t^{\mathbb{Q}}[e^{-m_{t,t+1}}] \right). \quad (101)$$

From equations (18) and (44), we can rewrite the pricing kernel as follows:

$$\begin{aligned} -m_{t,t+1} &= -\log \delta + \bar{\pi} + \rho_{\pi}(\pi_t - \bar{\pi}) + \gamma z_t + \rho_{\pi,z} \left( z_t - \frac{\nu_z c_z}{1 - \rho_z} \right) - \gamma E_t^{\mathbb{Q}}[z_{t+1}] + \sigma_{\pi} \epsilon_{\pi,t+1}^{\mathbb{Q}} \\ &\quad + \gamma f(z_t) + u_{\Lambda} z_{t+1} - u_{\Lambda} E_t^{\mathbb{Q}}[z_{t+1}], \end{aligned} \quad (102)$$

where

$$u_{\Lambda} = -\gamma \left( 1 + \frac{\sigma_g}{\sigma_t^{\mathbb{Q}}[z_{t+1}]} \right) - \sigma_{\pi,g}. \quad (103)$$

Consequently,

$$\begin{aligned} r_t &= -\log \delta + \bar{\pi} + \rho_{\pi}(\pi_t - \bar{\pi}) + \gamma z_t + \rho_{\pi,z} \left( z_t - \frac{\nu_z c_z}{1 - \rho_z} \right) - \gamma E_t^{\mathbb{Q}}[z_{t+1}] + \frac{1}{2} \sigma_{\pi}^2 \\ &\quad + \gamma f(z_t) + \log(E_t^{\mathbb{Q}}[e^{u_{\Lambda} z_{t+1}}]) - u_{\Lambda} E_t^{\mathbb{Q}}[z_{t+1}] \end{aligned} \quad (104)$$

If

$$f(z_t) = \mathcal{C} - (\gamma + \sigma_{\pi,g}) \sigma_g \sigma_t^{\mathbb{Q}}[z_{t+1}] - \frac{1}{\gamma} \log \left( \frac{E_t^{\mathbb{Q}}[e^{u_{\Lambda} z_{t+1}}]}{E_t^{\mathbb{Q}^G}[e^{u_{\Lambda} z_{t+1}}]} \right), \quad (105)$$

then

$$\gamma f(z_t) = \gamma \mathcal{C} - \gamma(\gamma + \sigma_{\pi,g}) \sigma_g \sigma_t^{\mathbb{Q}}[z_{t+1}] - \log(E_t^{\mathbb{Q}}[e^{u_{\Lambda} z_{t+1}}]) + u_{\Lambda} E_t^{\mathbb{Q}}[z_{t+1}] + \frac{1}{2} u_{\Lambda}^2 \sigma_t^{\mathbb{Q}}[z_{t+1}]^2. \quad (106)$$

Therefore:

$$\begin{aligned}
r_t &= -\log \delta + \bar{\pi} + \rho_\pi(\pi_t - \bar{\pi}) + \gamma z_t + \rho_{\pi,z} \left( z_t - \frac{\nu_z c_z}{1 - \rho_z} \right) - \gamma E_t^{\mathbb{Q}}[z_{t+1}] + \frac{1}{2} \sigma_\pi^2 \\
&\quad + \gamma \mathcal{C} - \gamma(\gamma + \sigma_{\pi,g}) \sigma_g \sigma_t^{\mathbb{Q}}[z_{t+1}] + \frac{1}{2} u_\Lambda^2 \sigma_t^{\mathbb{Q}}[z_{t+1}]^2.
\end{aligned} \tag{107}$$

The expression

$$u_\Lambda = -\gamma \left( 1 + \frac{\sigma_g}{\sigma_t^{\mathbb{Q}}[z_{t+1}]} \right) - \sigma_{\pi,g} \tag{108}$$

implies that

$$u_\Lambda \sigma_t^{\mathbb{Q}}[z_{t+1}] = -(\gamma + \sigma_{\pi,g}) \sigma_t^{\mathbb{Q}}[z_{t+1}] - \gamma \sigma_g, \tag{109}$$

so we have

$$\frac{1}{2} u_\Lambda^2 \sigma_t^{\mathbb{Q}}[z_{t+1}]^2 = \frac{1}{2} (\gamma + \sigma_{\pi,g})^2 \sigma_t^{\mathbb{Q}}[z_{t+1}]^2 + \frac{1}{2} \gamma^2 \sigma_g^2 + \gamma(\gamma + \sigma_{\pi,g}) \sigma_g \sigma_t^{\mathbb{Q}}[z_{t+1}]. \tag{110}$$

Therefore

$$\begin{aligned}
r_t &= -\log \delta + \bar{\pi} + \rho_\pi(\pi_t - \bar{\pi}) + \gamma z_t + \rho_{\pi,z} \left( z_t - \frac{\nu_z c_z}{1 - \rho_z} \right) - \gamma E_t^{\mathbb{Q}}[z_{t+1}] + \frac{1}{2} \sigma_\pi^2 \\
&\quad + \gamma \mathcal{C} + \frac{1}{2} \gamma^2 \sigma_g^2 + \frac{1}{2} (\gamma + \sigma_{\pi,g})^2 \sigma_t^{\mathbb{Q}}[z_{t+1}]^2.
\end{aligned} \tag{111}$$

Since the risk neutral mean and variance of  $z_{t+1}$  are linear in  $z_t$ ,

$$E_t^{\mathbb{Q}}[z_{t+1}] = \rho_z z_t + \nu_z c_z \tag{112}$$

$$\sigma_t^{\mathbb{Q}}[z_{t+1}]^2 = 2\rho_z c_z z_t + \nu_z c_z^2, \tag{113}$$

it follows that the short rate is linear in the state variables:

$$r_t = \delta_0 + \delta_z z_t + \delta_\pi \pi_t. \tag{114}$$

Collecting terms, we have:

$$\begin{aligned}
\delta_0 &= -\log \delta + (1 - \rho_\pi) \bar{\pi} - \rho_{\pi,z} \frac{\nu_z c_z}{1 - \rho_z} - \gamma \nu_z c_z + \gamma \mathcal{C} \\
&\quad + \frac{1}{2} \gamma^2 \sigma_g^2 + \frac{1}{2} (\gamma + \sigma_{\pi,g})^2 \nu_z c_z^2 + \frac{1}{2} \sigma_\pi^2
\end{aligned} \tag{115}$$

$$\delta_z = \gamma(1 - \rho_z) + \rho_{\pi,z} + (\gamma + \sigma_{\pi,z})^2 \rho_z c_z \tag{116}$$

$$\delta_\pi = \rho_\pi. \tag{117}$$

## E The Continuous Time Limit

### E.1 $z_t$

From our discussion of the  $DA_N^{\mathbb{Q}}(N)$  in section 3.2,  $z_t$  follows a CIR process in the time limit under the nominal risk-neutral measure  $\mathbb{Q}$ :

$$dz_t = \kappa_z(\theta_z - z_t)dt + \sigma_z\sqrt{z_t}dB_{z,t}^{\mathbb{Q}}. \quad (118)$$

In the limit, the total exposure of the nominal pricing kernel to changes in  $z_{t+1}$  approaches  $-\gamma\left(1 + \frac{\sigma_g^c}{\sigma_z\sqrt{z_t}}\right) - \sigma_{\pi,g}$ . The first term is the market price of z-risk. The second term accounts for the contemporaneous correlation between  $z_{t+1}$  and  $\pi_{t+1}$ . This means the difference between the drifts of  $z_t$  under  $\mathbb{P}$  and  $\mathbb{Q}$  must be:

$$\mu_{z,t}^{\mathbb{P}} - \mu_{z,t}^{\mathbb{Q}} = \sigma_z^2 z_t \left( -\gamma - \sigma_{\pi,g} - \frac{\gamma\sigma_g^c}{\sigma_z\sqrt{z_t}} \right). \quad (119)$$

Under  $\mathbb{P}$ , It follows that  $z_t$  approaches the process

$$dz_t = \left( \kappa_z\theta_z - (\kappa_z + \sigma_z^2(\gamma + \sigma_{\pi,g}))z_t - \gamma\sigma_g^c\sigma_z\sqrt{z_t} \right) dt + \sigma_z\sqrt{z_t}dB_{z,t}^{\mathbb{P}}. \quad (120)$$

### E.2 $\pi_t$

$\pi_t$  will approach the following dynamics under the  $\mathbb{Q}$  measure:

$$d\pi_t = (\kappa_{\pi}\bar{\pi} + \kappa_{\pi,z}\theta_z - \kappa_{\pi}\pi_t - \kappa_{\pi,z}z_t) dt - \sigma_{\pi,g}\sigma_z\sqrt{z_t}dB_{z,t}^{\mathbb{Q}} + \sigma_{\pi}^c dB_{\pi,t}^{\mathbb{Q}}. \quad (121)$$

Since the market price of inflation risk is 1, it follows that

$$dB_{\pi,t}^{\mathbb{Q}} = dB_{\pi,t}^{\mathbb{P}} + \sigma_{\pi}^c dt. \quad (122)$$

In addition, from the previous section, we know:

$$dB_{z,t}^{\mathbb{Q}} = dB_{z,t}^{\mathbb{P}} - (\sigma_z(\gamma + \sigma_{\pi,g})\sqrt{z_t} + \gamma\sigma_g^c)dt. \quad (123)$$

Therefore

$$d\pi_t = \left[ \kappa_{\pi}\bar{\pi} + \kappa_{\pi,z}\theta_z + \sigma_{\pi}^{c2} - \kappa_{\pi}\pi_t - (\kappa_{\pi,z} - \sigma_z^2(\gamma + \sigma_{\pi,g})\sigma_{\pi,g})z_t + \gamma\sigma_{\pi,g}\sigma_g^c\sigma_z\sqrt{z_t} \right] dt - \sigma_{\pi,g}\sigma_z\sqrt{z_t}dB_{z,t}^{\mathbb{P}} + \sigma_{\pi}^c dB_{\pi,t}^{\mathbb{P}}. \quad (124)$$

### E.3 $g_t$

Recalling that

$$f(z_t) = \mathcal{C} - (\gamma + \sigma_{\pi,g})\sigma_g\sigma_t^{\mathbb{Q}}[z_{t+1}] - \frac{1}{\gamma}\log\left(\frac{E_t^{\mathbb{Q}}[e^{u_{\Lambda}z_{t+1}}]}{E_t^{\mathbb{Q}^G}[e^{u_{\Lambda}z_{t+1}}]}\right), \quad (125)$$

if  $\mathcal{C} = \mathcal{C}^c \Delta$ , then in the continuous time limit  $f(z_t)$  approaches:

$$f(z_t) = [\mathcal{C}^c - (\gamma + \sigma_{\pi,g})\sigma_g^c \sigma_z \sqrt{z_t}] dt. \quad (126)$$

Note that the third term of equation (125) disappears in the limit because the two measures  $\mathbb{Q}$  and  $\mathbb{Q}^G$ , by construction, give rise to the same mean and variance - the two moments that matter in a continuous time setup. As a result, in the limit:

$$g_t = d \ln C_t \quad (127)$$

$$= [\mathcal{C}^c - (\gamma + \sigma_{\pi,g})\sigma_g^c \sigma_z \sqrt{z_t}] dt - \sigma_g^c dB_{z,t}^{\mathbb{Q}}. \quad (128)$$

Again, applying

$$dB_{z,t}^{\mathbb{Q}} = dB_{z,t}^{\mathbb{P}} - (\sigma_z(\gamma + \sigma_{\pi,g})\sqrt{z_t} + \gamma\sigma_g^c)dt, \quad (129)$$

we have

$$g_t = (\mathcal{C}^c + \gamma\sigma_g^{c2})dt - \sigma_g^c dB_{z,t}^{\mathbb{P}}. \quad (130)$$

## References

- Ahn, D., Dittmar, R., Gallant, A., 2002. Quadratic term structure models: Theory and evidence. *Review of Financial Studies* 15, 243–288.
- Ait-Sahalia, Y., 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., 2005. The term structure of real rates and expected inflation. Working paper, Columbia University.
- Ang, A., Dong, S., Piazzesi, M., 2007. No-arbitrage taylor rules. Unpublished working paper. Columbia University.
- Ang, A., Piazzesi, M., 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics* 50, 745–787.
- Bansal, R., Kiku, D., Yaron, A., 2007. Risks for the long run: Estimation and inference. Unpublished working paper. Duke University.
- Bansal, R., Shaliastovich, I., 2007. Risk and return in bond, currency and equity markets. Unpublished working paper. Duke University.
- Bansal, R., Yaron, A., 2004. Risks for the long run: A potential resolution of asset pricing puzzles. *Journal of Finance* 59, 1481–1509.

- Beaglehole, D. R., Tenney, M. S., 1991. General solutions of some interest rate-contingent claim pricing equations. *Journal of Fixed Income* September, 69–83.
- Bekaert, G., Cho, S., Moreno, A., 2006. New-keynesian macroeconomics and the term structure. Unpublished working paper. Columbia University.
- Boudoukh, J., 1993. An equilibrium model of nominal bond prices with inflation-output correlation and stochastic volatility. *Journal of Money, Credit and Banking* 25, 636–665.
- 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.
- Campbell, J., Cochrane, J., 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy* 107, 205–251.
- Campbell, J., Shiller, R., 1987. Cointegration and tests of present value models. *Journal of Political Economy* 95, 1062–1087.
- Cheng, P., Scaillet, O., 2002. Linear-quadratic jump-diffusion modelling with application to stochastic volatility. .
- Cheridito, R., Filipovic, D., Kimmel, R., 2005. Market price of risk in affine models: Theory and evidence. forthcoming, *Journal of Financial Economics*.
- Collin-Dufresne, P., Goldstein, R., Jones, C., 2008. Identification of maximal affine term structure models. Unpublished working paper. forthcoming, *Journal of Finance*.
- Constantinides, G., Ghosh, A., 2008. Asset pricing tests with long run risks in consumption growth. Unpublished working paper. Graduate School of Business, University of Chicago.
- Cox, J., Ingersoll, J., Ross, S., 1985. A theory of the term structure of interest rates. *Econometrica* 53, 385–407.
- Dai, Q., Philippon, T., 2005. Fiscal policy and the term structure of interest rates. Unpublished working paper. NYU and UNC.
- Dai, Q., Singleton, K., 2000. Specification analysis of affine term structure models. *Journal of Finance* 55, 1943–1978.
- , 2002. Expectations puzzles, time-varying risk premia, and affine models of the term structure. *Journal of Financial Economics* 63, 415–441.
- , 2003. Term structure dynamics in theory and reality. *Review of Financial Studies* 16, 631–678.

- Dai, Q., Singleton, K., Yang, W., 2005. Regime shifts in a dynamic term structure model of u.s. treasury bond yields. Working Paper, Stanford University.
- Darolles, S., Gouriéroux, C., Jasiak, J., 2006. Structural laplace transforms and compound autoregressive processes. forthcoming, *Journal of Time Series Analysis*.
- Duan, J., Simonato, J., 1999. Estimating and testing exponential-affine term structure models by kalman filter. *Review of Quantitative Finance and Accounting* 13, 111–135.
- Duarte, J., 2004. Evaluating an alternative risk preference in affine term structure models. *Review of Financial Studies* 17, 379–404.
- Duffee, G. R., 2002. Term premia and interest rate forecasts in affine models. *Journal of Finance* 57, 405–443.
- Duffie, D., Filipovic, D., Schachermayer, W., 2003. Affine processes and applications in finance. *Annals of Applied Probability* 13, 984–1053.
- Duffie, D., Kan, R., 1996. A yield-factor model of interest rates. *Mathematical Finance* 6, 379–406.
- Duffie, D., Pan, J., Singleton, K., 2000. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* 68, 1343–1376.
- Duffie, D., Pedersen, L., Singleton, K., 2003. Modeling credit spreads on sovereign debt: A case study of russian bonds. *Journal of Finance* 55, 119–159.
- Duffie, D., Singleton, K., 1993. Simulated moments estimation of markov models of asset prices. *Econometrica* 61, 929–952.
- Engsted, T., Moller, S., 2008. An iterated gmm procedure for estimating the campbell-cochrane habit formation model, with an application to danish stock and bond returns. Unpublished working paper. CREATES Research Paper No. 2008-12.
- Esscher, F., 1932. On the probability function in the collective theory of risk. *Skandinavisk Aktuarietidskrift* 15, 175–195.
- Fuhrer, J., 2000. Habit formation in consumption and its implications for monetary-policy models. *American Economic Review* 90, 367–390.
- Gallant, A. R., Tauchen, G., 1996. Which moments to match?. *Econometric Theory* 12, 657–681.
- Gallmeyer, M., Hollifield, B., Palomino, F., Zin, S., 2007. Bond pricing, habits, and a simple policy rule. Unpublished working paper. Canergie Mellon University.
- Gerber, H., Shiu, E., 1994. Option pricing by esscher transforms. *Transactions of Society of Actuaries* 46, 99–140.

- Gourieroux, C., Jasiak, J., 2006. Autoregressive gamma processes. forthcoming, *Journal of Forecasting*.
- Gourieroux, C., Monfort, A., Polimenis, V., 2002. Affine term structure models. Working paper, University of Toronto, Canada.
- Gurkaynak, R., Sack, B., Wright, J., 2006. The u.s. treasury yield curve: 1961 to the present. Unpublished working paper. FEDS Working Papers.
- Hordahl, P., Tristani, O., Vestin, D., 2006. A joint econometric model of macro-economic and term structure dynamics. *Journal of Econometrics* 131, 405–444.
- , 2007. The yield curve and macroeconomic dynamics. Working Paper, European Central Bank.
- Joslin, S., 2007. Pricing and hedging volatility in fixed income markets. Unpublished working paper. Working Paper, MIT.
- Kim, D. H., 2008. Challenges in macro-finance modeling. Unpublished working paper. Federal Reserve Board, Washington D.C.
- 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.
- 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.
- Nummelin, E., Tuominen, P., 1982. Geometric ergodicity of harris recurrent markov chains with applications to renewal 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.
- Piazzesi, M., Schneider, M., 2007. Equilibrium yield curves. in *NBER Macroeconomics Annual*, ed. by D. Acemoglu, K. Rogoff, and M. Woodford. MIT Press, Cambridge.
- Rudebusch, G., Wu, T., 2008. A macro-finance model of the term structure, monetary policy, and the economy. *Economic Journal* 118, 906–926.
- 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.
- Verdelhan, A., 2008. A habit-based explanation of the exchange rate risk premium. *Journal of Finance*.
- Wachter, J., 2005. A consumption-based model of the term structure of interest rates. forthcoming, *Journal of Financial Economics*.
- Wu, S., 2008. Consumption risk and the real yield curve. Unpublished working paper. The University of Kansas.
- Wu, T., 2005. Macro factors and the affine term structure of interest rates. forthcoming, *Journal of Money Credit and Banking*.