

# Expectation Puzzles, Time-varying Risk Premia, and Affine Models of the Term Structure

Qiang Dai and Kenneth J. Singleton\*

First version: June 7, 2000,  
This version: April 30, 2001

## Abstract

Though linear projections of returns on the slope of the yield curve have contradicted the implications of the traditional “expectations theory,” we show that these findings are not puzzling relative to a large class of richer dynamic term structure models. Specifically, we are able to match all of the key empirical findings reported by Fama and Bliss and Campbell and Shiller, among others, within large subclasses of affine and quadratic-Gaussian term structure models. Additionally, we show that certain “risk-premium adjusted” projections of changes in yields on the slope of the yield curve recover the coefficients of unity predicted by the models. Key to this matching are parameterizations of the market prices of risk that let the risk factors affect the market prices of risk directly, and not only through the factor volatilities. The risk premiums have a simple form consistent with Fama’s findings on the predictability of forward rates, and are also shown to be consistent with interest rate, feedback rules used by a monetary authority in setting monetary policy.

---

\*Stern School of Business, New York University, qdai@stern.nyu.edu; Graduate School of Business, Stanford University and NBER, ken@future.stanford.edu, respectively. We would like to thank Dave Backus, Greg Duffee, and Robert Engle for helpful comments, and the Gifford and Vivian Fong Financial Engineering Fund at Stanford University for Financial Support.

# 1 Introduction

Fama (1984b, 1984a) and Fama and Bliss [1987] present evidence of rich patterns of variation in expected returns across time and maturities that “stand as challenges or ‘stylized facts’” (Fama [1984b], page 545) to be explained by dynamic term structure models (*DTSMs*). A large literature has subsequently elaborated on the inconsistency of these patterns with the implications of the traditional expectations hypothesis – there is compelling evidence from yield (Campbell and Shiller [1991]) and forward-rate (Backus, Foresi, Mozumdar, and Wu [1997]) regressions for *time-varying* risk premiums. Still largely unresolved, however, is the broader question of whether, taken together, these historical patterns are “puzzling” within richer *DTSMs*, including those commonly implemented by academics and practitioners.

This paper takes up Fama’s challenge and uses several key *stylized facts* about excess returns on bonds to “draw out” the essential features of *DTSMs* that allow us to explain these stylized facts. Letting  $P_t^n$  denote the price of an  $n$ -period zero-coupon bond,  $R_t^n$  ( $\equiv -\ln P_t^n/n$ ) its corresponding yield, and  $r_t \equiv R_t^1$ , the empirical evidence shows that the estimated coefficients  $\phi_{nT}$  in the linear projections of  $R_{t+1}^{n-1} - R_t^n$  onto  $\frac{1}{n-1}(R_t^n - r_t)$  are negative and increasingly so with larger maturity  $n$ . This finding is often viewed as a puzzle by term structure modelers because, under the assumption of constant risk premiums, the expectations hypothesis implies that (in the population) the projection coefficients  $\phi_n$  are unity, for all  $n$ . We show that this pattern for  $\phi_{nT}$  is in fact not puzzling, but rather is generated by a large subclass (though not all) of affine *DTSMs* (Duffie and Kan [1996] and Dai and Singleton [2000] (hereafter *DS*)).<sup>1</sup>

More precisely, we document that the risk premiums, and associated expected excess holding period returns  $e_t^n \equiv E_t[\ln(P_{t+1}^{n-1}/P_t^n) - r_t]$ , implied by many affine models have the properties that:

- LPY:** (i) the pattern of *sample*  $\phi_{nT}$  is largely matched by the pattern of *model-implied population* coefficients  $\phi_n$  from the projections of  $R_{t+1}^{n-1} - R_t^n$  onto  $\frac{1}{n-1}(R_t^n - r_t)$ ;<sup>2</sup>
- (ii) *sample* coefficients  $\phi_{nT}^R$  from projections of the “risk-premium adjusted” yield changes  $R_{t+1}^{n-1} - R_t^n + \frac{1}{n-1}e_t^n$  onto  $\frac{1}{n-1}(R_t^n - r_t)$  that are insignificantly different than their model-implied population values of  $\phi_n^R = 1$ , for all  $n > 1$ .

Matching *LPY*(i) says that the *DTSM* describes the historical behavior of yields under the actual measure  $P$ , while matching *LPY*(ii) says that the *DTSM* essentially has the dynamics “right” under the risk-neutral measure  $Q$  used in pricing bonds. We show that these properties are not equivalent, and that the requirement that a *DTSM* match both is a powerful discriminator among models.

---

<sup>1</sup>We focus on affine models in part because the entire family of affine *DTSMs* imply that optimal forecasts of excess returns take the form of the linear projections extensively studied in the literature on the expectations hypothesis.

<sup>2</sup>We are presuming that the pattern of  $\phi_{nT}$  is not spurious, but rather is representative of the pattern of the “true” population  $\phi_n$ . As demonstrated by Bekaert, Hodrick, and Marshall [1997a] and Backus, Foresi, Mozumdar, and Wu [1997], the pattern of  $\phi_{nT}$  cannot be attributed to small-sample biases. Indeed, they find that the small-sample bias reinforces the puzzle by making the sample projection coefficients less negative than they would be in the absence of such bias.

The expectations puzzle, repeated in Table 1 for our treasury data set,<sup>3</sup> was anticipated by Fama [1984a] and Fama and Bliss [1987] who argued that excess returns are time-varying and typically *positively* correlated with the slope of the yield curve.<sup>4</sup> Matching *LPY*(i) requires that the population  $\phi_n$  generated by a *DTSM* largely match the downward sloping pattern in Table 1. We insist on matching the population  $\phi_n$  to the pattern in Table 1, because (as we document subsequently) this is a much more demanding requirement than that of satisfying *LPY*(i) using *fitted* yields from a *DTSM*.

Table 1: Campbell-Shiller Long Rate Regression

Estimated slope coefficients  $\phi_{nT}$  from the indicated linear projections using the smoothed Fama-Bliss data set. The maturities  $n$  are given in months and “s.e.” is the estimated standard error of  $\phi_{nT}$ .

	$R_{t+1}^{(n-1)} - R_t^n = \text{constant} + \phi_{nT}(R_t^n - r_t)/(n-1) + \text{residual}$									
Maturity	3	6	9	12	24	36	48	60	84	120
$\phi_{nT}$	-0.428	-0.883	-1.228	-1.425	-1.705	-1.190	-2.147	-2.433	-3.096	-4.173
s.e.	(.481)	(.640)	(.738)	(.825)	(1.120)	(1.295)	(1.418)	(1.519)	(1.705)	(1.985)

Much less attention has been given to requirement *LPY*(ii). We show formally below that, for *any* *DTSM* in which time variation in expected excess returns is due to time-varying risk premiums, there is a function  $D_{t+1}^{*n}$  of market risk premiums with the properties that  $e_t^n = E_t[D_{t+1}^{*n}]$  and the “risk-premium adjusted” *population* projection coefficients  $\phi_n^{\mathcal{R}}$  are unity. That is, once we adjust the yield changes  $R_{t+1}^{n-1} - R_t^n$  by  $\frac{1}{n-1}D_{t+1}^{*n}$ , we recover the coefficient of unity on  $\frac{1}{n-1}(R_t^n - r_t)$  desired by proponents of the expectations hypothesis. However, the corresponding *sample*  $\phi_{nT}^{\mathcal{R}}$ , obtained using historical yields for  $R_t^n$  and model-fitted risk premiums in constructing  $D_{t+1}^{*n}$ , will typically be close to unity only if the *DTSM* accurately describes the dynamic behavior of risk premiums; that is, only if the model captures the behavior of yields under the risk-neutral measure  $Q$ .

Matching both *LPY*(i) and (ii) simultaneously places substantial demands on specifications of the market prices of risk, correlations among the risk factors determining  $r$ , and the volatilities of these factors. This is particularly true if we insist that the model match other features of the conditional distributions of bond yields, say those summarized by the (model-implied) likelihood function of the data. Key to our success at matching *LPY* is the flexibility of our specification of the factor “market prices of risk”: we posit market prices of risk that depend not only on the factor volatilities, but also on (at least some of) the risk factors directly. For affine *DTSMs*, this specification follows Duffee [2000] in extending those affine models (see *DS* and the references therein) in which market prices of risk are proportional to factor volatilities alone.

Our empirical analysis focuses on three-factor models, motivated in part by that finding in Litterman and Scheinkman [1991] that three principle components capture well over 90% of

<sup>3</sup> We are grateful to Backus, Foresi, Mozumdar, and Wu [1997] for providing the smoothed Fama-Bliss data used in our analysis. The data are monthly from February, 1970 through December, 1995.

<sup>4</sup>Fama and Bliss [1987] focused on the slope of the forward rate curve, but as we shall see subsequently the basic intuition from their analysis carries over to the slope  $R_t^n - r_t$ .

the variation in U.S. treasury yields. All four of the canonical affine three-factor models (see *DS* and Duffee [2000]) are fit by the method of full-information *ML* and then the relevant population and sample versions of the risk-adjusted ( $\phi_n^{\mathcal{R}}$ ) and unadjusted ( $\phi_n$ ) projection coefficients are compared. Additionally, we assess the relative importance of time-varying risk premiums, time-varying factor volatilities, and non-zero factor correlations in matching *LPY*.

We show in Section 4 that a three-factor *CIR*-style model (in which the state variables follow square-root diffusions) is wholly incapable of matching *LPY*. We attribute this model failure to the constraint in *CIR*-style models that risk premiums are proportional to factor volatilities, so it is only through time-varying volatilities that risk premiums can vary. This finding complements those in Roberds and Whiteman [1999] who show (see their Figures 4 and 6) that one- and two-factor *CIR*-style models cannot match *LPY*(i) for their sample period and treasury yields, even when the parameters of their *DTSMs* are calibrated to match their counterparts of the  $\phi_{nT}$ . They also complement Backus, Foresi, Mozumdar, and Wu [1997]’s analytic demonstration that, in order for a (one-factor) *CIR*-style model potentially to match *LPY*(i), it must imply a *downward* sloping term structure of mean forward spreads contrary to historical experience.

In contrast, we find that multi-factor Gaussian models match *LPY* strikingly well at the historical *ML* estimates of the model. What is striking about this result is that standard formulations of Gaussian affine models have a constant market price of risk (see, e.g., Vasicek [1977] and *DS*) which implies that the expectations hypotheses  $\phi_n = 1$  is *true*! Therefore, it must be our (affine) state-dependent formulation of the market price of risk that leads to our success.<sup>5</sup>

Lying between the cases of *CIR*-style and Gaussian models are the affine models in which the number state variables driving factor volatilities,  $m$ , satisfies  $0 < m < 3$ . The  $m$  volatility factors have market prices of risk that are proportional to their respective volatilities as in *CIR*-style models, while the risk premiums of the remaining  $3 - m$  “non-volatility” factors may depend directly on these non-volatility risk factors. In these intermediate cases we have mixed success in matching *LPY*. Further exploration of the reasons reveals a tension in matching simultaneously the historical properties of the conditional means and variances of yields within affine *DTSMs*.

The remainder of this paper is organized as follows. In Section 2 we derive our fundamental “risk-premium adjusted” yield and forward rate projections that serve as the basis of our subsequent econometric analysis. Section 3 discusses in more depth our parameterizations of the market prices of risk and their link to *LPY*. The empirical assessment of the fit of three-factor affine *DTSMs* to *LPY* is presented in Section 4. Concluding remarks, and conjectures about other families of *DTSMs* that might generate the patterns *LPY*, are presented in Section 5. Some properties of the Gaussian family of models exploited in our analysis are collected in the Appendix.

---

<sup>5</sup>Fisher [1998] independently proposed a similar potential resolution of the puzzle *LPY* within a two-factor Gaussian model. However, he does not compare the model-implied and historical projection coefficients ( $\phi_n, \phi_n^{\mathcal{R}}$ ), as is done subsequently here in Section 4, to assess whether extended Gaussian models quantitatively match *LPY*. We are grateful to Greg Duffee for bringing this unpublished manuscript to our attention.

## 2 Risk-Premium Adjusted Projections

If the empirical failure of expectations hypothesis is due to time-varying risk premiums, then it would seem that accommodating risk premiums in these projection equations should restore slope coefficients of one. We begin our exploration of the links between *LPY* and *DTSMs* by showing a precise sense in which this intuition is correct. The resulting risk-premium adjusted projection equations serve as the fundamental relations underlying our subsequent empirical analysis.

### 2.1 Yield Projections

Letting  $D_{t+1}^n = \left( \ln \frac{P_{t+1}^{n-1}}{P_t^n} - r_t \right)$  denote the one-period excess return on an  $n$ -period bond, then from the basic price-yield relation, the *expected* excess return  $e_t^n \equiv E_t[D_{t+1}^n]$  can be expressed as

$$e_t^n = -(n-1)E_t[R_{t+1}^{n-1} - R_t^n] + (R_t^n - r_t), \quad (1)$$

where  $E_t$  denotes expectation conditioned on date  $t$  information. Rearranging (1) gives the fundamental relation<sup>6</sup>

$$E_t \left[ R_{t+1}^{n-1} - R_t^n + \frac{1}{n-1} D_{t+1}^n \right] = \frac{1}{n-1} (R_t^n - r_t). \quad (2)$$

There is no economic content to (2) as it holds by definition *even without the expectation operator*. Economic content is added by linking  $E_t[D_{t+1}^n]$  to the risk premiums implied by an economic model. Toward this end, we introduce two related notions of “term premiums:” the yield term premium

$$c_t^n \equiv R_t^n - \frac{1}{n} \sum_{i=0}^{n-1} E_t[r_{t+i}], \quad (3)$$

and the forward term premium

$$p_t^n \equiv f_t^n - E_t[r_{t+n}], \quad (4)$$

where  $f_t^n \equiv -\ln(P_t^{n+1}/P_t^n)$  denotes the forward rate for one-month loans commencing at date  $t+n$ . Since  $R_t^n \equiv \frac{1}{n} \sum_{i=0}^{n-1} f_t^i$ , the term premiums  $p_t^n$  and  $c_t^n$  are linked by the simple relation:

$$c_t^n \equiv \frac{1}{n} \sum_{i=0}^{n-1} p_t^i. \quad (5)$$

---

<sup>6</sup>Expression (2) is formally equivalent to equation (11) of Fama and Bliss [1987], which, in our notation, is:

$$E_t \left[ R_{t+1}^{n-1} - R_t^{n-1} + \frac{1}{n-1} D_{t+1}^n \right] = \frac{1}{n-1} (f_t^{n-1} - r_t).$$

We focus on (2) because it is more directly linked to the yield regressions in Campbell and Shiller [1991].

Throughout our analysis we assume that these variables are stationary stochastic processes with finite first and second moments.

The *realized* excess return  $D_{t+1}^n$  can be decomposed into a pure “premium” part,  $D_{t+1}^{*n}$ , and an “expectations” part:<sup>7</sup>

$$D_{t+1}^n = D_{t+1}^{*n} + \sum_{i=1}^{n-1} (E_t r_{t+i} - E_{t+1} r_{t+i}), \text{ where} \quad (6)$$

$$D_{t+1}^{*n} = -(n-1)(c_{t+1}^{n-1} - c_t^{n-1}) + p_t^{n-1}. \quad (7)$$

Since the  $(E_t r_{t+i} - E_{t+1} r_{t+i})$  have zero date- $t$  conditional means,  $e_t^n$  depends only on the premium term  $D_{t+1}^{*n}$ :

$$e_t^n = E_t[D_{t+1}^{*n}] = -(n-1)E_t[c_{t+1}^{n-1} - c_t^{n-1}] + p_t^{n-1}. \quad (8)$$

Equation (8) implies that  $E[e_t^n] = E[p_t^{n-1}] = E[f_t^{n-1} - r_t]$ , where the second equality follows from the definition of  $p_t^{n-1}$  and the stationarity of  $r_t$ . This equality seems to have been largely overlooked in the extant literature on the expectations hypothesis. For instance, Fama [1984b], drawing on results from Fama [1976], uses the relation (his equation (5) expressed in our notation)

$$p_t^{n-1} = E_t[D_{t+1}^n] + E_t[D_{t+2}^{n-1} - D_{t+1}^{n-1}] + \dots + E_t[D_{t+n-1}^2 - D_{t+n-2}^2] \quad (9)$$

to conclude that the forward rate  $f_t^{n-1}$  “contains” market expectations about the holding period return  $D_{t+1}^n$ . He then computed the sample means of  $p_t^{n-1}$  and  $(f_t^{n-1} - r_t)$  and expressed surprise at the finding that they were nearly the same (Fama [1984b], page 544). In fact, in the population, they are by definition the same.

In light of (8), we can replace  $D_{t+1}^n$  by  $D_{t+1}^{*n}$  in (2) to obtain

$$E_t \left[ R_{t+1}^{n-1} - R_t^n + \frac{1}{n-1} D_{t+1}^{*n} \right] = \frac{1}{n-1} (R_t^n - r_t). \quad (10)$$

From (10) it follows that the projection of the “premium-adjusted” change in yields,

$$R_{t+1}^{n-1} - R_t^n - (c_{t+1}^{n-1} - c_t^{n-1}) + \frac{1}{n-1} p_t^{n-1}, \quad (11)$$

onto the (scaled) slope of the yield curve,  $(R_t^n - r_t)/(n-1)$ , has a coefficient of one.<sup>8</sup> Yield-based regressions under the expectations hypothesis are obtained by setting the risk premiums in (11) to constants.

---

<sup>7</sup>Some of the intermediate steps in this derivation are:

$$\begin{aligned} D_{t+1}^n &\equiv nR_t^n - (n-1)R_{t+1}^{n-1} - r_t = nc_t^n - (n-1)c_{t+1}^{n-1} + \sum_{i=1}^{n-1} (E_t r_{t+i} - E_{t+1} r_{t+i}) \\ &= -(n-1)(c_{t+1}^{n-1} - c_t^{n-1}) + \sum_{j=0}^{n-1} p_t^j - \sum_{j=0}^{n-2} p_t^j + \sum_{i=1}^{n-1} (E_t r_{t+i} - E_{t+1} r_{t+i}). \end{aligned}$$

<sup>8</sup> There is an analogous set of yield projections for the forward rates:  $E_t[f_{t+1}^{n-1} - r_t] = (f_t^n - r_t) + (E_t[p_{t+1}^{n-1}] - p_t^n)$ . Thus, projection of the “premium-adjusted” forward rate,  $(f_{t+1}^{n-1} - r_t - (p_{t+1}^{n-1} - p_t^n))$ , onto  $(f_t^n - r_t)$  also gives a slope coefficient of one. In our empirical analysis we will focus on (10).

### 3 Risk Premiums, *DTSMs*, and *LPY*

The challenges set forth by Fama and from studies of the expectations hypothesis are about correlations among yields and, as such, are naturally studied using linear projections. Therefore, in attempting to match *LPY* within *DTSMs* we focus on models in which conditional expectations are linear in known functions of the state vector, a feature shared by both affine and quadratic-Gaussian *DTSMs*.

Consider first the case of affine *DTSMs* with the instantaneous short rate given by  $r(t) = a_0 + b'_0 Y(t)$  and the  $N$ -dimensional state vector  $Y$  following, under measure  $P$ , the affine diffusion

$$dY(t) = \kappa(\theta - Y(t)) dt + \Sigma\sqrt{S(t)}dW(t), \quad (12)$$

where  $W(t)$  is an  $N$ -dimensional vector of independent standard Brownian motions and  $S(t)$  is a diagonal matrix with the  $i^{\text{th}}$  diagonal element given by

$$[S(t)]_{ii} = \alpha_i + \beta'_i Y(t). \quad (13)$$

The risk-neutral representation of  $Y(t)$  used in pricing is obtained by subtracting  $\Sigma\sqrt{S(t)}\Lambda(t)$  from the drift of (12), where  $\Lambda(t)$  is the vector of “market prices of risk.” Standard formulations of affine *DTSMs* “close” this model by assuming that  $\Lambda(t)$  is proportional to  $\sqrt{S(t)}$ :

$$\Lambda(t) = \sqrt{S(t)}\ell_0, \quad (14)$$

where  $\ell_0$  is an  $N \times 1$  vector of constants. To assure the admissibility of an affine model—that it generate well-defined bond prices—we follow *DS* and work within their admissible subfamilies of models  $A_m(N)$ , where an admissible affine model is in  $A_m(N)$  if it has  $m$  state variables driving all  $N$  conditional variances  $[S(t)]_{ii}$  (more precisely, the rank of  $(\beta_1, \dots, \beta_N)$  is  $m$ ).

Since at the heart of matching *LPY* is the specification of the factor risk premiums, additional flexibility is obtained by following Duffee [2000] and extending the specification of  $\Lambda(t)$  in  $A_m(N)$  models, for  $m < N$ , to satisfy

$$\Lambda(t) = \sqrt{S(t)}\lambda^0 + \sqrt{S^-(t)}\lambda^Y Y(t), \quad (15)$$

where  $\lambda^0$  is an  $N \times 1$  vector and  $\lambda^Y$  is a  $N \times N$  matrix of constants; and the diagonal matrix  $S^-(t)$  has zeros in its first  $m$  diagonal entries and  $1/(\alpha_i + \beta'_i Y(t))$  in entries  $i = m+1, \dots, N$ , under the presumption that  $\inf(\alpha_i + \beta'_i Y(t)) > 0$ . The case of  $\lambda^Y = 0$  corresponds to the standard risk premium specification (14). That the added flexibility of  $\lambda^Y \neq 0$  translates directly into flexibility in matching *LPY* follows from the observation that the *instantaneous* expected excess return on a  $\tau$ -period zero-coupon bond is

$$\mu^e(t, \tau) = -B(\tau)' \Sigma \sqrt{S(t)} \Lambda(t), \quad (16)$$

where  $B(\tau)$  is the “factor loading” on the state vector from the affine pricing relation  $P_t^\tau = e^{-A(\tau) - B(\tau)' Y_t}$ . Clearly the specification of  $\Lambda(t)$  can have a significant effect on the model-implied properties of  $\mu^e$  and, hence, the matching of *LPY*.

At the same time, we see that the dynamic properties of excess returns are also influenced by the degree of factor correlations  $\Sigma$ , the nature of the factor volatilities  $S(t)$ , and the factor dynamics as reflected in the factor loadings  $B(\tau)$ . Moreover, there is an important interaction effect between  $S(t)$  and  $\Lambda(t)$ : the richer the (admissible) specification of factor volatilities (equivalently, the larger is  $m$ ), the less flexibility there is in specifying  $\Lambda(t)$ . In the case of  $m = N$  (*CIR*-style models), admissibility requires that  $\lambda^Y$  in (15) is zero so the risk premiums are constrained to be of the form (14). For the cases  $m < N$ , then the first  $m$  rows and columns of  $\lambda^Y$  are set to zero to assure admissibility. Thus, maximal flexibility for having state-dependent market prices of risk is obtained in the case of  $m = 0$  where  $\lambda^Y$  is unconstrained. One of the issues we explore empirically is the relative contributions of non-diagonal  $\Sigma$  and non-zero elements of  $\lambda^Y$  to matching *LPY* in multi-factor models.

Our strategy for assessing whether a specific *DTSM* can match *LPY* is to estimate the model parameters; compute the model-implied  $p_t^n$  and  $c_t^n$ , evaluated at the estimated parameters; and, finally, to examine the relevant term structures of projection coefficients. For *LPY*(i), we compare the model-implied population coefficients,

$$\phi_n \equiv \frac{\text{cov}(R_{t+1}^{n-1} - R_t^n, (R_t^n - r_t)/(n-1))}{\text{var}((R_t^n - r_t)/(n-1))}, \quad (17)$$

to their sample counterparts displayed in Table 1. The population  $\phi_n$  are computed by treating the estimates of the model parameters as “truth” and then using analytic formulas to compute the second moments in (17). The data enters these calculations only indirectly through the estimates of the model parameters.

For *LPY*(ii), we examine whether the sample counterparts,  $\phi_{nT}^{\mathcal{R}}$ , of the coefficients

$$\phi_n^{\mathcal{R}} \equiv \frac{\text{cov}(R_{t+1}^{n-1} - R_t^n + D_{t+1}^{*n}/(n-1), (R_t^n - r_t)/(n-1))}{\text{var}((R_t^n - r_t)/(n-1))} \quad (18)$$

are statistically different from a horizontal line at 1, the model-implied values of  $\phi_n^{\mathcal{R}}$ . Here we use historical yields  $R_t^n$  and model-implied  $D_{t+1}^{*n}$ , where the latter are computed by evaluating the expected excess returns at the fitted state variables. The resulting sample  $\phi_{nT}^{\mathcal{R}}$  will be close to one if the sample correlations between the fitted premium terms  $(c_{t+1}^{n-1} - c_t^{n-1}) + p_t^{n-1}/(n-1)$  and the slopes  $(R_t^n - r_t)/(n-1)$  offset the negative coefficients in Table 1.

Comparing (17) and (18), we see that *LPY*(i) and *LPY*(ii) embody distinct sets of economic restrictions. Which (if either) of these patterns is well matched by a particular *DTSM* will depend in part on the structure of the model and in part on the criterion function used in estimation (i.e., on which moments are well matched in estimation). Since *LPY* refers to the properties of the first moments of yields, and the family of Gaussian models (family  $A_0(3)$ ) gives the most flexibility to the structure of factor correlations and conditional means, one might conjecture *a priori* that these models would perform at least as well as other affine models. Similarly, though *CIR*-style model offer the most flexibility with regard to factor volatilities, they are the most constrained with regard to the correlation structure of the factors and the specification of risk premiums, so they might be expected to perform relatively poorly. Beyond these general remarks, the dimensionality of our models and the richness of the moments used in *ML* estimation seem to preclude more precise theoretical statements about which models will best match *LPY*(i) or *LPY*(ii).



Finally, to link our analysis to early empirical studies of excess returns, notice that, from (2) and (8), we can write

$$E_t [R_{t+1}^{n-1} - R_t^n] = \frac{1}{n-1}(R_t^n - r_t) + E_t [c_{t+1}^{n-1} - c_t^{n-1}] - \frac{1}{n-1}p_t^{n-1}. \quad (19)$$

Thus, given parameterizations of the  $p_t^n$ , we have fully determined the yield projections as well as the dependence of  $e_t^n = E_t[D_{t+1}^n]$  on agents' information set. Within the affine family of models,  $p_t^n$  is an affine function of the state  $Y_t$ , as are all yields and forward rates (assuming no pricing errors). (This is shown explicitly in the Appendix for the case of the Gaussian family  $A_0(3)$ .) Therefore, all of the affine *DTSMs* examined imply that

$$p_t^n = \delta_n + \alpha_n(f_t^n - r_t) + \gamma_{1n}X_{1t} + \gamma_{2n}X_{2t}, \quad (20)$$

where  $(X_{1t}, X_{2t})$  are affine functions of  $Y_t$ . Thus, (20) has us conditioning on the same or richer information sets than those used by Fama [1984a] and Fama and Bliss [1987], among others, in constructing risk premia and excess-return forecasting relations. However, an important difference is that, for all  $n$ , our *DTSMs* imply that the parameters  $(\delta_n, \alpha_n, \gamma_{1n}, \gamma_{2n})$  are functions of the same set of underlying primitive parameters of the model (those describing the evolution of the risk factors and the dependence of  $r$  and  $\Lambda$  on  $Y$ ). We impose these constraints as part of our *ML* estimation.

## 4 *LPY* and Multi-Factor Affine Models

Initially, we focus on the ‘‘canonical’’  $A_m(3)$  models, denoted  $A_{mC}(3)$ , and defined as

$$r(t) = \delta_0 + Y_1(t) + Y_2(t) + Y_3(t), \quad (21)$$

where  $Y(t)$  follows the affine diffusion (12) with volatilities given as in (13) with the normalization  $\alpha_i + \beta_{i1} = 1$ ,  $\Sigma$  is a diagonal matrix of free parameters, and

$$\kappa = \begin{pmatrix} \mathcal{K}_{11} & 0 \\ \mathcal{K}_{21} & \mathcal{K}_{22} \end{pmatrix}, \quad \theta = \begin{pmatrix} \Theta_1 \\ 0 \end{pmatrix}, \quad (22)$$

where  $\mathcal{K}_{11}$  is an  $m \times m$  diagonal matrix,  $\mathcal{K}_{21}$  and  $\mathcal{K}_{22}$  are  $(N-m) \times m$  and  $(N-m) \times (N-m)$  matrices of free parameters, and  $\Theta_1$  is an  $m \times 1$  vector of free parameters.<sup>9</sup> For the case  $m = 0$ ,  $\kappa$  is normalized to be lower triangular and  $\theta = 0$ . For the cases  $m \geq 2$ , we assume that the state variables following square-root diffusions are mutually independent.<sup>10</sup> The

<sup>9</sup>See *DS* and Duffee [2000] for discussions of canonical affine models. In contrast to *DS*, we normalize the  $\delta$  weights on  $Y(t)$  in the definition of  $r(t)$  to unity and treat the factor volatilities (the diagonal elements of  $\Sigma$ ) as free parameters. Additionally, for models with  $m > 0$  we normalized  $\alpha_i + \beta_{i1} = 1$ ,  $i = 1, \dots, m$ , instead of  $\alpha_i = 1$  in the volatility specifications. Adopting the normalization in *DS* instead gave virtually identical values of the likelihood.

<sup>10</sup>This is the standard assumption in models with state variables following square-root diffusions. As noted in *DS*, it is possible to extend these formulations to allow for *positively* correlated state variables through the drifts. However, our *ML* estimation method (see below) does not accommodate this extension.

market prices of risk are given by (15) with the first  $m$  rows of  $\lambda^Y$  set to zeros in the  $A_{mC}(3)$  model to assure admissibility, and the  $\lambda_{i1}^Y$  normalized to zero for  $i = m + 1, \dots, 3$ .

For estimation we used 312 monthly observations on U.S. treasury zero-coupon bond yields for maturities six months and two, three, five, seven, and ten years over the sample period 1970 through 1995. The yields on bonds with six months and two and ten years to maturity were assumed to be measured without error, while the yields of the remaining maturities differed from the model-implied yields by an *i.i.d.* normally distributed error with mean zero. This specification assures that, evaluated at the maximum likelihood estimates, the differences between the observed and model-implied yields have sample means equal to zero at all maturities. Additionally, the assumption that pricing errors are serially independent forces the *DTSM* to capture, as best it can, all aspects of the forecastability of observed yields – a model must match *LPY* without assistance from measurement errors.

*ML* estimation of the  $A_{0C}(3)$  and  $A_{3C}(3)$  models proceeded using the known conditional Gaussian and non-central chi-square density functions of  $r$ , respectively. Full information *ML* estimation of the  $A_{1C}(3)$  and  $A_{2C}(3)$  models proceeded using the methods proposed by Duffie, Pedersen, and Singleton [2000]. They exploit the affine structure of the model to approximate the true, unknown conditional density of  $Y$  and use this approximate density function in constructing the likelihood function of the data. In all cases, standard errors were computed using the sample “outer product” of the scores of the log-likelihood function. For most cases, comparable standard errors were obtained from the sample Hessian matrix. Out of concern that model  $A_{0C}(3)$  is over-parameterized, we re-estimated it after setting to zero the coefficients with the largest relative standard errors,  $(\kappa_{13}, \lambda_1^0, \lambda_2^0, \lambda_{21}^Y, \lambda_{13}^Y, \lambda_{33}^Y)$ . This led to virtually the same value of the likelihood function, so we henceforth study this constrained version of  $A_{0C}(3)$ .<sup>11</sup>

The *ML* estimates of the models  $A_{mC}(3)$ , along with their estimated standard errors and the values of the log-likelihood functions, are displayed in Table 2.<sup>12</sup> Focusing first on  $\kappa$  we see that, consistent with many previous studies (e.g., Chen and Scott [1993] and *DS*), all of the canonical models with  $m > 0$  have one state variable with very slow mean reversion (a  $\kappa_{ii}$  close to zero). This is the “level” factor that is correlated most highly with the first principal component of U.S. treasury yields (see, e.g., Litterman and Scheinkman [1991]). (For instance, in model  $A_{1C}(3)$  for our sample period, the correlation between  $Y_1$  and the first principal component of changes in yields is 0.93.) In contrast, the minimal  $\kappa_{ii}$  in model  $A_{0C}(3)$  is notably larger:  $\kappa_{22} = 0.378$ . Moreover, as we discuss more extensively below, the state vector in model  $A_{0C}(3)$  is a stationary stochastic process under both the  $P$  and  $Q$  measures while, for the models with  $m > 0$ ,  $Y$  is nonstationary under  $Q$ . ( $Q$ -nonstationarity is not a conceptual problem for valuation or estimation.)

Finally, some of the estimated elements of  $\lambda^Y$  are statistically different from zero in models  $A_{0C}(3)$  and  $A_{1C}(3)$ , but all of the non-zero elements in  $\lambda^Y$  for model  $A_{2C}(3)$  are insignificant

<sup>11</sup>For the case of model  $A_{1C}(3)$ , we found little change by constraining  $\lambda_{32}^Y = 0$ , and constraining other parameters to zero led to a notable decline in the value of the likelihood function, so we proceed to study the unconstrained model  $A_{1C}(3)$ .

<sup>12</sup>We have suppressed the estimates of  $\delta_0$ ,  $\theta$  (the long-run means), and the  $\beta_i$  (the coefficients on  $Y$  in the factor volatility specifications), because they do not play central roles in subsequent discussions of matching *LPY*. They are displayed in Appendix C for completeness.

Param.	$A_{0C}(3)$	$A_{1C}(3)$	$A_{2C}(3)$	$A_{3C}(3)$
$\kappa_{11}$	3.012 (.403)	0.002 (.001)	0.628 (.088)	2.714 (.121)
$\kappa_{21}$	0	0.204 (.075)	0	0
$\kappa_{31}$	0	0.295 (.175)	-5.55 (1.18)	0
$\kappa_{12}$	2.081 (.463)	0	0	0
$\kappa_{22}$	0.378 (.122)	0.983 (.355)	0.006 (.002)	0.005 (.002)
$\kappa_{32}$	0	-2.740 (.953)	-0.349 (.320)	0
$\kappa_{23}$	-0.154 (.081)	-0.403 (.204)	0	0
$\kappa_{33}$	0.466 (.110)	2.510 (.499)	2.01 (.285)	0.526 (.082)
$\Sigma_{11}$	0.011 (.002)	0.030 (.000)	0.055 (.005)	0.063 (.002)
$\Sigma_{22}$	0.005 (.002)	0.048 (.011)	0.028 (.000)	0.030 (.000)
$\Sigma_{33}$	0.025 (.001)	0.080 (.013)	0.192 (.023)	0.105 (.013)
$\lambda_1^0$	0	-0.256 (.049)	-1.29 (1.58)	-0.669 (.022)
$\lambda_2^0$	0	-6.521 (.772)	-0.389 (.048)	-0.328 (.090)
$\lambda_3^0$	-0.361 (.269)	-8.490 (1.42)	-0.514 (2.41)	-0.646 (.764)
$\lambda_{11}^Y$	-101.8 (36.91)	0	0	0
$\lambda_{31}^Y$	158.0 (30.90)	0	-0.996 (2.41)	0
$\lambda_{12}^Y$	-100.6 (27.21)	0	0	0
$\lambda_{22}^Y$	-58.09 (26.12)	-19.35 (9.18)	0	0
$\lambda_{32}^Y$	49.81 (22.50)	9.920 (5.17)	-0.643 (.938)	0
$\lambda_{23}^Y$	21.91 (8.92)	18.36 (5.51)	0	0
$\lambda_{33}^Y$	0	-2.071 (5.81)	-0.069 (1.50)	0
ML	33.43	33.54	33.54	33.14

Table 2: Maximum Likelihood Estimates of Canonical  $A_{mC}(3)$  Models. Standard errors of the estimates are given in parentheses. The row “ML” gives the maximized values of the log-likelihood functions.  $\kappa_{13}$ ,  $\lambda_{13}^Y$ , and  $\lambda_{21}^Y$  are zero in all four models.

from zero at conventional significance levels. (Admissibility requires that  $\lambda^Y = 0$  for model  $A_{3C}(3)$ .) Thus, extending the risk premium specification in affine models as in Duffee [2000] is potentially material for matching  $LPY$  with models  $A_{0C}(3)$  and  $A_{1C}(3)$ , but the point estimates suggest extended risk premiums are less important for model  $A_{2C}(3)$ .

#### 4.1 Matching $LPY(i)$

Matching  $LPY(i)$  requires a *positive* correlation between excess returns and the term spread. For a one-factor model with mean-reverting short rate under the risk-neutral measure, this positive correlation is associated with a *negative* correlation between  $r$  and the expected excess return  $\mu^e(t, \tau)$ . A similar negative correlation is also typical in multi-factor models under parameter values consistent with historical behavior of yield curves. The intuition for this is that if risk premiums are *negatively* correlated with the short-term rate, then an expected rise in the short-term rate has two opposing effects: first, holding the risk premium fixed, the prices of long-term bonds will fall; and second, a falling risk premium tends to *increase* the values of long-term bonds. The expectations puzzle in Table 1 arises whenever

the second effect dominates the first, causing the slope of the yield curve to *fall* as interest rates rise – in which case the slope of the yield curve and the risk premium are *positively* correlated.

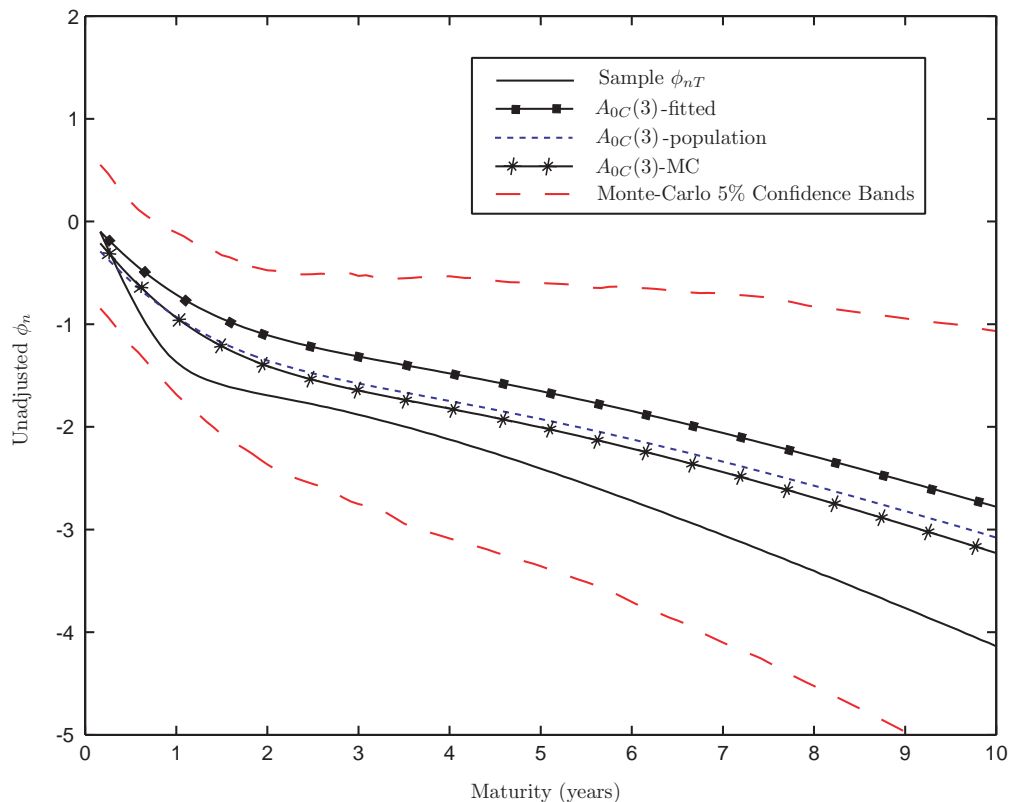


Figure 1: Model Implied Estimates of Unadjusted  $\phi_n$  from the Projection of  $R_{t+1}^{(n-1)} - R_t^n$  onto  $(R_t^n - r_t)/(n - 1)$  for the  $A_{0C}(3)$  Models.

Given the persistent nature of yields, we are naturally led to inquire at the outset of our analysis of  $LPY(i)$  whether small-sample biases might distort our assessments of goodness-of-fit. Toward this end, we focus on model  $A_{0C}(3)$  and its projection coefficients displayed in Figure 1. The graph labeled “ $A_{0C}(3)$ -population” displays the population  $\phi_n$  implied by this model taking the  $ML$  estimates as the true parameters of the data-generating process. To examine the small sample properties of projection coefficients computed from data generated by model  $A_{0C}(3)$ , we conducted the following, limited Monte-Carlo exercise: five hundred samples of length 312 (the length of our sample of treasury yields) were simulated from model  $A_{0C}(3)$  and, for each sample, the  $\phi_{n,312}$  were estimated. The mean of these estimates across the five hundred samples (displayed in Figure 1 as “ $A_{0C}(3)$ -MC”) lies very close to  $A_{0C}(3)$ -population, suggesting that small-sample biases are negligible for this model when estimated by the method of  $ML$ .

Proceeding with our assessment of fit, we see that the historical estimates  $\phi_{nT}$  from Table 1 (displayed as graph “Sample  $\phi_{nT}$ ” in Figure 1) lie well inside the Monte Carlo confidence bands – the 5% quantiles of the small sample distribution of the  $\phi_{nT}$ . Thus, based on the model-implied population results, we conclude that model  $A_{0C}(3)$  is successful

in matching  $LPY(i)$ . From Figure 2 we see that the remaining three  $A_{mC}(3)$  models do not fair nearly as well at matching  $LPY(i)$  in the population. Indeed, none of these graphs have the characteristic downward sloping pattern of *Sample*  $\phi_{nT}$  and graphs for models  $A_{2C}(3)$  and  $A_{3C}(3)$  are approximately horizontal lines at unity! From these population values we conclude that only model  $A_{0C}(3)$  is successful at matching  $LPY(i)$ .

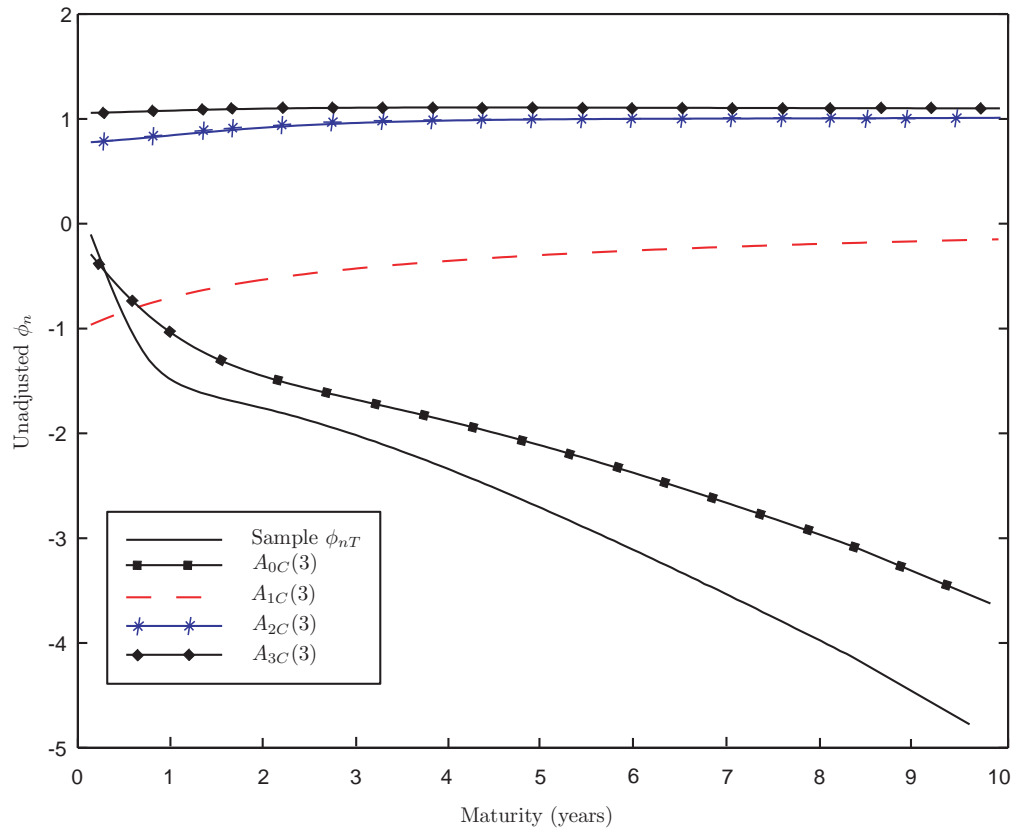


Figure 2: Model-Implied Population Estimates of  $\phi_n$  from the Projection of  $R_{t+1}^{(n-1)} - R_t^n$  onto  $(R_t^n - r_t)/(n - 1)$  for the  $A_{mC}(3)$  Models.

Figure 1 also displays the model-implied  $\phi_n$  computed from the sample of model-implied fitted yields (graph “ $A_{0C}(3)$ -fitted”). These are obtained by inverting the model for the fitted state variables, computing model-implied fitted zero-coupon bond yields, and then estimating the regressions in Table 1 using these fitted yields. It turns out that the corresponding “fitted” graphs for *all three* of the other canonical models are virtually on top of “ $A_{0C}(3)$ -fitted.” Yet we have just seen that the population counterparts,  $\phi_n$ , for models  $A_{mC}(3)$ ,  $m > 0$ , look very different than “ $A_{0C}(3)$ -population.” The reason there is this discrepancy between the  $\phi_n$  computed in the population and with fitted yields is that the latter mix properties of the model with those of the historical data (which may not coincide). We conclude that assessments of fit based on fitted yields can give very misleading impressions of the actual population distributions implied by *DTSMs* and this motivates our focus on the population  $\phi_n$ .

## 4.2 Matching $LPY(ii)$

Whether or not a  $DTSM$  adequately captures the persistence of expected excess returns, or equivalently of the market risk premiums, is measured in part by its effectiveness at matching  $LPY(ii)$ . Given well-specified risk premiums, the term structure of risk-adjusted  $\phi_n^{\mathcal{R}}$  should be a horizontal line at unity. In this case, the most interesting  $\phi_n^{\mathcal{R}}$  to examine are those computed from (18) using actual historical yields  $R_t^n$  and risk premiums  $c_t$  and  $p_t$  evaluated at the fitted state variables. Were we instead to compute population model-implied  $\phi_n^{\mathcal{R}}$ , they would be identically equal to unity for any  $DTSM$  regardless of its descriptive qualities.

Figure 3 displays the model-implied  $\phi_n^{\mathcal{R}}$  along with the historical results from Table 1 ( $Sample-\phi_{nT}$ ). We see that models  $A_{0C}(3)$  and  $A_{1C}(3)$  imply risk premiums that fully meet the challenge  $LPY(ii)$ , at least beyond maturities of about two years. In contrast, as with  $LPY(i)$ , models  $A_{2C}(3)$  and  $A_{3C}(3)$  fail entirely to match  $LPY(ii)$ ; adjusting for risk premiums in these models gives virtually the same results as in the unadjusted regressions. Since the  $Sample-\phi_{nT}$  results challenge expectations theories most dramatically at the longer end of the yield curve, we focus on the results for two years and beyond, deferring until Section 4.6 further discussion of the fit to under two years to maturity.

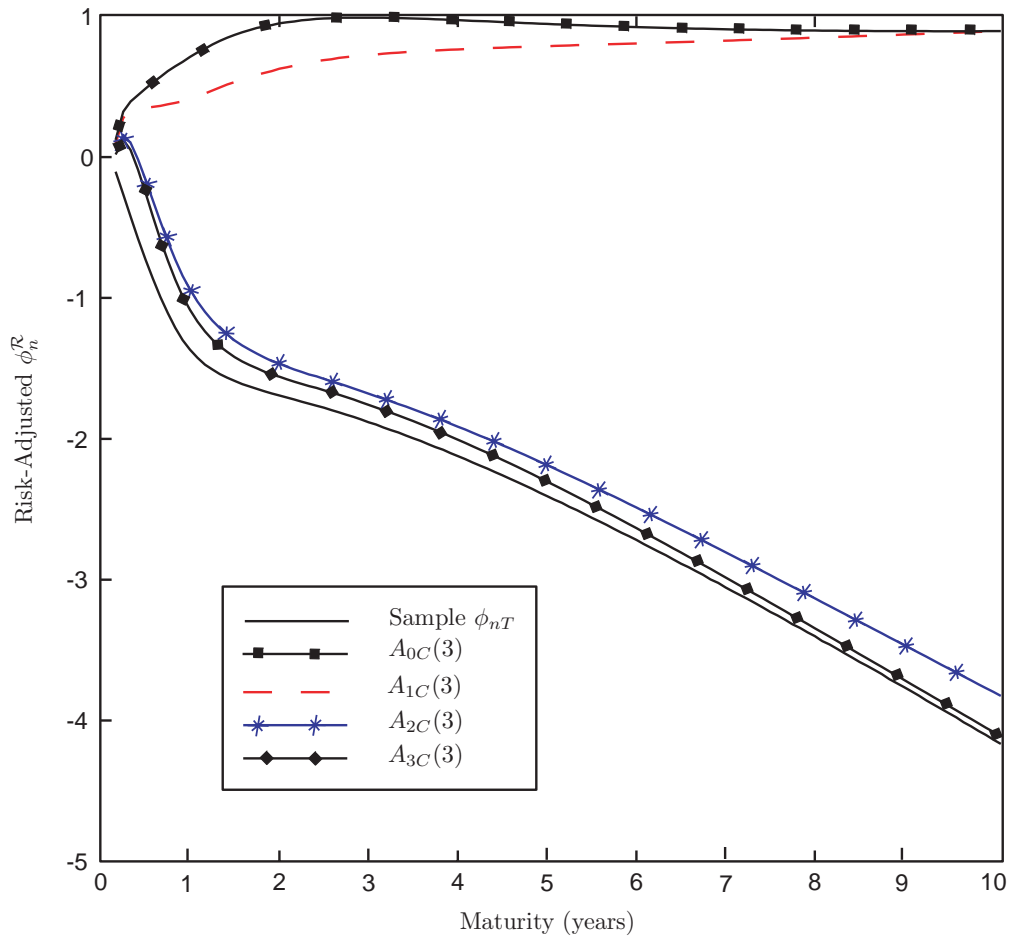


Figure 3: Model-Implied Estimates of  $\phi_n^{\mathcal{R}}$  From the Models  $A_{mC}(3)$ .

### 4.3 Further Observations On Matching *LPY*

Why does model  $A_{0C}(3)$  outperform the other canonical models in its ability to match both dimensions of *LPY*? There are at least two notable differences between model  $A_{0C}(3)$  and the other canonical models: (1) models  $A_{mC}(3)$ ,  $m \geq 1$ , accommodate stochastic volatility and model  $A_{0C}(3)$  does not, and (2) in model  $A_{0C}(3)$ ,  $Y(t)$  is a stationary process under both the  $P$  and  $Q$  measures,<sup>13</sup> while  $Y$  is  $P$ -stationary, but  $Q$ -nonstationary in models  $A_{mC}(3)$ ,  $m \geq 1$ .

Param.	$A_{0C}(3)$	$A_{1C}(3)$	$A_{1S}(3)$	$A_{1L}(3)$
$\kappa_{11}$	3.01 (1.91)	0.002 (-0.005)	0.653 (0.574)	1.95 (1.94)
$\kappa_{21}$	0	0.204 (-0.107)	-5.45 (-6.33)	-0.44 (0.08)
$\kappa_{31}$	0	0.295 (-0.384)	0.029 (0.039)	1.01 (3.99)
$\kappa_{12}$	2.08 (0.99)	0	0	0
$\kappa_{22}$	0.378 (0.062)	0.983 (0.062)	1.50 (1.80)	0.13 (0.002)
$\kappa_{32}$	0	-2.740 (-1.95)	-0.022 (-0.011)	0.25 (-0.03)
$\kappa_{23}$	-0.154 (-0.035)	-0.403 (0.471)	-16.6 (-38.2)	-0.26 (0.005)
$\kappa_{33}$	0.466 (0.466)	2.510 (2.340)	0.500 (0.244)	0.61 (0.58)
ML	33.43	33.54	33.54	33.42

Table 3: Maximum Likelihood Estimates of  $\kappa$  from Constrained  $A_m(3)$  Models. Risk-neutral  $\kappa$ 's are given in parentheses.

The stationarity properties of  $Y$  are documented in columns two and three of Table 3 for models  $A_{0C}(3)$  and  $A_{1C}(3)$ , where the actual and risk-neutral (in parentheses) values of  $\kappa$  are presented. To gain some insight into the role of  $Q$ -nonstationary  $Y$  for matching *LPY*, we re-estimated model  $A_{1C}(3)$  under the constraint that  $Y$  be  $Q$ -stationary, model  $A_{1S}(3)$ , and the resulting  $\kappa$  matrix is displayed in column four of Table 3. Imposing  $Q$ -stationarity leads to a “rotation” of the risk factors: in model  $A_{1C}(3)$  the first, volatility factor is the “level” of the yield curve (the factor with slowest mean reversion), whereas it becomes “slope” (the factor with intermediate mean reversion) in model  $A_{1S}(3)$ . The consequences of this change for matching *LPY* are displayed in Figure 4. The population  $\phi_n$  now exhibit the downward sloping pattern of their sample counterparts. However, they do not come close to achieving the large negative numbers for long maturities exhibited in Table 1. Moreover, now the coefficients  $\phi_n^R$  are notably less than unity for all maturities, suggesting that the model-implied risk premiums no longer have the requisite properties to match *LPY*(ii).

Still unresolved is the role of stochastic volatility in matching *LPY* within the  $A_1(3)$  family. The third model,  $A_{1L}(3)$ , displayed in Table 3 and Figure 4 is obtained by further constraining model  $A_{1S}(3)$  to have  $\beta_{21} = \beta_{31} = 0$ . In this case, only  $Y_1$  exhibits instantaneous stochastic volatility. These constraints lead to another factor rotation with the volatility

<sup>13</sup>The relatively large values of the  $\kappa_{ii}$  in  $A_{0C}(3)$  is clearly a consequence of our having extended the usual specification of risk premiums in Gaussian models to allow  $\lambda^Y \neq 0$ . When we re-estimate model  $A_{0C}(3)$  with  $\lambda^Y = 0$ , the smallest diagonal element of  $\kappa$  is  $\kappa_{22} = .002$ . So by letting  $\lambda^Y \neq 0$  we fundamentally change the persistence of  $Y$  under the  $P$  probability measure which, as we have seen, has material implications for matching *LPY*. Duarte [1999] reports a similar change in persistence due to a different reparameterization of the risk premium in a three-factor square-root diffusion model.

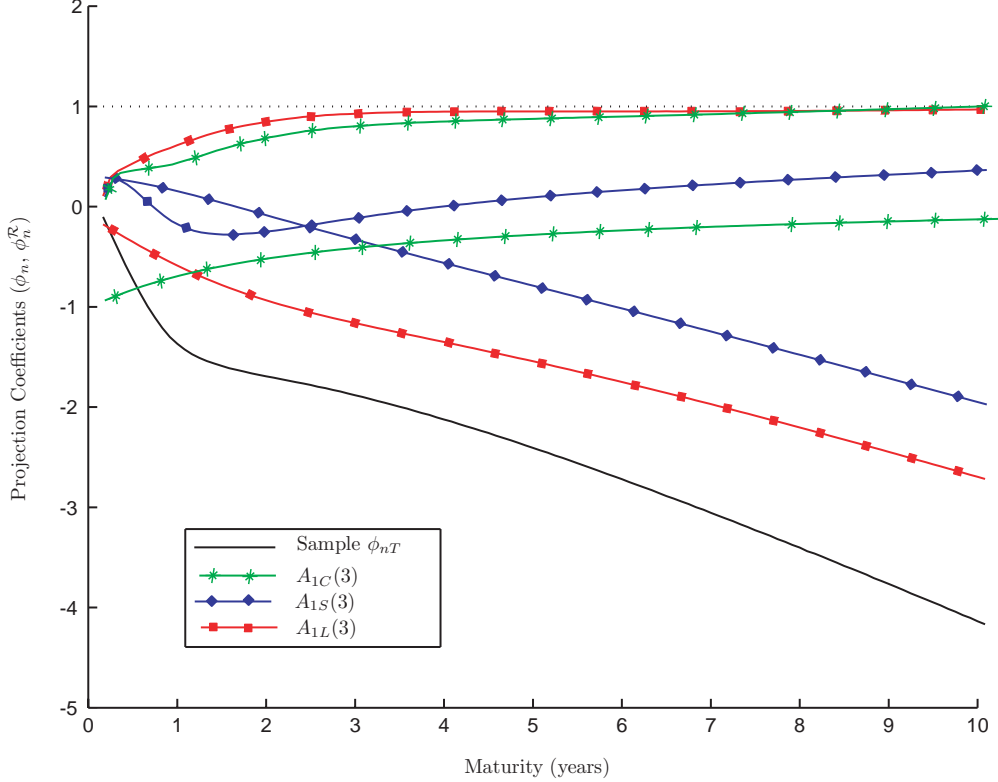


Figure 4: Model implied projection coefficients from constrained versions of model  $A_{1C}(3)$ . For each model the upper (lower) graph is a plot of  $\phi_n^{\mathcal{R}}$  ( $\phi_n$ ).

factor  $Y_1$  becoming the factor with the fastest rate of mean reversion. From Figure 4 we see that this model matches quite well *both*  $LPY(ii)$  for longer maturities and the downward sloping pattern of the unadjusted  $\phi_n$ ,  $LPY(i)$ .

Taken together, these results suggest two broad conclusions. First, at least for the families  $A_0(3)$  and  $A_1(3)$ , we have found strong support for our heuristic that time-varying risk premiums of a form consistent with the Fama-Bliss evidence resolve the expectations puzzles for longer maturity bonds. Second, the reason the model  $A_{1C}(3)$ , with or without  $Q$ -stationary  $Y$ , cannot fully match  $LPY$  is not the presence of stochastic volatility *per se*. Rather it appears to be the *desire* of the likelihood function to trade off matching  $LPY$ , and related features of the conditional means of yields, with fitting the stochastic volatility in the yield data. Given the flexibility to introduce stochastic volatilities for  $Y_2$  and  $Y_3$  (set  $(\beta_{21}, \beta_{31}) \neq 0$ ), the likelihood optimization will do just that in order to achieve a higher value of the likelihood (compare the likelihood values for models  $A_{1C}(3)$  and  $A_{1L}(3)$ ). The higher value of the likelihood and better fit to the conditional second moments of returns is achieved at the expense of matching  $LPY$ , however. In other words, affine models do not appear to have the flexibility simultaneously to fit  $LPY$  and the volatility properties of returns.

The worst performing model is the three-factor  $CIR$ -style model  $A_{3C}(3)$  in which  $\Lambda_i(t) = \ell_{0i}\sqrt{Y_i(t)}$ , so that  $Y$  affects the market prices of risk only through factor volatilities. Models in the families  $A_N(N)$  are not easily “fixed up” to match  $LPY$  with market prices of risk of



the form (15) without introducing arbitrage opportunities (Cox, Ingersoll, and Ross [1985]). Model  $A_{2C}(3)$  also has more flexibility in principle than model  $A_{3C}(3)$ . However, the non-zero admissible elements of  $\lambda^Y$  are all insignificantly different from zero and this shows up in Figure 3 as risk-adjusted projection coefficients that are nearly the same as in model  $A_{3C}(3)$ .

#### 4.4 Does Sample Period Matter?

Our sample period, 1970 - 1995, includes the period of historically high interest rate volatility associated with the monetary experiment of the Federal Reserve in the late 1970's. Many have argued that this period is best viewed as a different "regime" (e.g., Gray [1996]). In the light of the preceding remarks regarding trade-offs in fitting conditional means and volatilities, might the inclusion of this volatile period have affected our findings? We address this issue by re-estimating model  $A_{1C}(3)$  for the post-experiment period January, 1983 - December, 1995 (156 monthly observations). Model  $A_{1C}(3)$  was chosen, because it was the best performing model with stochastic volatility (in terms of matching  $LPY$ ) for the full sample period. Figure 5 shows that the corresponding  $\phi_n$  and  $\phi_n^{\mathcal{R}}$  for the full and post-1983 sample periods are nearly identical.

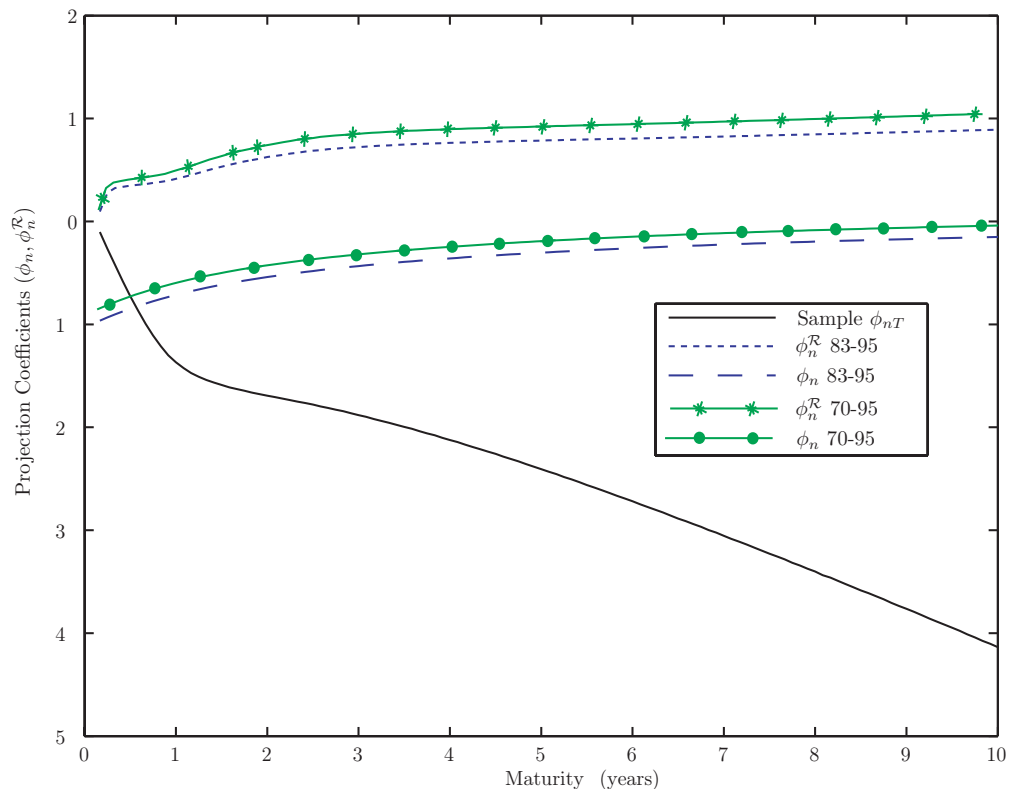


Figure 5: Model-implied projection coefficients from model  $A_{1C}(3)$  for the sample periods January, 1970 through December, 1995 and February, 1983 through December, 1995.

## 4.5 Risk Premia Versus Factor Correlations in Matching *LPY*

Though we have focused on the role of risk premiums in matching *LPY*, as stressed by *DS*, there are also important differences between affine models in terms of the nature of the factor correlations accommodated. In particular, model  $A_{0C}(3)$  offers the most flexibility in specifying factor correlations, so it is of interest to explore the relative contributions of non-diagonal  $\kappa$  and nonzero  $\lambda^Y$  in matching *LPY* in Gaussian models.

Figure 6 displays the model-implied  $\phi_n^{\mathcal{R}}$  from four different  $A_0(3)$  models. Models  $A_{0VU}(3)$  and  $A_{0V}(3)$  are standard ‘‘Vasicek’’ models in which  $\lambda^Y = 0$  (market prices of risk are constants); see Langetieg [1980]. The former imposes zero factor correlations (diagonal  $\kappa$ ), while the latter allows maximal flexibility in the correlation structure. Clearly neither model matches *LPY*(ii).

Maintaining the assumption of zero factor correlation through the drift, model  $A_{0U}(3)$  relaxes the assumption that  $\lambda^Y = 0$  (this is model  $A_{0C}(3)$  with diagonal  $\kappa$ ). We see a notable improvement in matching *LPY* in this model, but state-dependent risk premiums *per se* are clearly not sufficient. Rather, comparing the results for models  $A_{0C}(3)$  and  $A_{0U}(3)$ , we see that it is the combination of non-zero factor correlations through both the drift,  $\kappa$ , and state-dependent market prices of risk,  $\lambda^Y$ , that allow model  $A_{0C}(3)$  to match *LPY*.

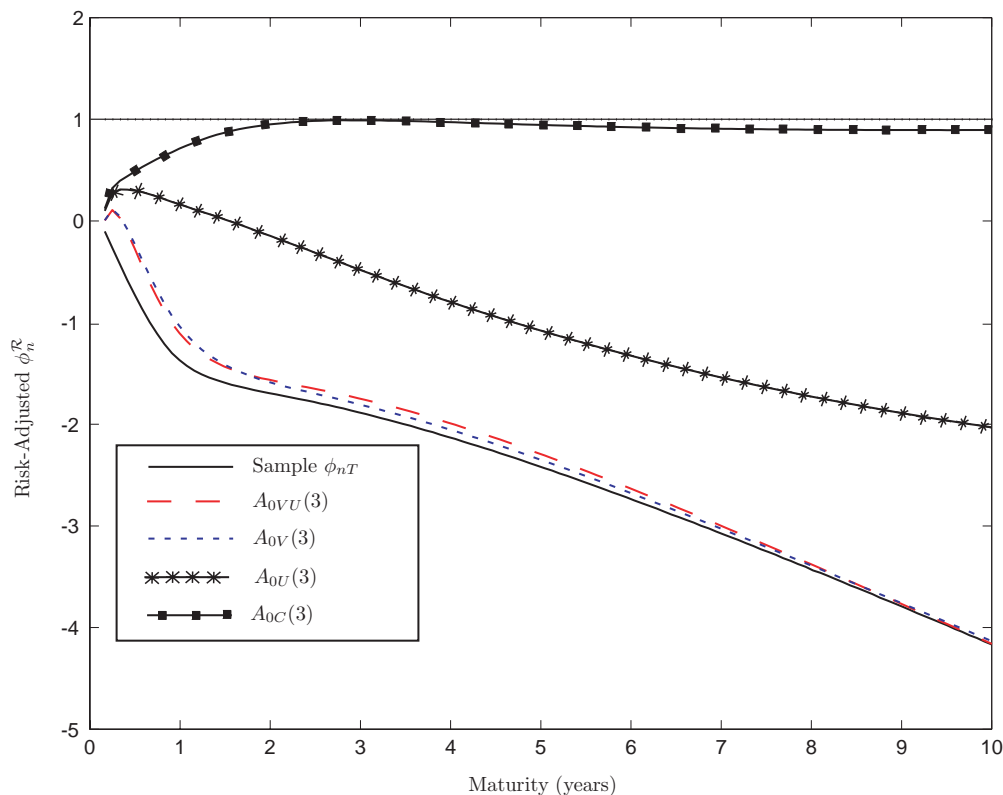


Figure 6: Estimated  $\phi_n^{\mathcal{R}}$  implied by *ML* Estimates of Constrained  $A_0(3)$  Models.

## 4.6 The Short End of the Yield Curve

We have deferred examination of the failure of affine models to match  $LPY(ii)$  at the short end of the yield curve, under two years to maturity. We now show that the mismatch at the short end is rectified by the addition of a fourth “short-end” factor. That is, failing to match  $LPY$  at all maturities was simply the consequence of an omitted factor with relevance primarily for very short-dated bonds. To show this, we proceed to estimate the canonical Gaussian four-factor model,  $A_{0C}(4)$ , assuming the one- and six-month, and two- and ten-year yields are measured without errors. The match to  $LPY(ii)$  at the short end is now nearly perfect as can be seen from Figure 7.<sup>14</sup>

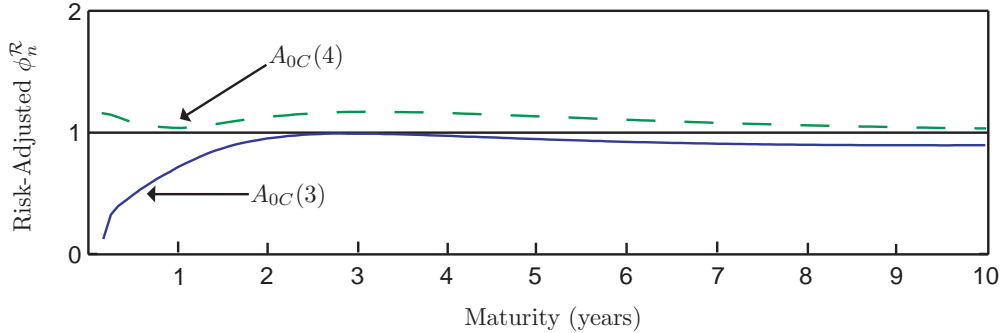


Figure 7: Estimated  $\phi_n^R$  implied by  $ML$  Estimates of Models  $A_{0C}(3)$  and  $A_{0C}(4)$ .

That the omission of the fourth factor from model  $A_{0C}(3)$  can potentially lead to its failure to match  $LPY(ii)$  at maturities under two years can be seen intuitively as follows. Equations (2), (6), and (7) imply that

$$R_{t+1}^{n-1} - R_t^n - (c_{t+1}^{n-1} - c_t^{n-1}) + \frac{1}{n-1}p_t^n = \frac{1}{n-1}(R_t^n - r_t) - \sum_{i=1}^{n-1}(E_{t+1}r_{t+i} - E_t r_{t+i}) \quad (23)$$

The estimated  $\phi_n^R$  displayed in Figures 3 and 7 are obtained by projecting the left-hand-side of (23) onto  $(R_t^n - r_t)/(n-1)$ , where the  $R_t^n$  are historical yields and  $c_t^n$  and  $p_t^n$  are model-implied risk premiums evaluated at the fitted state variables. If a model is correctly specified, then (in large samples) the projection of the second term on the right-hand-side of (23) onto  $(R_t^n - r_t)/(n-1)$  gives zero, so we get  $\phi_n^R = 1$ . However, if the time-series properties of the fitted  $r_t$  do not match those of the historical one-month rate, then the model-implied forecast error  $\sum_i(E_t r_{t+i} - E_{t+1} r_{t+i})$  and historical slope  $(R_t^n - r_t)/(n-1)$  will be correlated. In fact, the model-implied forecast error  $\sum_i(E_{t+1} r_{t+i} - E_t r_{t+i})$  is predictable using historical data over a maturity range that induces a downward bias in  $\phi_n^R$  out to about two years. The fourth factor in model  $A_{0C}(4)$  corrects for this mispricing at the very short end and, as such, it fully resolves the expectations puzzles.

<sup>14</sup>In a different context, Longstaff, Santa-Clara, and Schwartz [2000] find that a four-factor model, and in particular a model with a factor dedicated to the very short end of the curve, is necessary to model the term structure of LIBOR and swap rates.

## 5 Conclusion

We began this exploration of expectations puzzles with the conjecture that richer risk premiums than those accommodated by traditional “Vasicek” or “CIR” models will give the requisite flexibility for *DTSMs* to match *LPY* – thereby resolving the puzzles that have arisen in testing the expectations hypothesis. We showed that this is indeed the case for the Gaussian sub-family of affine *DTSMs* with correlated factors and state-dependent risk premiums. They fully matched *LPY*, and in particular resolved the Campbell-Shiller expectations puzzles, *at the maximum likelihood estimates for this model*.

On the other hand, models with state-dependent factor volatilities (those in the families  $A_m(3)$  with  $m > 0$ ) fared less well in matching *LPY*. The reason seemed to be a difficulty in simultaneously matching the conditional first-moment properties of yields, as summarized by *LPY*, and the conditional volatilities of yields. Gaussian models resolve this tension (trivially) by positing constant volatilities and thereby ignoring time-varying volatility in the likelihood function. Once stochastic volatility is admitted as a possibility in the families  $A_m(3)$  with  $m > 0$ , then the likelihood function seems to give substantial weight to fitting volatility at the expense of matching *LPY*.

More generally, we conjecture that there is a much larger class of *DTSMs* with sufficient flexibility to match *LPY*. For instance, the family of  $N$ -factor quadratic-Gaussian models studied by Leippold and Wu [1998], Lu [1999], and Ahn, Dittmar, and Gallant [2000] has the instantaneous short rate  $r$  given by  $r(t) = a_0 + Y'b_0 + Y'c_0Y$ , where  $c_0$  is an  $N \times N$  symmetric matrix of constants and  $Y$  follows a Gaussian diffusion. The market price of risk in the canonical  $N$ -factor quadratic-Gaussian model takes exactly the same form as (15) for the Gaussian case of  $m = 0$  (Ahn, Dittmar, and Gallant [2000]). Thus, one might expect that this model would have comparable flexibility to model  $A_{0C}(N)$  in matching *LPY*. Evidence consistent with this conjecture is reported in Dai and Singleton [2001] for the case of  $N = 1$ .

Alternatively, Naik and Lee [1997] introduce Markov regime switching into an affine (*CIR*-style) model (see Evans [2000] for the analogous result for discrete-time *CIR*-style models). Bansal and Zhou [2000] study Markov switching in the context of richer discrete-time affine models. In all of these cases, the presence of Markov switching introduces additional free parameters into the pricing relation, thereby giving the model more flexibility to match *LPY*.<sup>15</sup> Another family of *DTSMs* that might match *LPY* are the models proposed by Duarte [1999] in which the state vector follows the affine diffusion (12) and  $\Lambda(t) = \sqrt{S(t)}\ell_0 + \Sigma^{-1}c$ , for some constant  $N$ -vector  $c$ . The only state-dependence of  $\Lambda(t)$  in Duarte’s model is through the factor volatilities.

Given these findings, a natural next step is providing economic underpinnings for our parameterization of  $\Lambda(t)$  in (15). While this is beyond the scope of this paper, we briefly discuss two possible structural underpinnings of this affine parameterization within a one-factor Gaussian model. (Obviously, moving to more factors only expands the possible structural interpretations.) First, it turns out that McCallum [1994]’s resolution of the expectations puzzle based on the behavior of a monetary authority is substantively equivalent to our

---

<sup>15</sup>Bekaert, Hodrick, and Marshall [1997b] also explore “peso problem” interpretations of the failure of the expectations hypothesis by positing a regime switching model for the short rate and exploring the implications for the Campbell-Shiller regressions.

affine parameterization of the market price of risk. McCallum [1994] starts by exogenously specifying the yield premium as an AR(1) process, and the riskless rate process as an AR(1) process augmented by an linear policy reaction rule:  $r_t = \sigma r_{t-1} + \lambda(R_t - r_t) + \zeta_t$ , where the first term is a mean-reverting or “smoothing” component, the second term is a “policy reaction” component with  $0 \leq \lambda \leq 2$  to rule out bubble solutions, and  $\zeta_t$  is a policy shock. Under the assumptions that (i)  $\sigma = 1$  (which is the case studied by Kugler [1997]), and (ii) the bond yield is linear in the short rate (i.e.,  $R_t = b_0 + b_1 r_t$ ), the monetary policy rule implies that  $r_t$  is an AR(1) process with mean reversion coefficient  $\kappa = (1 - b_1)\lambda/[1 + (1 - b_1)\lambda]$ . Supposing that  $r$  is also an AR(1) process under the risk-neutral measure (with mean reversion coefficient  $\tilde{\kappa}$ ), then  $b_1 \approx 1 - \tilde{\kappa}/2$  and  $\lambda \approx 2\kappa/\tilde{\kappa}$ . Thus, the condition  $0 \leq \lambda \leq 2$  translates into the condition  $\tilde{\kappa} \geq \kappa > 0$ . In other words, the constraints on  $\lambda$  that produce McCallum’s “policy reaction” interpretation of interest rate behavior are equivalent to our state-dependent formulation of the market price of risk.

An alternative motivation comes from the general equilibrium production economy with stochastic habit formation studied in Dai [2000]. He shows that, in a neoclassical setting of consumption, saving, and wealth accumulation with risky production, if an infinitely lived representative agent has a time-nonseparable preference induced by stochastic habit formation, then the implied negative correlation between  $r_t$  and the Sharpe ratio of risky production technology allows the model to generate the pattern in Table 1. The models with affine, state-dependent market price of risk that we studied can be interpreted as approximations to the (intrinsically nonlinear) interest rate dynamics implied by Dai’s model.

## Appendices

### A Multi-factor Gaussian Model – Some Basic Facts

This appendix outlines the basic features of Gaussian *DTSMs* that we use in our analysis. Assume that the *instantaneous* short rate  $r_0(t)$  is a linear function of the  $N \times 1$  state vector  $Y(t)$ :

$$r_0(t) = a_0 + b_0'Y(t), \quad (24)$$

where  $a_0$  is a constant, and  $b_0$  is a  $N \times 1$  vector.

The state dynamics under the physical measure is given by

$$dY(t) = \kappa(\theta - Y(t))dt + \sigma dW(t), \quad (25)$$

where  $\kappa$  and  $\sigma$  are  $N \times N$  matrices and  $\theta$  is a  $N \times 1$  vector.

The market price of risk<sup>16</sup> is given by

$$\Lambda(t) = \sigma^{-1}(\lambda^0 + \lambda^Y Y(t)), \quad (26)$$

where  $\lambda^0$  is a  $N \times 1$  vector and  $\lambda^Y$  is a  $N \times N$  matrix of constants. If the Girsanov's theorem applies,<sup>17</sup> the risk neutral dynamics of the state vector is given by

$$dY(t) = \tilde{\kappa}(\tilde{\theta} - Y(t))dt + \sigma d\tilde{W}(t), \quad (27)$$

where  $\tilde{\kappa} = \kappa + \lambda^Y$  and  $\tilde{\theta} = \tilde{\kappa}^{-1}(\kappa\theta - \lambda^0)$ .

We assume that  $\kappa$  can be decomposed as  $\kappa = X^{-1}\kappa_d X$ , where  $\kappa_d$  is a diagonal matrix with strictly positive diagonal elements  $\kappa_i$ ,  $1 \leq i \leq N$ ,  $X$  is a non-singular real matrix, with diagonal elements normalized to 1.<sup>18</sup> Similarly, we assume that  $\tilde{\kappa}$  can also be decomposed as  $\tilde{\kappa} = \tilde{X}^{-1}\tilde{\kappa}_d\tilde{X}$ , where  $\tilde{\kappa}_d$  is diagonal with diagonal elements  $\tilde{\kappa}_i$ ,  $1 \leq i \leq N$ , and  $\tilde{X}$  is a non-singular normalized matrix.

The relevant properties of the Gaussian model we need for later development are the following. First, the conditional mean of the state vector is given by

$$E[Y(t + \tau)|Y(t)] = e^{-\kappa\tau}Y(t) + (I - e^{-\kappa\tau})\theta. \quad (28)$$

The conditional variance is given by

$$\text{Var}(Y(t + \tau)|Y(t)) = X^{-1}\Omega(\tau)X'^{-1}, \quad (29)$$

---

<sup>16</sup>The pricing kernel is given by

$$\frac{dM(t)}{M(t)} = -r_0(t)dt - \Lambda(t)dW(t).$$

<sup>17</sup>See Appendix B for a proof that this is indeed the case.

<sup>18</sup>Alternatively, one could normalize the Euclidean length of each column vector of  $X$  to 1. If  $\sigma$  is completely free, then we can choose to normalize  $\kappa$  to be diagonal. In which case, we set  $X \equiv I$ .

where

$$\Omega_{ij}(\tau) = \Sigma_{ij} \frac{1 - e^{-(\kappa_i + \kappa_j)\tau}}{\kappa_i + \kappa_j}, \quad (30)$$

$$\Sigma = X\sigma\sigma'X'. \quad (31)$$

The zero coupon bond price and yield (with term to maturity  $\tau$ ) are given by

$$P(t, \tau) = e^{-A(\tau) - B(\tau)'Y(t)}, \quad (32)$$

$$R(t, \tau) = a(\tau) + b(\tau)'Y(t), \quad (33)$$

where  $a(\tau) = A(\tau)/\tau$ ,  $b(\tau) = B(\tau)/\tau$ ,

$$b(\tau) = (I - e^{-\tilde{\kappa}'\tau})(\tilde{\kappa}'\tau)^{-1}b_0, \quad (34)$$

$$a(\tau) = a_0 + (b_0 - b(\tau))'\tilde{\theta} - \frac{1}{2}\text{Tr} \left[ \Xi(\tau)\tilde{X}'^{-1}\tilde{\kappa}'^{-1}b_0b_0'\tilde{\kappa}^{-1}\tilde{X}^{-1} \right], \quad (35)$$

$$\Xi_{ij}(\tau) = \tilde{\Sigma}_{ij} \left[ 1 - \frac{1 - e^{-\tilde{\kappa}_i\tau}}{\tilde{\kappa}_i\tau} - \frac{1 - e^{-\tilde{\kappa}_j\tau}}{\tilde{\kappa}_j\tau} + \frac{1 - e^{-(\tilde{\kappa}_i + \tilde{\kappa}_j)\tau}}{(\tilde{\kappa}_i + \tilde{\kappa}_j)\tau} \right], \quad (36)$$

$$\tilde{\Sigma} = \tilde{X}\sigma\sigma'\tilde{X}'. \quad (37)$$

## A.1 Risk Premiums

Let us fix  $\Delta$  as the length of a period, and define  $a_n \equiv a(n\Delta)$ ,  $b_n \equiv b(n\Delta)$ ,  $A_n \equiv A(n\Delta)$ , and  $B_n \equiv B(n\Delta)$ . We will also frequently use the short hand  $t + n$  to represent  $t + n\Delta$ , whenever there is no confusion. Then the  $n$ -period zero yield is given by  $R_t^n = a_n + b_n Y_t$  and we let  $r_t \equiv R_t^1$ . The conditional mean of the short rate is given by

$$E_t[r_{t+n}] = \mu_n + \nu_n'Y(t), \text{ where} \quad (38)$$

$$\mu_n = a_1 + \theta'(I - e^{-\kappa'n})b_1 \quad (39)$$

$$\nu_n = e^{-\kappa'n}b_1 \quad (40)$$

The one-period forward rate, delivered  $n$ -period hence,  $f_t^n$ , is given by

$$f_t^n \equiv -\frac{1}{\Delta} \ln \frac{P(t, (n+1)\Delta)}{P(t, n\Delta)} = A_n^\Delta + B_n^{\Delta'}Y(t), \quad (41)$$

where

$$A_n^\Delta \equiv \frac{A_{n+1} - A_n}{\Delta} \text{ and } B_n^\Delta \equiv \frac{B_{n+1} - B_n}{\Delta}. \quad (42)$$

Thus, the forward risk premium is given by

$$p_t^n \equiv f_t^n - E_t[r_{t+n}] = (A_n^\Delta - \mu_n) + (B_n^\Delta - \nu_n)'Y(t), \quad (43)$$

which is linear in the state vector. It follows that the yield risk premium,  $c_t^n$ , defined by  $c_t^n \equiv \frac{1}{n} \sum_{i=0}^{n-1} p_t^i$ , is also linear in the state vector.

If we have  $N$  observed yields (or related yield curve variables, such as term spreads), we can substitute out  $Y(t)$  by these yields. This is the general procedure for obtaining an  $N$ -factor risk premium model in which the forward term premium is predicted by  $N$  observed yields.

## B Conditions for Girsanov's Theorem

The goal is to show that

$$Z(t) = e^{\int_0^t \Lambda_s' dW_s - \frac{1}{2} \int_0^t \Lambda_s' \Lambda_s ds}, \quad (44)$$

is a Martingale, when  $\Lambda_s$  is an affine function of a Gaussian state-vector. It can be shown that the standard Novikov condition imposes a strong restriction on model parameters. We use a weaker condition to show that  $Z(t)$  is a Martingale without imposing parametric restrictions.

According to Corollary 5.16 of Karatzas and Shreve [1988], if,  $\Lambda_t$  is a progressively measurable function of the Brownian motion, and for arbitrary  $T > 0$ , there exists a  $K_T > 0$ , such that

$$|\Lambda_t| \leq K_T(1 + W^*(t)), \quad 0 \leq t \leq T, \quad (45)$$

where  $W^*(t) = \max_{0 \leq s \leq t} |W(s)|$ , then  $Z(t)$  is a martingale.

For simplicity, consider the one-dimensional case (extension to the multi-dimensional case is straightforward.) Without loss of generality, we can assume that the long-run mean of  $Y(t)$  is zero, and its volatility is 1. Then it can be shown that

$$Y_t = \int_0^t e^{-\kappa(t-u)} dW_u = W_t + \int_0^t W_u de^{-\kappa(t-u)}.$$

It follows that

$$\begin{aligned} |Y_t| &\leq |W_t| + \int_0^t |W_u| de^{-\kappa(t-u)} \\ &\leq W_t^*(1 + \int_0^t de^{-\kappa(t-u)}) = W_t^*(2 - e^{-\kappa t}) \\ &\leq (2 - e^{-\kappa T})W_t^* \leq (2 - e^{-\kappa T})(1 + W_t^*) \end{aligned}$$

Since  $\Lambda_t$  is an affine function of  $Y(t)$ , it is obvious that (45) holds.

## C Estimates of Auxiliary Parameters in $A_{mC}(3)$ Models

Following are the estimations of the auxiliary parameters for the  $A_{mC}(3)$  models.



Param.	$A_{0C}(3)$	$A_{1C}(3)$	$A_{2C}(3)$	$A_{3C}(3)$
$\delta_0$	0.055	-0.142	-0.035	-0.499
$\theta_1$	NA	0.197	0.017	0.372
$\theta_2$	NA	NA	0.073	0.101
$\theta_3$	NA	NA	NA	0.081
$\beta_{21}$	NA	1.00	NA	NA
$\beta_{31}$	NA	1.00	1.00	NA
$\beta_{32}$	NA	NA	$1.22 \times 10^{-12}$	NA

## References

- Ahn, D.-H., R. F. Dittmar, and A. R. Gallant (2000). Quadratic gaussian models: Theory and evidence. Working paper, University of North Carolina.
- Backus, D., S. Foresi, A. Mozumdar, and L. Wu (1997, April). Predictable changes in yields and forward rates. Working Paper, New York University.
- Bansal, R. and H. Zhou (2000, June). Term structure of interest rates with regime shifts. Working Paper, Duke University.
- Bekaert, G., R. Hodrick, and D. Marshall (1997a). On biases in tests of the expectations hypothesis of the term structure of interest rates. *Journal of Financial Economics* 44, 309–348.
- Bekaert, G., R. Hodrick, and D. Marshall (1997b). ‘Peso problem’ explanations for term structure anomalies. Working paper, Columbia University.
- Campbell, J. Y. and R. J. Shiller (1991). Yield spreads and interest rate movements: A bird’s eye view. *Review of Economic Studies* 58, 495–514.
- Chen, R. and L. Scott (1993). Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates. *Journal of Fixed Income* 3, 14–31.
- Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica* 53, 385–408.
- Dai, Q. (2000). From equity premium puzzle to expectations puzzle: A general equilibrium production economy with stochastic habit formation. working paper, New York University.
- Dai, Q. and K. Singleton (2000, October). Specification analysis of affine term structure models. *Journal of Finance* 50(5), 1943–1978.
- Dai, Q. and K. Singleton (2001). Term structure modeling and fixed-income pricing. Working Paper.
- Duarte, J. (1999). The relevance of parameterization of market price of risk in affine term structure models. Working paper, University of Chicago.
- Duffee, G. R. (2000, November). Term premia and interest rate forecasts in affine models. Working paper, University of California, Berkeley.

- Duffie, D. and R. Kan (1996). A yield-factor model of interest rates. *Mathematical Finance* 6, 379–406.
- Duffie, D., L. Pedersen, and K. Singleton (2000). Modeling credit spreads on sovereign debt. working paper, Stanford University.
- Evans, B. D. (2000, March). Regime shifts, risk and the term structure. Working Paper, Georgetown University.
- Fama, E. (1976). Forward rates as predictors of future spot rates. *Journal of Financial Economics* 3, 361–377.
- Fama, E. (1984a). The information in the term structure. *Journal of Financial Economics* 13, 509–528.
- Fama, E. (1984b). Term premiums in bond returns. *Journal of Financial Economics* 13, 529–546.
- Fama, E. F. and R. R. Bliss (1987). The information in long-maturity forward rates. *American Economic Review* 77(4), 680–692.
- Fisher, M. (1998, January). A simple model of the failure of the expectations hypothesis. Working Paper, Federal Reserve Board.
- Gray, S. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* 42, 27–62.
- Karatzas, I. and S. Shreve (1988). *Brownian Motion and Stochastic Calculus*. New York: Springer-Verlag.
- Kugler, P. (1997). Central bank policy reaction and expectations hypothesis of the term structure. *International Journal of Financial Economics* 2, 217–224.
- Langetieg, T. (1980). A multivariate model of the term structure. *Journal of Finance* 35, 71–97.
- Leippold, M. and L. Wu (1998). Quadratic term structure models. Working paper, Fordham University.
- Litterman, R. and J. Scheinkman (1991). Common factors affecting bond returns. *Journal of Fixed Income* 1, 54–61.
- Longstaff, F., P. Santa-Clara, and E. Schwartz (2000). The relative valuation of caps and swaptions: Theory and empirical evidence. working paper, UCLA.
- Lu, B. (1999). An empirical analysis of the constantinides model of the term structure. Working paper, University of Michigan.
- McCallum, B. T. (1994). Monetary policy and the term structure of interest rates. NBER Working Paper No. 4938.
- Naik, V. and M. H. Lee (1997). Yield curve dynamics with discrete shifts in economic regimes: Theory and estimation. Working paper, Faculty of Commerce, University of British Columbia.

- Roberds, W. and C. Whiteman (1999). Endogenous term premia and anomalies in the term structure of interest rates: Explaining the predictability smile. *Journal of Monetary Economics* 44, 555–580.
- Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics* 5, 177–188.