NBER WORKING PAPER SERIES

EXPECTATION PUZZLES, TIME-VARYING RISK PREMIA, AND DYNAMIC MODELS OF THE TERM STRUCTURE

Qiang Dai
Kenneth J. Singleton

Working Paper 8167
http://www.nber.org/papers/w8167

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2001

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. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

Expectation Puzzles, Time-varying Risk Premia, and Dynamic Models of the Term Structure
Qiang Dai and Kenneth J. Singleton
NBER Working Paper No. 8167
March 2001
JEL No. G0

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 shown to also be consistent with interest rate, feedback rules used by a monetary authority in setting monetary policy.

Qiang Dai
Stern School of Business
New York University
44 West 4th Street
New York, NY 10012
qdai@stern.nyu.edu

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

1 Introduction

Fama [1984b], Fama [1984a], Fama [1984b], 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 (*DTSM*s). 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 *DTSM*s, 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 *DTSM*s that allow us to explain these *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

LPY: the estimated coefficients ϕ_{nT} in the linear projections[1] 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.

LPY 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 ϕ_n are unity, for all n. We show that *LPY* is in fact not puzzling, but rather is generated by at least two important classes of *DTSM*s: (1) a large subclass (though not all) of affine *DTSM*s (Duffie and Kan [1996] and Dai and Singleton [2000] (hereafter *DS*)), and (2) the family of quadratic-Gaussian term structure models (Beaglehole and Tenney [1991] and Ahn, Dittmar, and Gallant [2000]). 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 these models give:

MPY: **(i)** *population* coefficients ϕ_n from the projections of $R_{t+1}^{n-1} - R_t^n$ onto $\frac{1}{n-1}(R_t^n - r_t)$ that largely match the historical pattern of the sample coefficients from *LPY*;[2]

(ii) *sample* coefficients $\phi_{nT}^{\mathcal{R}}$ 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 $\phi_n^{\mathcal{R}} = 1$, for all $n > 1$.

The puzzle *LPY*, repeated in Table 1 for our treasury data set,[3] was anticipated by Fama [1984a] and Fama and Bliss [1987] who argued that excess returns are time-varying

[1] The mnemonic **LPY** stands for "sample-based **L**inear **P**rojection coefficients in **Y**ield-based regressions." Similarly, **MPY** stands for "**M**odel-based **LPY**'s." There are two versions of MPYs, one with risk-premium adjustment, and the other one without.

[2] We are presuming that the historical pattern *LPY* is not spurious, but rather is representative of the pattern of the population ϕ_n. As demonstrated by Bekaert, Hodrick, and Marshall [1997a] and Backus, Foresi, Mozumdar, and Wu [1997], the patterns *LPY* cannot be attributed to small-sample bias in the relevant linear projections. Indeed, they find that the small-sample bias reinforces the puzzle by making the projection coefficients under *LPY* less negative than they would be in the absence of such bias.

[3] We are grateful to Backus, Foresi, Mozumdar, and Wu [1997] for providing the smoothed Fama-Bliss

and typically *positively* correlated with the slope of the yield curve.[4] For a *DTSM* to imply *population* values of the ϕ_n that match the downward sloping pattern in Table 1, it must accurately capture the historical distribution of yields under the actual probability measure P, requirement MPY(i).[5]

Table 1: Campbell-Shiller Long Rate Regression

Estimated slope coefficients ϕ_{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 ϕ_{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
ϕ_{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 MPY(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 MPY(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 MPY 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 on (at least some of) the risk factors directly. For affine *DTSM*s, this specification follows Duffee [2000][6] in extending

data used in our analysis. The data are monthly from February, 1970 through December, 1995. Letting $f_t^n \equiv -\ln(P_t^{n+1}/P_t^n)$ denote the forward rate for one-month loans commencing at date $t+n$, Backus, Foresi, Mozumdar, and Wu [1997] also present empirical evidence against the related expectations null hypothesis of $d_n^f = 1$ in the linear regression $f_{t+1}^{(n-1)} - r_t = \text{constant} + d_n^f \left(f_t^{(n)} - r_t\right) + \text{residual}$, particularly at the shorter maturities.

[4] 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 $R_t^n - r_t$.

[5] As we will see, focusing on ϕ_n makes this a much more demanding requirement than that of explaining LPY using sample ϕ_{nT} obtained with *fitted* yields from a *DTSM*.

[6] Duffee shows that extending the risk-premium specifications in standard affine models improves their forecasting performance and helps in matching the coefficients of variation of yields. He does not formally

those affine models (see *DS* and the references therein) in which market prices of risk are proportional to factor volatilities alone. The market prices of risk in quadratic-Gaussian models are shown to inherently have this flexibility. Our successful models also exhibit non-zero correlations among the risk factors.

We present our quantitative resolution of the expectations puzzle *LPY* in two steps. First, we highlight the role of the factor risk premiums by undertaking an illustrative calibration exercise within one-factor Gaussian and quadratic-Gaussian *DTSM*s. 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, we "extend" the standard Gaussian model by allowing the market price of risk to be an affine function of the state.[7] This state-dependence, in turn, implies that the term premium $p_t^n \equiv f_t^n - E_t[r_{t+n}]$ is an affine function of the slope of the forward curve, $f_t^n - r_t$, where the forward rate is defined as $f_t^n \equiv -\ln(P_t^{n+1}/P_t^n)$. This formulation is reminiscent of the projections in Fama [1984a] and Fama and Bliss [1987] of excess returns onto $f_t^n - r_t$. It turns out that the basic structure of the market price of risk in our quadratic- and extended-Gaussian *DTSM*s is the same (see Ahn, Dittmar, and Gallant [2000] and Section 3.1.2). However, because of the squared state variable in quadratic-Gaussian models, the risk premium p_t^n is an affine function of both the forward slope $f_t^n - r_t$ and the rate r_t, as in two-factor Gaussian models.

We proceed to calibrate the parameters of these one-factor models so that *MPY(i)* is satisfied (the model matches *LPY*) by construction, and then we verify that our specification of risk premiums allows these models to match requirement *MPY(ii)*. This exercise illustrates in a simple and intuitive way the role of state-dependent risk premiums in resolving expectations puzzles: factor volatilities are constant in Gaussian models, so any time-variation of the market prices of risk is due to their direct dependence on the risk factors.[8] Moreover, we find that (for the purpose of matching *MPY*(ii)) one-factor Gaussian and quadratic-Gaussian models offer essentially equivalent flexibility – neither seems to dominate the other.

Of course, we do not presume that one-factor models capture the rich variation over time in yield curves. Nor is finding admissible parameters that match *MPY*(ii), conditional on *MPY*(i) being satisfied, the same as finding that *MPY* is matched at the maximum likelihood (*ML*) estimates of the model. Both of these concerns are addressed in an extensive exploration of *MPY* within the families of three-factor "affine" *DTSM*s.[9] We fit all four of the canonical three-factor models (see *DS* and Duffee [2000]) by the method of full-information

address the matching of *MPY*.

[7]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 *MPY*. We are grateful to Greg Duffee for bringing this unpublished manuscript to our attention.

[8]Backus, Foresi, Mozumdar, and Wu [1997] argue that the expectations puzzles can be resolved using a "negative CIR" process. Our resolution shares some of the same features as their negative *CIR* process. However, we believe that the models studied here more clearly highlight the essential features of *DTSM*s that generate *LPY*. In addition, we provide a link to, and reinterpretation of, the modeling implications of the forward-rate regressions in Fama and Bliss [1987].

[9]Like the special case of Gaussian models, the entire family of affine *DTSM*s imply that optimal forecasts of excess returns take the form of the linear projections extensively studied in the literature on the expectations hypothesis.

ML and then compare the relevant population and sample versions of the risk-adjusted ($\phi_n^{\mathcal{R}}$) and unadjusted (ϕ_n) projection coefficients. Additionally, we assess the relative importance of time-varying risk premiums, time-varying factor volatilities, and non-zero factor correlations in matching MPY.

Following *DS* we classify three-factor affine models according to the number of state variables m that drive the volatilities of all three state variables. *CIR*-style models (models in which the state variables follow square-root diffusions) have risk premiums that are proportional to factor volatilities, so $m = 3$. As such, it is only through time-varying volatilities that risk premiums can vary.[10] Consequently, they do not meet our heuristic conditions for matching MPY. In fact, we show in Section 4 that a three-factor *CIR*-style model, evaluated at maximum likelihood estimates of the parameters, is wholly incapable of matching MPY.[11] In contrast, we find that multi-factor Gaussian models– with constant conditional volatilities ($m = 0$), but state-dependent market prices of risk– match MPY strikingly well at the historical *ML* estimates of the model.

Lying between the cases of CIR-style and Gaussian models are the affine models with $0 < m < 3$. 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 MPY. Further exploration of the reasons reveals a tension in matching simultaneously the historical properties of the conditional means and variances of yields within affine $DTSM$s.

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, and undertakes the calibration exercises with one-factor models. A more formal and extensive empirical assessment of the fit of three-factor affine $DTSM$s to MPY is presented in Section 4. Concluding remarks are presented in Section 5. Technical details are collected in an appendix.

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 $DTSM$s 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

[10] See *DS* for a discussion of canonical square-root $DTSM$s, and Chen and Scott [1993], Pearson and Sun [1994], and Duffie and Singleton [1997] for empirical applications.

[11] These findings complement those in Roberds and Whiteman [1999] who show (see their Figures 4 and 6) that one- and two-factor *CIR*-style models cannot match MPY(i) for their sample period and treasury yields, even when the parameters of their $DTSM$s are calibrated to match their counterparts of the ϕ_{nT}. Backus, Foresi, Mozumdar, and Wu [1997] demonstrate analytically that, in order for a (one-factor) *CIR*-style model to potentially match MPY(i), it must imply a *downward* sloping term structure of mean forward spreads $\{E[f_t^n - r_t]\}$, contrary to historical experience.

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

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

where E_t denotes expectation conditioned on date t information. Rearranging (1) gives the fundamental relation[12]

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

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

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

and the forward term premium

$$p_t^n \equiv f_t^n - E_t[r_{t+n}]. \tag{4}$$

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:

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

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},

[12] 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].

and an "expectations" part:[13]

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

$$D^{*n}_{t+1} = -(n-1)(c^{n-1}_{t+1} - c^{n-1}_t) + p^{n-1}_t. \quad (7)$$

Since the $(E_t r_{t+i} - E_{t+1} r_{t+i})$ have zero date-t conditional means,[14] e^n_t depends only on the premium term D^{*n}_{t+1}:[15]

$$e^n_t = E_t[D^{*n}_{t+1}] = -(n-1)E_t[c^{n-1}_{t+1} - c^{n-1}_t] + p^{n-1}_t. \quad (8)$$

Thus, we can replace D^n_{t+1} by D^{*n}_{t+1} in (2) to obtain

$$E_t\left[R^{n-1}_{t+1} - R^n_t + \frac{1}{n-1}D^{*n}_{t+1}\right] = \frac{1}{n-1}(R^n_t - r_t). \quad (9)$$

From (9) it follows that the projection of the "premium-adjusted" change in yields,

$$R^{n-1}_{t+1} - R^n_t - (c^{n-1}_{t+1} - c^{n-1}_t) + \frac{1}{n-1}p^{n-1}_t, \quad (10)$$

onto the (scaled) slope of the yield curve, $(R^n_t - r_t)/(n-1)$, has a coefficient of one.[16] Yield-based regressions under the expectations hypothesis are obtained by setting the risk premiums in (10) to constants.

[13] Some of the intermediate steps in this derivation are:

$$\begin{aligned} D^n_{t+1} &\equiv nR^n_t - (n-1)R^{n-1}_{t+1} - r_t = nc^n_t - (n-1)c^{n-1}_{t+1} + \sum_{i=1}^{n-1}(E_t r_{t+i} - E_{t+1} r_{t+i}) \\ &= -(n-1)(c^{n-1}_{t+1} - c^{n-1}_t) + \sum_{j=0}^{n-1} p^j_t - \sum_{j=0}^{n-2} p^j_t + \sum_{i=1}^{n-1}(E_t r_{t+i} - E_{t+1} r_{t+i}). \end{aligned}$$

[14] The projection of $\sum_i (E_t r_{t+i} - E_{t+1} r_{t+i})$ onto date t information will in general be zero only if the model correctly captures the dynamic properties of r_t, in our case the one-month treasury bill rate. We expand on this point in Section 4.5 in explaining why three-factor models fail to match MPY for small n.

[15] Equation (8) implies that $E[e^n_t] = E[p^{n-1}_t] = E[f^{n-1}_t - r_t]$, where the second equality follows from the definition of p^{n-1}_t 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)

$$p^{n-1}_t = E_t[D^n_{t+1}] + E_t[D^{n-1}_{t+2} - D^{n-1}_{t+1}] + \ldots + E_t[D^2_{t+n-1} - D^2_{t+n-2}]$$

to conclude that the forward rate f^{n-1}_t "contains" market expectations about the holding period return D^n_{t+1}. He then computed the sample means of p^{n-1}_t and $(f^{n-1}_t - r_t)$ 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.

[16] There is an analogous set of yield projections for the forward rates. Specifically, from the definition of p^n_t it follows that $f^{n-1}_{t+1} - f^n_t = E_{t+1}(r_{t+n} - E_t[r_{t+n}]) + (p^{n-1}_{t+1} - p^n_t)$. Subtracting r_t from both sides, rearranging, and taking conditional expectations gives $E_t[f^{n-1}_{t+1} - r_t] = (f^n_t - r_t) + (E_t[p^{n-1}_{t+1}] - p^n_t)$. Thus, projection of the "premium-adjusted" forward rate, $(f^{n-1}_{t+1} - r_t - (p^{n-1}_{t+1} - p^n_t))$, onto $(f^n_t - r_t)$ also gives a slope coefficient of one. In our empirical analysis we will focus on (9). Results for forward rate projections are available from the authors upon request.

3 Risk Premiums, *DTSM*s, and *LPY*

The challenges set forth by Fama and the studies of *LPY* in the literature on the expectations hypothesis are statements about correlations among yields and, as such, are naturally studied using linear projections. Therefore, in attempting to match *MPY* within *DTSM*s 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 *DTSM*s.

Consider first the case of affine *DTSM*s with the instantaneous short rate given by $r_0(t) = a_0 + b_0'Y(t)$ and the N-dimensional state vector Y following, under measure P, the affine diffusion

$$dY(t) = \kappa\,(\theta - Y(t))\,dt + \Sigma\sqrt{S(t)}dW(t), \tag{11}$$

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

$$[S(t)]_{ii} = \alpha_i + \beta_i'Y(t). \tag{12}$$

The risk-neutral representation of $Y(t)$ used in pricing is obtained by subtracting $\Sigma\sqrt{S(t)}\Lambda(t)$ from the drift of (11), where $\Lambda(t)$ is the vector of "market prices of risk." Standard formulations of affine *DTSM*s "close" this model by assuming that $\Lambda(t)$ is proportional to $\sqrt{S(t)}$:

$$\Lambda(t) = \sqrt{S(t)}\ell_0, \tag{13}$$

where ℓ_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 follows *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 $(\beta_1, \ldots, \beta_N)$ is m).

Since at the heart of matching *MPY* 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

$$\Lambda(t) = \sqrt{S(t)}\lambda^0 + \sqrt{S^-(t)}\lambda^Y Y(t), \tag{14}$$

where λ^0 is an $N \times 1$ vector and λ^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 $i = m+1, \ldots, N$, 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 (13). That the added flexibility of $\lambda^Y \neq 0$ translates directly into flexibility in explaining *LPY* follows from the observation that the *instantaneous* expected excess return on a τ-period zero-coupon bond is

$$\mu^e(t,\tau) = -B(\tau)'\Sigma\sqrt{S(t)}\Lambda(t), \tag{15}$$

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 μ^e and, hence, the matching of *MPY*.

At the same time, we see that the dynamic properties of excess returns are also influenced by the degree of factor correlations Σ, 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 λ^Y in (14) is zero so the risk premiums are constrained to be of the form (13). For the cases $m < N$, then the first m rows and columns of λ^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 λ^Y is unconstrained. One of the issues we explore empirically is the relative contributions of non-diagonal Σ and non-zero elements of λ^Y to matching *MPY* in multi-factor models.

Another family of *DTSM*s with the potential to match *MPY* is the family of N-factor quadratic-Gaussian models with the instantaneous short rate r_0 given by $r_0(t) = a_0 + Y'b_0 + Y'c_0Y$, where c_0 is an $N \times N$ symmetric matrix of constants and Y follows the Gaussian special case of (11) with $S(t) = I_N$. Ahn, Dittmar, and Gallant [2000] show that the market price of risk in their canonical N-factor quadratic-Gaussian model takes exactly the same form as (14) for the Gaussian case of $m = 0$.

Our strategy for assessing whether a specific *DTSM* can match *MPY* 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 *MPY*(i), we compare the model-implied population coefficients,

$$\phi_n \equiv \frac{\mathrm{cov}(R_{t+1}^{n-1} - R_t^n, (R_t^n - r_t)/(n-1))}{\mathrm{var}((R_t^n - r_t)/(n-1))}, \tag{16}$$

to their sample counterparts displayed in Table 1. The population ϕ_n are computed by treating the estimates of the model parameters as "truth" and then using analytic formulas to compute the second moments in (16). The data enters these calculations only indirectly through the estimates of the model parameters.

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

$$\phi_n^{\mathcal{R}} \equiv \frac{\mathrm{cov}(R_{t+1}^{n-1} - R_t^n + D_{t+1}^{*n}/(n-1), (R_t^n - r_t)/(n-1))}{\mathrm{var}((R_t^n - r_t)/(n-1))} \tag{17}$$

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 pattern of *LPY*.

3.1 *LPY* and One-Factor Models

To highlight the role of the market prices of risk in generating time-varying risk premiums that are consistent with the sample distribution of the data, we proceed initially with a simple calibration exercise with one-factor affine and quadratic-Gaussian models. Our calibration strategy chooses model parameters so that *MPY*(i) is satisfied, and then we check to see whether we can match *MPY*(ii).

We focus on forward term premiums p_t^n, which is is equivalent (see footnote 15) to parameterizing the dependence of $e_t^n = E_t[D_{t+1}^n]$ on agents information set, as in Fama [1984a] and Fama and Bliss [1987]. From (2) and (8), we can write

$$E_t\left[R_{t+1}^{n-1} - R_t^n\right] = \frac{1}{n-1}(R_t^n - r_t) + E_t\left[c_{t+1}^{n-1} - c_t^{n-1}\right] - \frac{1}{n-1}p_t^{n-1}. \tag{18}$$

Thus, given parameterizations of the p_t^n, we have fully determined the yield projections as well. As shown formally in appendices, all of our illustrative one-factor models imply that

$$p_t^n = \delta_n + \alpha_n(f_t^n - r_t), \tag{19}$$

where the (δ_n, α_n) are model-dependent functions of the underlying primitive parameter vector ς describing the state vector and the dependence of $r_0(t)$ and $\Lambda(t)$ on $Y(t)$.

For the one-factor extended and quadratic Gaussian models, calibration is based on the moment equation (see footnote 16)

$$E_t[f_{t+1}^{n-1} - r_t] - (f_t^n - r_t) + (E_t[p_{t+1}^{n-1}] - p_t^n) \equiv E_t[u_{t+1}] = 0. \tag{20}$$

For these models, the risk premium parameter α_n turns out to depend only on the scalar parameters κ and λ^Y. However, δ_n depends, as well, on other parameters of our illustrative models. Therefore, we proceed by "concentrating" out δ_n from the empirical analysis using the observation that (19) and the assumption of stationarity imply $\delta_n = (1-\alpha_n)E[f_t^n - r_t]$. Thus, if the model is correctly specified, δ_n can be inferred from α_n and the sample means of the forward-spot spread and one-period short rate. This procedure forces our one-factor models to match *LPY*, while ignoring the restrictions implicit in the dependence of δ_n on ς. In so doing, we highlight the roles of the model parameters λ^Y and κ in matching *MPY*(ii).

The parameters (κ, λ^Y) were chosen to minimize a standard *GMM* objective function (Hansen [1982]) based on the moment conditions

$$E[u_{t+1}^n z_t] = 0, \text{ with } z_t = (f_t^n - r_t, r_t)', \; n = 6, 12, 24, 60, 84, 120. \tag{21}$$

Only a subset of the maturities between 1 and 120 months are used because the smoothed Fama-Bliss dataset is interpolated.

3.1.1 One-Factor Gaussian Models

In the one-factor Gaussian *DTSM* the instantaneous short rate is given by $r_{0t} = a_0 + b_0 Y_t$; Y_t follows a one-dimensional Gaussian process (11) with $N = 1$, $S(t) = 1$, and $\Sigma = 1$;[17] and p_t^n can be represented as in (19) with[18]

$$\alpha_n = \frac{e^{-\kappa n\Delta} - e^{-\tilde{\kappa} n\Delta}}{1 - e^{-\tilde{\kappa} n\Delta}}, \tag{22}$$

$$\delta_n = (1-\alpha_n)(A_n^\Delta - a_1) + (1-\alpha_n)(B_n^\Delta - b_1)\theta, \tag{23}$$

[17] The latter is a normalization, imposed without loss of generality (*DS*).

[18] The linearity of the one-factor Gaussian model implies that the yield and forward risk premiums are both affine in Y, so we are free to parameterize p_t^n as an affine function of the a yield spread.

where $\tilde{\kappa} = \kappa + \lambda^Y$ is the mean reversion coefficient under the risk neutral measure, Δ is the length of each period, A_n^Δ and B_n^Δ are the intercept and the factor loading on the one-period forward rate delivered n periods hence, and a_1 and b_1 are the intercept and the factor loading on the one-period zero coupon yield (the short rate). The precise definitions of these loadings in terms of basic model parameters are given in Appendix A. This one-factor model maps directly, using (7), to Fama [1976]'s regression model of excess returns, which implicitly assumes that, in our notation, $E[D_{t+1}^n | f_t^{n-1} - r_t] = E[D_{t+1}^{*n} | f_t^{n-1} - r_t]$ is linear in $f_t^{n-1} - r_t$. Of course Fama does not impose the dynamic restrictions (22) and (23), because (19) is essentially the starting point of his empirical analysis.

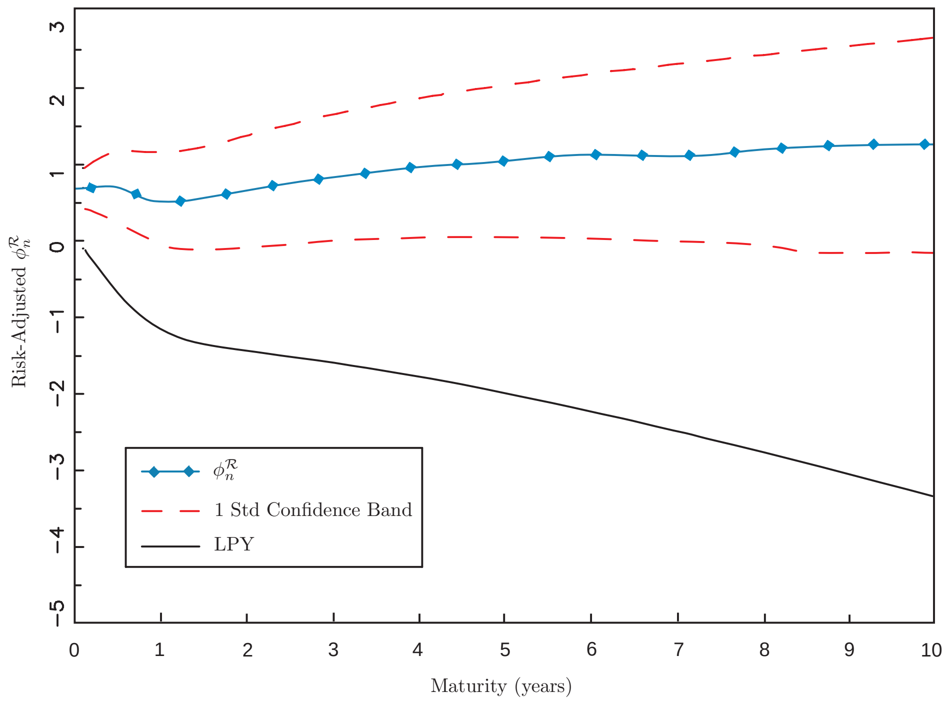

Figure 1: Projections Coefficients $\phi_n^{\mathcal{R}}$ Implied by the One-factor Gaussian Model.

The calibrated values of (κ, λ^Y) are $(0.0012, 0.0008)$. Using these values, we compute the model-implied $\phi_{nT}^{\mathcal{R}}$ and plot them in Figure 1, along with the ϕ_{nT} from Table 1 estimated under the null hypothesis that p_t^n is constant for all n. For all but the shortest maturities, $\phi_{nT}^{\mathcal{R}}$ lie within one sample standard error of one.[19] Thus, our (calibrated) risk premiums largely match MPY and, thereby, resolve the expectations puzzles.[20]

[19] These standard error bands reflect the sampling variation of the parameter estimates, but not of the sample moments used in estimating $\phi_{nT}^{\mathcal{R}}$. Accounting for the latter would most likely widen these bands.

[20] In the light demonstrated limitations of one-factor *CIR*-style models (Backus, Foresi, Mozumdar, and Wu [1997]), a natural question at this junction is: can we generate an upward sloping mean yield curve with our extended, one-factor Gaussian model? As demonstrated in an earlier version of this paper, this does not present a serious challenge for the one-factor Gaussian model, because the three parameters θ, λ^0, and σ have not been used in matching MPY. The reason the Gaussian model out-performs the square-root model is the restrictive form of the market price of risk in the latter, together with the fact that (as we discuss more extensively below) allowing for time-varying volatility is not central to matching MPY.

3.1.2 One-Factor Quadratic-Gaussian Models

Shifting attention to the quadratic-Gaussian models, the zero coupon bond price $P(t,\tau)$ is given by

$$-\log P(t,\tau) = A(\tau) + Y'B(\tau) + Y'C(\tau)Y \tag{24}$$

where

$$B(\tau) = \left(\frac{\tilde{\kappa}}{\Gamma}\frac{e^{\Gamma\tau}-1}{e^{\Gamma\tau}+1}+1\right)Q(\tau)\,b_0 + \left(\frac{2\tilde{\kappa}\tilde{\theta}}{\Gamma}\frac{e^{\Gamma\tau}-1}{e^{\Gamma\tau}+1}\right)Q(\tau)\,c_0 \tag{25}$$

$$C(\tau) = Q(\tau)\,c_0 \tag{26}$$

$$Q(\tau) = \frac{e^{2\Gamma\tau}-1}{(\Gamma+\tilde{\kappa})(e^{2\Gamma\tau}-1)+2\Gamma}, \tag{27}$$

with $\Gamma^2 = \tilde{\kappa}^2 + 2c_0\sigma^2$. The expected short rate is

$$E_t[r_{t+n}] = \mu_n + \nu_n Y_t + \omega_n Y_t^2, \tag{28}$$

with the coefficients expressed as functions of the primitive parameters in Appendix B. Letting $a_1 \equiv A(\Delta)/\Delta$, $b_1 \equiv B(\Delta)/\Delta$, $c_1 \equiv C(\Delta)/\Delta$, $A_n^\Delta \equiv [A((n+1)\Delta) - A(n\Delta)]/\Delta$, $B_n^\Delta \equiv [B((n+1)\Delta) - B(n\Delta)]/\Delta$, and $C_n^\Delta \equiv [C((n+1)\Delta) - C(n\Delta)]/\Delta$, we show in Appendix B that the forward risk premiums in this model can be expressed as $p_t^n = \delta_n + \alpha_n(f_t^n - r_t) + \beta_n r_t$ with coefficients

$$\alpha_n = 1 - \frac{\nu_n/b_1 - \omega_n/c_1}{B_n^\Delta/b_1 - C_n^\Delta/c_1} \tag{29}$$

$$\beta_n = (B_n^\Delta/b_1 - \nu_n/b_1) - (B_n^\Delta/b_1 - 1)\alpha_n. \tag{30}$$

Thus, the one-factor quadratic Gaussian model implies "two-factor" risk premium model in that p_t^n depends linearly on both f_t^n and r_t. In fact, it is easy to verify that the forward risk premiums in the two-factor Gaussian model with $\lambda^Y \neq 0$ and one-factor quadratic-Gaussian model have the same structure, but they are not identical because the *dynamic restrictions* imposed on the parameters α_n and β_n are different.

Figure 2 displays the forward-rate projection coefficients $\phi_{nT}^{\mathcal{R}f}$ (see footnotes 3 and 16) implied by the one-factor quadratic-Gaussian model calibrated with the same moments used in estimating the one- and two-factor Gaussian models. We see that the quadratic-Gaussian model also does a good job of matching the forward version of *MPY*(ii) at all but the shortest maturities. What is perhaps most striking about Figure 2 is that the $\phi_{nT}^{\mathcal{R}f}$ from the one-factor extended Gaussian and quadratic-Gaussian *DTSM*s are virtually on top of each other. In other words, with regard to their abilities to match *MPY*, these two one-factor models perform equally well. This is because the estimated mean reversion coefficients and the quadratic constant c_0 are small.[21]

[21] Strictly speaking, the quadratic model does not nest the one-factor Gaussian model, since in the limit as $c_0 \to 0$ the forward risk premium model implied by the quadratic model maintains its two-factor structure, but with $f_t^n - r_t$ and r_t being perfectly collinear. Consequently, α_n and β_n are not identified in this limiting case.

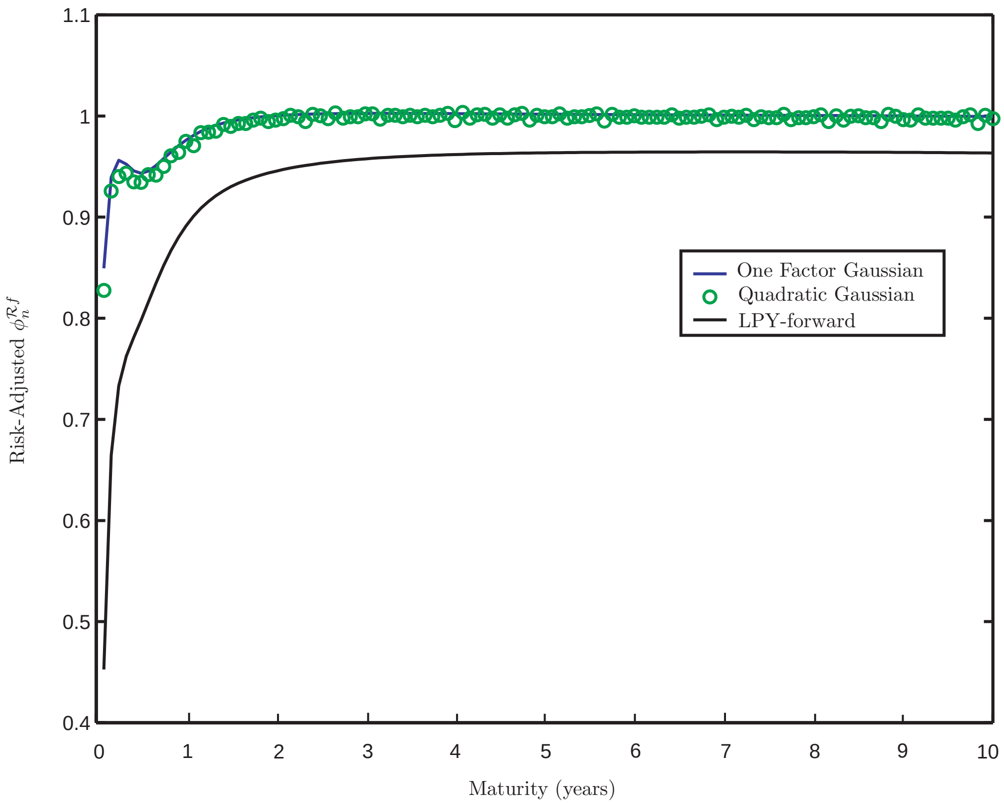

Figure 2: Model-implied Estimates of Forward Projection Coefficients $\phi_n^{\mathcal{R}f}$ From the One-factor Gaussian and Quadratic-Gaussian *DTSM*s.

4 *MPY* and Multi-Factor Affine Models

The preceding calibration of one-factor models, while demonstrating that *MPY* can be matched by judicious choice of admissible parameters in certain *DTSM*s, leaves open the question of whether we can *simultaneously* match *MPY* and other, higher-order moments of yield distributions. We turn next to a more demanding assessment of affine *DTSM*s by computing maximum likelihood (*ML*) estimates of models within the families $A_m(3) : 0 \leq m \leq 3$ and examining whether the implied risk premiums, computed at the *ML* estimates, resolve the expectations puzzles. Our shift to three-factor models is in recognition of the widely documented observation that more than one risk factor is necessary to describe yield curve dynamics.

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

$$r(t) = \delta_0 + Y_1(t) + Y_2(t) + Y_3(t), \tag{31}$$

where $Y(t)$ follows the affine diffusion (11) with volatilities given as in (12) with the normalization $\alpha_i + \beta_{i1} = 1$, Σ is a diagonal matrix of free parameters, and

$$\kappa = \begin{pmatrix} \mathcal{K}_{11} & 0 \\ \mathcal{K}_{21} & \mathcal{K}_{22} \end{pmatrix}, \quad \theta = \begin{pmatrix} \Theta_1 \\ 0 \end{pmatrix}, \tag{32}$$

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 Θ_1 is an $m \times 1$ vector of free parameters.[22] For the case $m = 0$, κ 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.[23] The market prices of risk are given by (14) with the first m rows of λ^Y set to zeros in the $A_{mC}(3)$ model to assure admissibility, and the λ^Y_{i1} normalized to zero for $i = m+1, \ldots, 3$.

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 respective observed and model-implied yields of all maturities are equal on average. 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, $(\kappa_{13}, \lambda^0_1, \lambda^0_2, \lambda^Y_{21}, \lambda^Y_{13}, \lambda^Y_{33})$. This led to virtually the same value of the likelihood function, so we henceforth study this constrained version of $A_{0C}(3)$.[24]

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.[25] Focusing first on κ 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 κ_{ii}

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

[23] 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.

[24] For the case of model $A_{1C}(3)$, we found little change by constraining $\lambda^Y_{32} = 0$, 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 $A_{1C}(3)$.

[25] We have suppressed the estimates of δ_0, θ (the long-run means), and the β_i (the coefficients on Y in the factor volatility specifications), because they do not play central roles in subsequent discussions of matching *MPY*. They are available from the authors upon request.

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]). (Continuing with this analogy, the factor with the intermediate (fastest) rate of mean reversion is most highly correlated with the "slope" ("curvature") principal component.) In contrast, the minimal κ_{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 the risk-neutral measure (hereafter Q-nonstationary).[26]

Param.	$A_{0C}(3)$	$A_{1C}(3)$	$A_{2C}(3)$	$A_{3C}(3)$
κ_{11}	3.012 (.403)	0.002 (.001)	0.628 (.088)	2.714 (.121)
κ_{21}	0	0.204 (.075)	0	0
κ_{31}	0	0.295 (.175)	-5.55 (1.18)	0
κ_{12}	2.081 (.463)	0	0	0
κ_{22}	0.378 (.122)	0.983 (.355)	0.006 (.002)	0.005 (.002)
κ_{32}	0	-2.740 (.953)	-0.349 (.320)	0
κ_{23}	-0.154 (.081)	-0.403 (.204)	0	0
κ_{33}	0.466 (.110)	2.510 (.499)	2.01 (.285)	0.526 (.082)
Σ_{11}	0.011 (.002)	0.030 (.000)	0.055 (.005)	0.063 (.002)
Σ_{22}	0.005 (.002)	0.048 (.011)	0.028 (.000)	0.030 (.000)
Σ_{33}	0.025 (.001)	0.080 (.013)	0.192 (.023)	0.105 (.013)
λ_1^0	0	-0.256 (.049)	-1.29 (1.58)	-0.669 (.022)
λ_2^0	0	-6.521 (.772)	-0.389 (.048)	-0.328 (.090)
λ_3^0	-0.361 (.269)	-8.490 (1.42)	-0.514 (2.41)	-0.646 (.764)
λ_{11}^Y	-101.8 (36.91)	0	0	0
λ_{31}^Y	158.0 (30.90)	0	-0.996 (2.41)	0
λ_{12}^Y	-100.6 (27.21)	0	0	0
λ_{22}^Y	-58.09 (26.12)	-19.35 (9.18)	0	0
λ_{32}^Y	49.81 (22.50)	9.920 (5.17)	-0.643 (.938)	0
λ_{23}^Y	21.91 (8.92)	18.36 (5.51)	0	0
λ_{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. κ_{13}, λ_{13}^Y, and λ_{21}^Y are zero in all four models.

Finally, some of the estimated elements of λ^Y are statistically different from zero in models $A_{0C}(3)$ and $A_{1C}(3)$, but all of the non-zero elements in λ^Y for model $A_{2C}(3)$ are insignificant 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 *MPY* with models $A_{0C}(3)$ and $A_{1C}(3)$, but the point estimates suggest extended risk premiums are less important for model $A_{2C}(3)$.

[26] Q-nonstationary state vectors do not present a conceptual problem for valuation or estimation.

4.1 Matching $\boldsymbol{MPY}$(i)

Matching MPY(i) requires a *positive* correlation between excess returns and the term spread, which is typically associated with a *negative* correlation between r and the expected excess return $\mu^e(t, \tau)$.[27] 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 LPY 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.

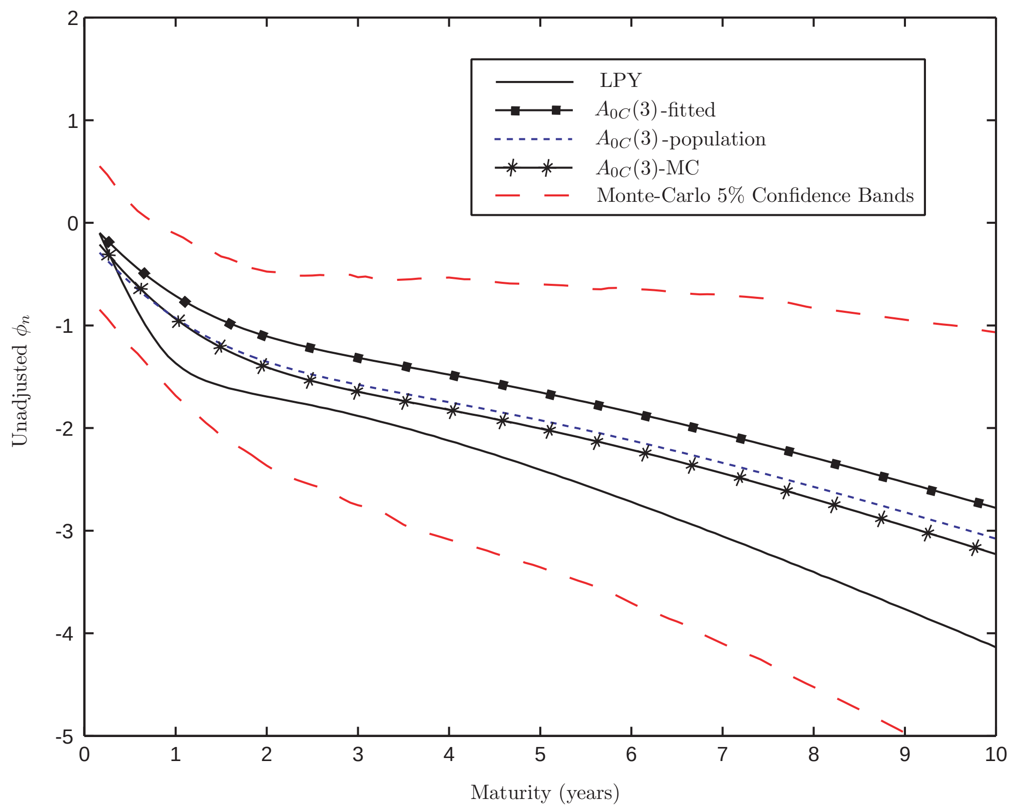

Figure 3: Model Implied Estimates of Unadjusted ϕ_n from the Projection of $R_{t+1}^{(n-1)} - R_t^n$ onto $(R_t^n - r_t)/(n-1)$ for the $A_{0C}(3)$ Models.

In assessing the goodness-of-fit of a model based on whether it matches the pattern LPY we are led to confront several important issues, including: (i) should we compare the model-implied population ϕ_n or sample ϕ_n computed from fitted yields to the historical ϕ_{nT} in Table 1; and (ii) might the ϕ_n computed from fitted yields be "biased" due to the highly persistent nature of the fitted yields? To motivate our answers to these questions, consider model $A_{0C}(3)$ and the projection coefficients implied by this model that are displayed

[27]This is unambiguously true for a one-factor model with mean-reverting short rate under the risk-neutral measure. It is usually true for more general models under reasonable parameter values.

in Figure 3. Focusing first on the issue of small sample bias, the graph labeled "$A_{0C}(3)$-population" displays the population ϕ_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 the ϕ_{nT} generated by this model, 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 ϕ_n were estimated. The mean of these estimates across the five hundred samples (displayed in Figure 3 as "$A_{0C}(3)$-MC") lies very close to "$A_{0C}(3)$-population," suggesting that small-sample biases are negligible for this model. Moreover, historical estimates of these projection coefficient from Table 1 (displayed as graph "LPY" in Figure 3) lie well inside the Monte Carlo confidence bands – the 5% quantiles of the small sample distribution of the ϕ_{nT}. Thus, based on the model-implied population results, we conclude that model $A_{0C}(3)$ is successful in matching MPY(i).

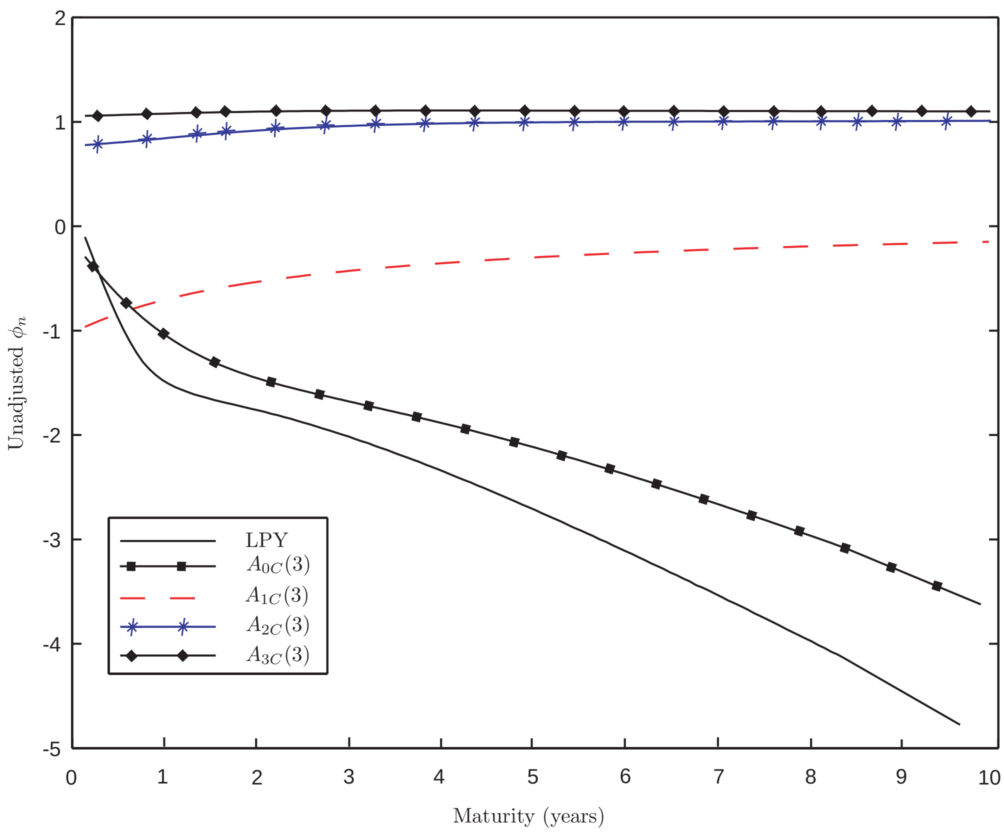

Figure 4: Model-Implied Population Estimates of ϕ_n from the Projection of $R_{t+1}^{(n-1)} - R_t^n$ onto $(R_t^n - r_t)/(n-1)$ for the $A_{mC}(3)$ Models.

From Figure 4 we see that the remaining three $A_{mC}(3)$ models do not fair nearly as well at matching MPY(i) in the population. Indeed, none of these graphs have the characteristic downward sloping pattern of LPY 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 MPY(i) in the population.

A very different, and we feel misleading, assessment comes from analysis of the "fitted" ϕ_n

obtained by inverting the model for the fitted state variables, computing model-implied fitted zero-coupon bond yields, and then estimating the regressions for Table 1 with these fitted yields. The results for model $A_{0C}(3)$ (graph "$A_{0C}(3)$-fitted" in Figure 3) show that it looks very much like "$A_{0C}(3)$-population." 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 argued that their population counterparts look very different than "$A_{0C}(3)$-population." We conclude that assessments of fit based on fitted yields can give very misleading impressions of the actual population distributions implied by *DTSM*s and, therefore, we henceforth focus on the population ϕ_n. [28]

4.2 Matching *MPY*(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 it effectiveness at matching *MPY*(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 (17) 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 5 displays the model-implied $\phi_n^{\mathcal{R}}$ along with the historical results from Table 1 ("LPY"). We see that models $A_{0C}(3)$ and $A_{1C}(3)$ imply risk premiums that fully meet the challenge *MPY*(ii), at least beyond maturities of about two years. In contrast, as with *MPY*(i), models $A_{2C}(3)$ and $A_{3C}(3)$ fail entirely to match *MPY*(ii); adjusting for risk premiums in these models gives virtually the same results as in the unadjusted regressions. Since the "LPY" 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 4.5 further discussion of the fit to under two years to maturity.

4.3 Further Observations On Matching *MPY*

Why does model $A_{0C}(3)$ outperform the other canonical models in its ability to match both dimensions of *MPY*? 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,[29] while Y is P-stationary, but Q-nonstationary in models $A_{mC}(3)$, $m \geq 1$.

[28] This point was stressed in a related context by *DS*. The reason the "fitted" graphs are misleading is that the fitted yields give the model the benefit of using the actual data, through the inversion of the model to obtain the fitted state variables, when matching *MPY*(i).

[29] The relatively large values of the κ_{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 κ 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 *MPY*. 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.

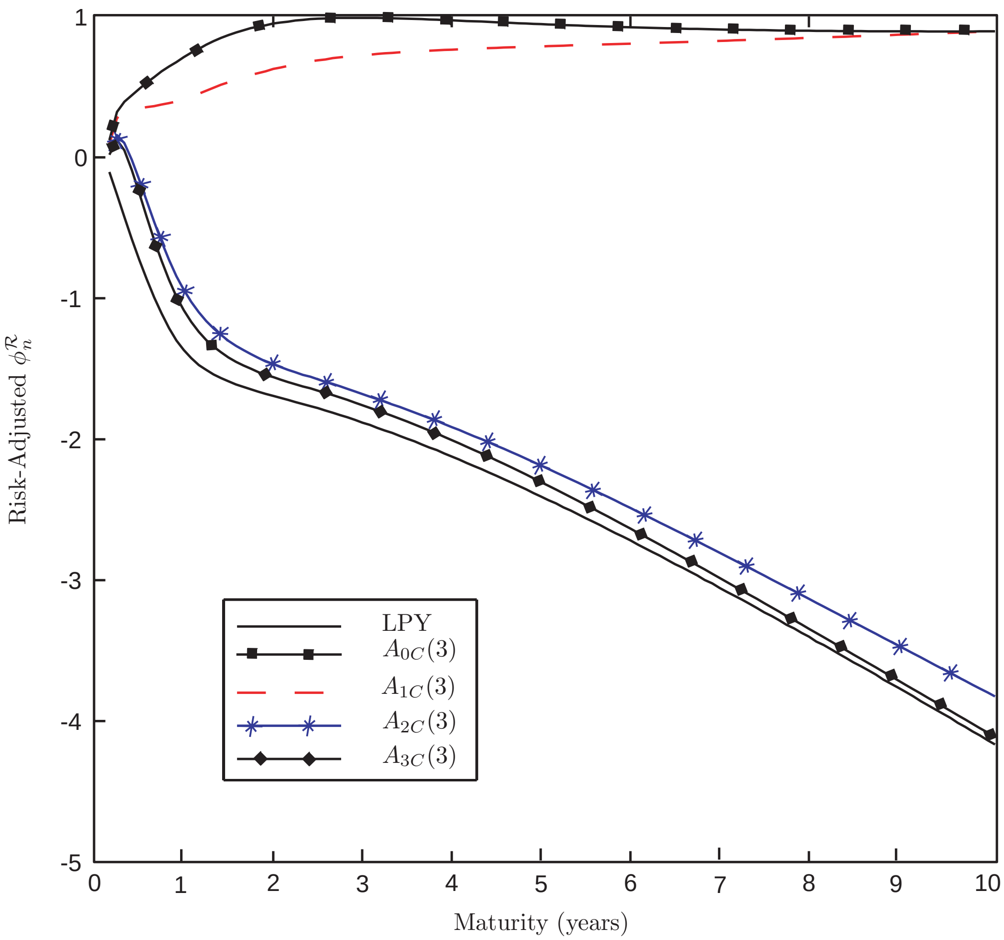

Figure 5: Model-Implied Estimates of $\phi_n^{\mathcal{R}}$ From the Models $A_{mC}(3)$.

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 κ are presented. To gain some insight into the role of Q-nonstationary Y for matching MPY, we re-estimated model $A_{1C}(3)$ under the constraint that Y be Q-stationary, model $A_{1S}(3)$, and the resulting κ matrix in 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 MPY are displayed in Figure 6. The population ϕ_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 $\phi_n^{\mathcal{R}}$ are notably less than unity for all maturities, suggesting that the model-implied risk premiums no longer have the requisite properties to match MPY(ii).

Still unresolved in the role of stochastic volatility in matching MPY within the $A_1(3)$ family. The third model, $A_{1L}(3)$, displayed in Table 3 and Figure 6 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

Param.	$A_{0C}(3)$	$A_{1C}(3)$	$A_{1S}(3)$	$A_{1L}(3)$
κ_{11}	3.01 (1.91)	0.002 (-0.005)	0.653 (0.574)	1.95 (1.94)
κ_{21}	0	0.204 (-0.107)	-5.45 (-6.33)	-0.44 (0.08)
κ_{31}	0	0.295 (-0.384)	0.029 (0.039)	1.01 (3.99)
κ_{12}	2.08 (0.99)	0	0	0
κ_{22}	0.378 (0.062)	0.983 (0.062)	1.50 (1.80)	0.13 (0.002)
κ_{32}	0	-2.740 (-1.95)	-0.022 (-0.011)	0.25 (-0.03)
κ_{23}	-0.154 (-0.035)	-0.403 (0.471)	-16.6 (-38.2)	-0.26 (0.005)
κ_{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 κ from Constrained $A_m(3)$ Models. Risk-neutral κ's are given in parentheses.

factor Y_1 becoming the "curvature" factor (the one with the fastest rate of mean reversion). From Figure 6 we see that this model matches quite well *both* MPY(ii) for longer maturities and the downward sloping pattern of the unadjusted ϕ_n, MPY(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 MPY is not the presence of stochastic volatility *per se*. Rather it appears to be the *desire* of the likelihood function to trade off matching MPY, 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 MPY, however. In other words, affine models do not appear to have the flexibility to simultaneously fit MPY and the volatility properties of returns.[30]

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 MPY with market prices of risk of the form (14) 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 nonzero admissible elements of λ^Y are all insignificantly different from zero and this shows up in Figure 5 as risk-adjusted projection coefficients that are nearly the same as in model $A_{3C}(3)$.

[30]Complementary evidence on the limitations of $A_1(3)$ models for fitting the conditional variances of yields is presented in Ahn, Dittmar, and Gallant [2000].

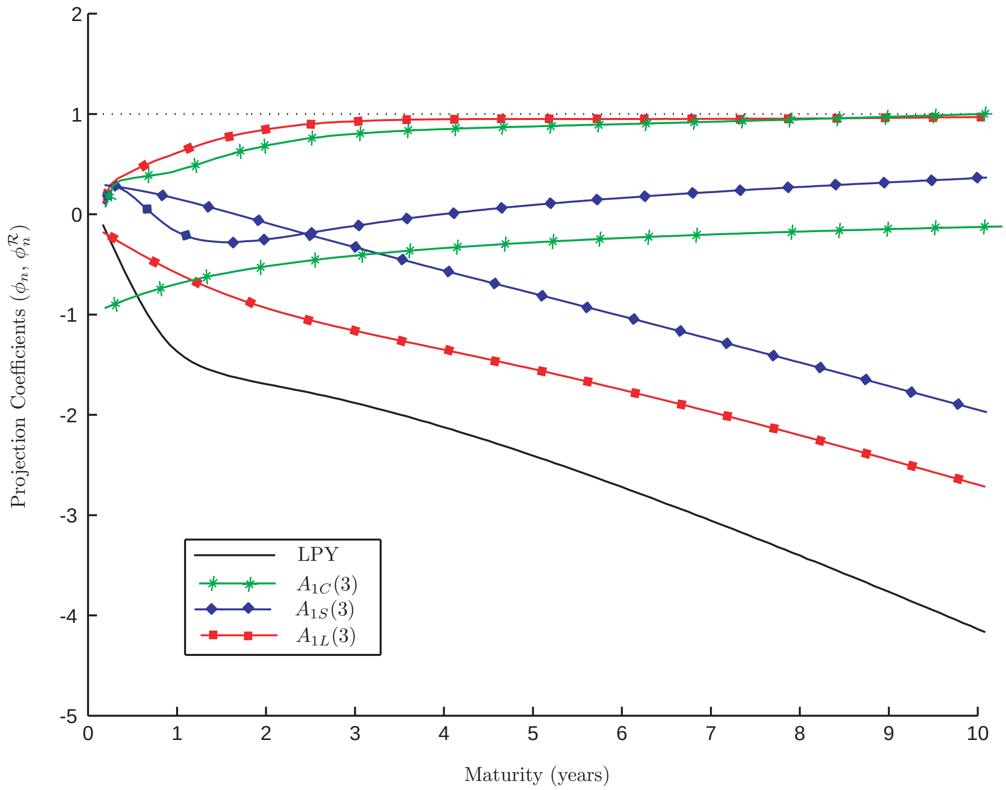

Figure 6: Model implied projection coefficients from constrained versions of model $A_{1C}(3)$. For each model the upper (lower) graph is a plot of $\phi_n^{\mathcal{R}}$ (ϕ_n).

4.4 Risk Premia Versus Factor Correlations in Matching *MPY*

Though we have focused on the role of risk premiums in matching *MPY*, 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 κ and nonzero λ^Y in matching *MPY* in Gaussian models.

Figure 8 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 κ), while the latter allows maximal flexibility in the correlation structure. Clearly neither model matches *MPY*(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 κ). We see a notable improvement in matching *MPY* 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, κ, and state-dependent market prices of risk, λ^Y, that allow model $A_{0C}(3)$ to match *MPY*.

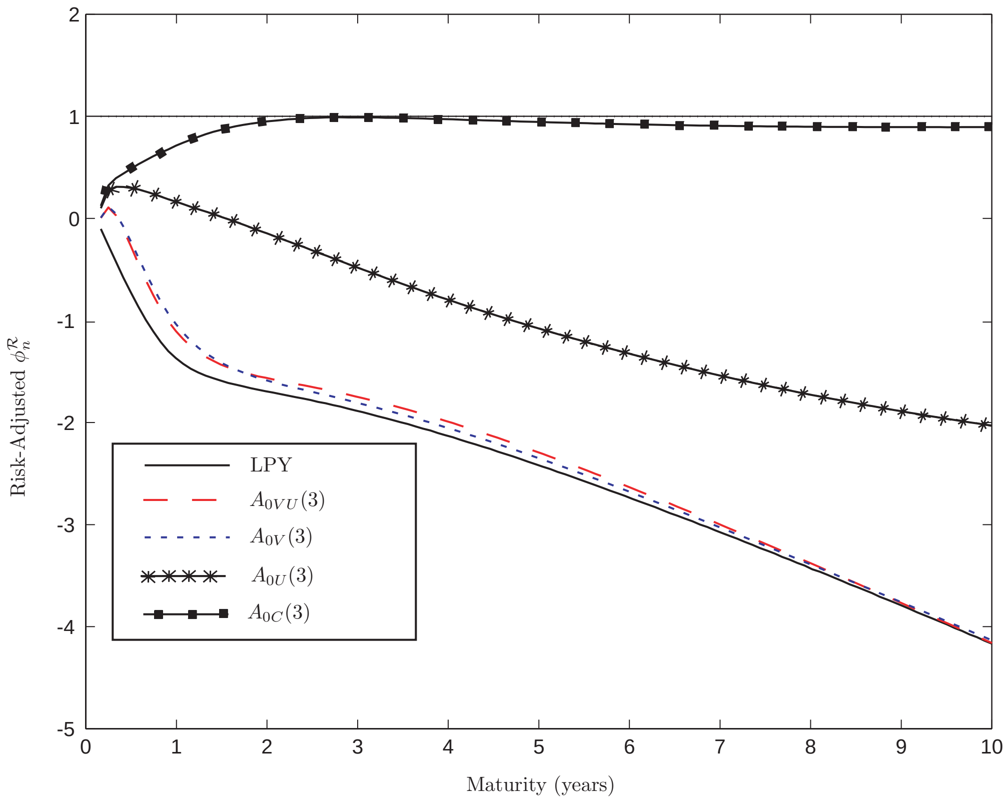

Figure 7: Estimated $\phi_n^{\mathcal{R}}$ implied by *ML* Estimates of Constrained $A_0(3)$ Models.

4.5 The Short End of the Yield Curve

We have deferred examination of the failure of affine models to match *MPY*(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 *MPY* 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 *MPY*(ii) at the short end is now nearly perfect as can been seen from Figure 8.[31]

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

$$R_{t+1}^{n-1} - R_t^n - (c_{t+1}^{n-1} - c_t^{n-1}) + \frac{1}{n-1} p_t^n = \frac{1}{n-1}(R_t^n - r_t) - \sum_{i=1}^{n-1}(E_{t+1}r_{t+i} - E_t r_{t+i}) \quad (33)$$

The estimated $\phi_n^{\mathcal{R}}$ displayed in Figures 5 and 8 are obtained by projecting the left-hand-

[31] 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.

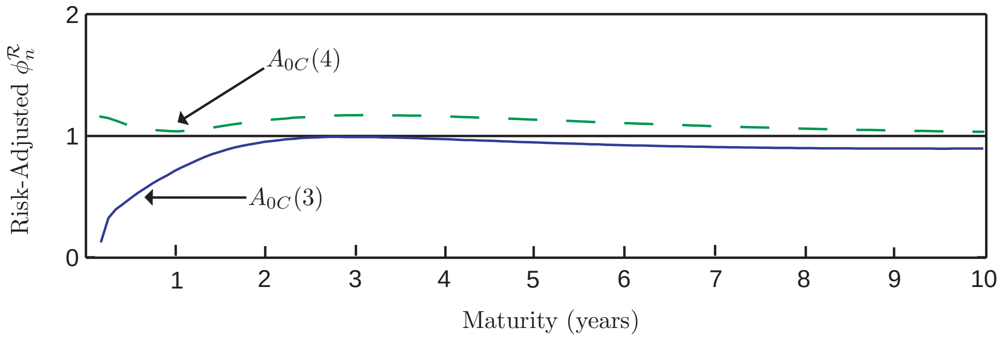

Figure 8: Estimated $\phi_n^{\mathcal{R}}$ implied by *ML* Estimates of Models $A_{0C}(3)$ and $A_{0C}(4)$.

side of (33) 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 (33) onto $(R_t^n - r_t)/(n-1)$ gives zero, so we get $\phi_n^{\mathcal{R}} = 1$. 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 $\phi_n^{\mathcal{R}}$ 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.

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 *DTSM*s to match *MPY* – thereby explaining *LPY*. For several of the popular families of one-factor *DTSM*s, we showed that this is indeed the case as these models were calibrated to match *MPY* quite closely. We then took up the more demanding challenge of formulating models that match *MPY* and at the same time match other features of the conditional distributions of bond yields as summarized by the scores of the model-implied log-likelihood functions. Focusing on the case of affine models, we showed that Gaussian models with correlated factors and state-dependent risk premiums fully resolved the Campbell-Shiller expectations puzzles *at the maximum likelihood estimates for this model.*

In the process, several observations about the empirical fits of affine and quadratic-Gaussian models emerged. First, since quadratic-Gaussian and "extended" Gaussian affine models have the same structure of their risk premiums, they are both capable in principle of matching *MPY*. In our calibration exercise, we found that one-factor versions of these models perform equally well in this regard. Second, seemingly central to the goodness-of-fit of the multi-factor Gaussian models to *MPY* was the non-zero correlation among the factors: allowing for state-dependent risk premiums, but maintaining zero factor correlations fell notably short of matching *MPY*.

Third, we identified a clear trade-off within affine models between matching the conditional first-moment properties of yields, as summarized by *MPY*, and matching the conditional volatilities of yields. Gaussian models resolve this trade-off (trivially) by positing constant volatilities. 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 *MPY*. Our findings that quadratic-Gaussian models are also capable of matching *MPY*, together with the recent findings by Ahn, Dittmar, and Gallant [2000] that quadratic-Gaussian models seem to fit the volatility properties of yields better than affine models, suggest that quadratic-Gaussian models may be more successful than affine models at matching both features of return distributions.

More generally, we conjecture that there is a much larger class of *DTSM*s with sufficient flexibility to match *MPY*. For instance, 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 *MPY*.[32] Another family of *DTSM*s that might match *MPY* are the models proposed by Duarte [1999] in which the state vector follows the affine diffusion (11) 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 (14). 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 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 $0 \le \lambda \le 2$ to rule out bubble solutions, and ζ_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 $0 \le \lambda \le 2$ translates into the condition $\tilde{\kappa} \ge \kappa > 0$. In other words, the constraints on λ 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

[32] 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.

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 explain LPY. 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 *DTSM*s 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 $Y(t)$:

$$r_0(t) = a_0 + b_0' Y(t), \tag{34}$$

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

$$dY(t) = \kappa(\theta - Y(t))dt + \sigma dW(t), \tag{35}$$

where κ and σ are $N \times N$ matrices and θ is a $N \times 1$ vector.

The market price of risk[33] is given by

$$\Lambda(t) = \sigma^{-1}(\lambda^0 + \lambda^Y Y(t)), \tag{36}$$

where λ^0 is a $N \times 1$ vector and λ^Y is a $N \times N$ matrix of constants. If the Girsanov's theorem applies,[34] the risk neutral dynamics of the state vector is given by

$$dY(t) = \tilde{\kappa}(\tilde{\theta} - Y(t))dt + \sigma d\tilde{W}(t), \tag{37}$$

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

We assume that κ can be decomposed as $\kappa = X^{-1}\kappa_d X$, where κ_d is a diagonal matrix with strictly positive diagonal elements κ_i, $1 \le i \le N$, X is a non-singular real matrix, with diagonal elements normalized to 1.[35] 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 $\tilde{\kappa}_i$, $1 \le i \le N$, 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

$$E\left[Y(t+\tau)|Y(t)\right] = e^{-\kappa\tau}Y(t) + (I - e^{-\kappa\tau})\theta. \tag{38}$$

The conditional variance is given by

$$\mathrm{Var}(Y(t+\tau)|Y(t)) = X^{-1}\Omega(\tau)X'^{-1}, \tag{39}$$

[33] The pricing kernel is given by

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

[34] See Appendix C for a proof that this is indeed the case.

[35] Alternatively, one could normalize the Euclidean length of each column vector of X to 1. If σ is completely free, then we can choose to normalize κ to be diagonal. In which case, we set $X \equiv I$.

where

$$\Omega_{ij}(\tau) = \Sigma_{ij} \frac{1 - e^{-(\kappa_i + \kappa_j)\tau}}{\kappa_i + \kappa_j}, \tag{40}$$

$$\Sigma = X\sigma\sigma' X'. \tag{41}$$

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

$$P(t,\tau) = e^{-A(\tau) - B(\tau)'Y(t)}, \tag{42}$$
$$R(t,\tau) = a(\tau) + b(\tau)'Y(t), \tag{43}$$

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

$$b(\tau) = (I - e^{-\tilde{\kappa}'\tau})(\tilde{\kappa}'\tau)^{-1} b_0, \tag{44}$$

$$a(\tau) = a_0 + (b_0 - b(\tau))'\tilde{\theta} - \frac{1}{2}\mathrm{Tr}\left[\Xi(\tau)\tilde{X}'^{-1}\tilde{\kappa}'^{-1} b_0 b_0' \tilde{\kappa}^{-1}\tilde{X}^{-1}\right], \tag{45}$$

$$\Xi_{ij}(\tau) = \tilde{\Sigma}_{ij}\left[1 - \frac{1 - e^{-\tilde{\kappa}_i\tau}}{\tilde{\kappa}_i\tau} - \frac{1 - e^{-\tilde{\kappa}_j\tau}}{\tilde{\kappa}_j\tau} + \frac{1 - e^{-(\tilde{\kappa}_i + \tilde{\kappa}_j)\tau}}{(\tilde{\kappa}_i + \tilde{\kappa}_j)\tau}\right], \tag{46}$$

$$\tilde{\Sigma} = \tilde{X}\sigma\sigma'\tilde{X}'. \tag{47}$$

A.1 Risk Premiums

Let us fix Δ 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

$$E_t[r_{t+n}] = \mu_n + \nu_n' Y(t), \text{ where} \tag{48}$$
$$\mu_n = a_1 + \theta'(I - e^{-\kappa' n})b_1 \tag{49}$$
$$\nu_n = e^{-\kappa' n} b_1 \tag{50}$$

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

$$f_t^n \equiv -\frac{1}{\Delta}\ln\frac{P(t,(n+1)\Delta)}{P(t,n\Delta)} = A_n^\Delta + B_n^{\Delta\prime} Y(t), \tag{51}$$

where

$$A_n^\Delta \equiv \frac{A_{n+1} - A_n}{\Delta} \text{ and } B_n^\Delta \equiv \frac{B_{n+1} - B_n}{\Delta}. \tag{52}$$

Thus, the forward risk premium is given by

$$p_t^n \equiv f_t^n - E_t[r_{t+n}] = (A_n^\Delta - \mu_n) + (B_n^\Delta - \nu_n)' Y(t), \tag{53}$$

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.

A.2 One-factor Case

The formulas for the factor loadings in the one-factor Gaussian model are

$$A(\tau) = a_0\tau + (\tau - B(\tau))(\tilde{\theta} - \frac{\sigma^2}{2\tilde{\kappa}^2}) + \frac{\sigma^2}{4\tilde{\kappa}} B(\tau)^2, \tag{54}$$

$$B(\tau) = \frac{1 - e^{-\tilde{\kappa}\tau}}{\tilde{\kappa}} b_0. \tag{55}$$

The forward-spot spread is given by

$$f_t^n - r_t = (A_n^\Delta - a_1) + (B_n^\Delta - b_1) Y(t). \tag{56}$$

Substituting (56) into (53), we have

$$f_t^n - E_t[r_{t+n}] = \delta_n + \alpha_n (f_t^n - r_t), \tag{57}$$

where

$$\delta_n = \frac{A_n^\Delta - \mu_n}{B_n^\Delta - b_1}, \tag{58}$$

$$\alpha_n = \frac{B_n^\Delta - e^{-\kappa' n\Delta} b_1}{B_n^\Delta - b_1} = \frac{e^{-\tilde{\kappa} n} - e^{-\kappa n}}{e^{-\tilde{\kappa} n} - 1} \tag{59}$$

Since $E[p_t^n] = E[f_t^n - r_t]$, δ_n can be related to the sample mean of the forward spread: $\delta_n = (1 - \alpha_n) E(f_t^n - r_t)$.

B Quadratic-Gaussian Model

Starting with the specification of the one-factor quadratic-Gaussian model in Section 3.1.2, we let $A_n \equiv A(n\Delta)$, $B_n = B(n\Delta)$, $C_n = C(n\Delta)$, $a_1 = \frac{A_1}{\Delta}$, $b_1 = \frac{B_1}{\Delta}$, $c_1 = \frac{C_1}{\Delta}$, $A_n^\Delta \equiv \frac{A_{n+1} - A_n}{\Delta}$, $B_n^\Delta \equiv \frac{B_{n+1} - B_n}{\Delta}$, and $C_n^\Delta \equiv \frac{C_{n+1} - C_n}{\Delta}$. Then the one-period short rate is given by

$$r_t = a_1 + b_1 Y_t + c_1 Y_t^2, \tag{60}$$

and the one-period forward rate, delivered n periods from t, is given by

$$f_t^n \equiv -\frac{1}{\Delta}\log\frac{P_t^{n+1}}{P_t^n} = A_n^\Delta + B_n^\Delta Y_t + C_n^\Delta Y_t^2. \tag{61}$$

The expected short rate is given by

$$E_t[r_{t+n}] \;=\; \mu_n + \nu_n Y_t + \omega_n Y_t^2, \tag{62}$$

where

$$\begin{aligned}
\mu_n &= a_1 + b_1\theta(1-e^{-\kappa n\Delta}) + c_1\theta^2(1-e^{-\kappa n\Delta})^2 + c_1\mathrm{Var}_t(Y_{t+n}) \\
\nu_n &= b_1 e^{-\kappa n\Delta} + 2c_1\theta(1-e^{-\kappa n\Delta})e^{-\kappa n\Delta} \\
\omega_n &= c_1 e^{-2\kappa n\Delta}.
\end{aligned}$$

From above, we can deduce the functional forms of constant coefficients in a two-factor forward risk premium model generated by the Quadratic-Gaussian model:

$$f_t^n - E_t[r_{t+n}] = \delta_n + \alpha_n(f_t^n - r_t) + \beta_n r_t, \tag{63}$$

where

$$\begin{pmatrix} B_n^\Delta - b_1 & b_1 \\ C_n^\Delta/c_1 - 1 & 1 \end{pmatrix}\begin{pmatrix} \alpha_n \\ \beta_n \end{pmatrix} \;=\; \begin{pmatrix} B_n^\Delta - \nu_n \\ C_n^\Delta/c_1 - \omega_n/c_1 \end{pmatrix}, \tag{64}$$

or

$$\alpha_n \;=\; 1 - \frac{\nu_n/b_1 - \omega_n/c_1}{B_n^\Delta/b_1 - C_n^\Delta/c_1} \tag{65}$$

$$\beta_n \;=\; (B_n^\Delta/b_1 - \nu_n/b_1) - (B_n^\Delta/b_1 - 1)\alpha_n \tag{66}$$

and

$$\delta_n \;=\; (1-\alpha_n)E[f_t^n - r_t] - \beta_n E_t[r_t]. \tag{67}$$

Note that due to the existence of invariant transformations, we can normalize $\theta = 0$, $b = 1$. Now, the parameters σ, c, and λ^0 appear only in the combinations $c\sigma^2$ and $c\lambda^0$ in our moment conditions. So one of the three parameters is not independently identified and must be normalized to 1. Consistent estimators of the "true" parameter values can be inferred once one of the parameters is identified through other means.[36]

[36]For an example, suppose that, under the normalization $\sigma = 1$, the estimators for c and λ^0 are c_T and λ_{0T}, respectively. If we subsequently have a consistent estimator of σ, σ_T, then the consistent estimators for c and λ^0 would be c_T/σ_T^2 and $\lambda_{0T}\sigma_T^2$, respectively. For our purpose, however, only c_T and λ_T matter, although they should not be interpreted as consistent estimators of the population coefficients for the underlying DGP.

C Conditions for Girsanov's Theorem

The goal is to show that

$$Z(t) = e^{\int_0^t \Lambda_s' dW_s - \frac{1}{2}\int_0^t \Lambda_s' \Lambda_s ds}, \tag{68}$$

is a Martingale, when Λ_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, Λ_t is a progressively measurable function of the Brownian motion, and for arbitrary $T > 0$, there exists a $K_T > 0$, such that

$$|\Lambda_t| \le K_T(1 + W^*(t)), \ 0 \le t \le T, \tag{69}$$

where $W^*(t) = \max_{0 \le s \le t} |W(s)|$, 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

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

Since Λ_t is an affine function of $Y(t)$, it is obvious that (69) holds.

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). Predictable changes in yields and forward rates. Working Paper, New York University.

Bansal, R. and H. Zhou (2000). Term structure of interest rates with regime shifts. Working Paper, Duke University.

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

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). Specification analysis of affine term structure models. *Journal of Finance 50*(5), 1943–1978.

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). 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.

Duffie, D. and K. Singleton (1997). An econometric model of the term structure of interest rate swap yields. *Journal of Finance 52*, 1287–1321.

Evans, B. D. (2000). 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). A simple model of the failure of the expectations hypothesis. Working Paper, Federal Reserve Board.

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*, 1029–1286.

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.

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.

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.

Pearson, N. D. and T. Sun (1994). Exploiting the conditional density in estimating the term structure: An application to the Cox, Ingersoll, and Ross model. *Journal of Finance 49*, 1279–1304.

Roberds, W. and C. Whiteman (1999). Endogenous term premia and anamolies 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.