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

Qiang Dai and Kenneth J. Singleton*

September 18, 2000

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. Key to this matching are parameterizations of the market prices of risk that let us separately "control" the shape of the mean yield curve and the correlation structure of excess returns with the slope of the yield curve. 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.

*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 (1984a, 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'"[1] 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 theory (*EH*) – there is compelling evidence from yield[2] and forward-rate[3] 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 its corresponding yield, r_t the one-period short rate, and f_t^n the forward rate for a one-period loan commencing at date $t+n$,[4] our discussion is organized around the following empirical observations:

LPY (i) Linear projections of $R_{t+1}^{n-1} - R_t^n$ onto $\frac{1}{n-1}(R_t^n - r_t)$ give negative, often statistically significant slope coefficients d_n^y. (ii) Moreover, the d_n^y typically become increasingly negative with maturity.

LPF Linear projections of $f_{t+1}^{n-1} - r_t$ onto $f_t^{(n)} - r_t$ give slope coefficients d_n^f that are significantly less than one, particularly for short maturities.

UMY On average, the term structure of treasury bond yields is upward sloping.

We refer to these patterns collectively as *LPEH*.

Though *LPEH* has often been viewed as a "puzzle" by term structure modelers,[5] we show that the patterns are in fact generated by two important classes of models: (1) a large subclass (though not all) of affine *DTSM*s,[6] and (2) the family of quadratic-Gaussian term structure models.[7] Key to our success at matching *LPEH* is having market prices of risk that affect both the (risk-neutral) long-run mean and the rate of mean-reversion of the state variables. Heuristically, the effect of risk premiums on the long-run mean of the state gives us the flexibility to match *UMY*, while their effect on the rate of mean reversion allows us to match the correlations between holding period returns and the slope of the yield curve (*LPY-LPF*). To substantiate this claim, we show empirically that, for plausible parameter values, *LPEH* is well matched by several one- and two-factor *DTSM*s.

[1] See Fama (1984b), page 545.

[2] See, e.g., Campbell and Shiller (1991).

[3] See, e.g., Backus et al. (1997).

[4] $R_t^n \equiv -\ln P_t^n/n$, $r_t \equiv R_t^1$, and $f_t^n \equiv -\ln(P_t^{n+1}/P_t^n)$.

[5] As demonstrated by Bekaert et al. (1997a) and Backus et al. (1997), the patterns *LPY* and *LPF* 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.

[6] See Duffie and Kan (1996) and Dai and Singleton (2000) for a discussion of affine *DTSM*s.

[7] See, e.g., Beaglehole and Tenney (1991) and Ahn et al. (2000).

At the heart of the *EH* puzzles are the findings, repeated in Table 1,[8] that coefficients in the linear projection of $R^{n-1}_{t+1} - R^n_t$ onto $\frac{1}{n-1}(R^n_t - r_t)$ are statistically significantly *negative* – observation LPY(i). These findings were 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.[9] The challenge we set for *DTSM*s is not simply to have a model imply the negative coefficients d^y_n in *EH* yield projections – essentially any *DTSM* that implies a positive correlation between excess returns and the slope of the yield curve will give this result.[10] Instead, we seek to jointly match the sign (observation LPY (i)) and maturity structure (observation LPY(ii)) of the d^y_n, *and* observation *UMY*.

Table 1: Campbell-Shiller Long Rate Regression

Estimated slope coefficients d_n from the indicated linear projections using the smoothed Fama-Bliss data set. The maturities n are given in months. See also Backus et al. (1997), column 1 of Tables 1 and 6.

$R^{(n-1)}_{t+1} - R^n_t = \text{constant} + d^y_n\,(R^n_t - r_t)/(n-1) + \text{residual}$										
Maturity	3	6	9	12	24	36	48	60	84	120
d^y_n	-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)

$f^{(n-1)}_{t+1} - r_t = \text{constant} + d^f_n \left(f^{(n)}_t - r_t\right) + \text{residual}$										
Maturity	3	6	9	12	24	36	48	60	84	120
d^f_n	0.731	0.797	0.851	0.891	0.946	0.958	0.962	0.964	0.964	0.963
s.e.	(.091)	(.057)	(.046)	(.039)	(.024)	(.017)	(.014)	(.012)	(.011)	(.010)

"Affine" *DTSM*s are a natural place to begin our exploration of *LPEH* both because of their historical prominence and the fact that optimal forecasts of excess returns take the form of the linear projections extensively studied in the literature on *EH*. Standard formulations of affine *DTSM*s (e.g., as parameterized in Dai and Singleton (2000)) have the market prices of risk proportional to factor volatilities. As such, the flexibility of a particular affine model to match *LPEH* will depend in part on how many of the state variables have time-varying volatilities.

Within CIR-style models, non-zero risk premiums affect the rate of mean reversion, but

[8] We are grateful to Backus et al. (1997) for providing the smoothed Fama-Bliss data used in our analysis. The data are monthly from February, 1970 through December, 1995.

[9] 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 $R^n_t - r_t$.

[10] If the risk premium were constant, then an expected rise in the short-term rate would lead to an expected capital loss on long-term bonds. Consequently, the slope of the yield curve would increase. This is the intuition used to justify the expectations hypothesis. If the risk premium is time-varying, and in particular is *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 price of long-term bonds will fall. Second, a falling risk premium tends to *increase* the values of long-term bonds. The expectations puzzle 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.

not the long-run mean, of the state process.[11] Consequently, they do not meet our conditions for matching *LPEH*. In fact, Figures 4 and 6 in Roberds and Whiteman (1999) show that (one- and two-factor) *CIR*-style models are wholly incapable of matching *LPY* for their sample period and treasury yields, even when the parameters of their *DTSM*s are calibrated to match their counterparts of the d_n^y. Moreover, Backus et al. (1997) demonstrate analytically that, in order for a (one-factor) *CIR*-style model to potentially match *LPY*, it must imply a *downward* sloping term structure of mean forward spreads $\{E[f_t^n - r_t]\}$, contrary to *UMY*.

At the opposite end of the spectrum of conditional volatility is the case of Gaussian models. With their constant volatilities and market prices of risk, the *EH* null hypotheses $d_n^y = 1$ is *true*! Thus, if we are to be successful at matching *LPY* within a Gaussian model we must step outside the Dai-Singleton family of affine models. Accordingly, we focus on an "extended" Gaussian model in which the market prices of risk are affine functions of the state. 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, $f_t^n - r_t$, which is reminiscent of the projections in Fama (1984a) and Fama and Bliss (1987) of excess returns onto $f_t^n - r_t$. In fact, we show that it both generates projection equations consistent with their empirical findings and resolves *LPEH*. Fisher (1998) independently proposed a similar potential resolution of *LPY* within a two-factor Gaussian model.[12]

Lying between these two cases – CIR-style and Gaussian models – are what we refer to as the $A_m(N)$ families of N-factor models,[13] where m is the number of state variables driving the volatilities of all N state variables and $0 < m < N$. The m volatility factors have market prices of risk that affect the rates of mean reversion of the states as in CIR-style models, while the risk premiums of the remaining $N - m$ "non-volatility" factors affect the long-run means of the states as in standard Gaussian models. Thus, these models meet our heuristic criteria for matching *LPEH*. Additional flexibility is obtained by extending a standard $A_m(N)$ model to allow the $N - m$ non-volatility factors to have state-dependent risk premiums as in the "extended" Gaussian case.

These extended affine models are special cases of the "essentially" affine family of models proposed by Duffee (1999).[14] 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. We provide the complementary, formal assessment of whether affine models also match *LPEH* and, in particular, the maturity patterns displayed in Table 1. Of equal interest, we assess the relative importance, within the affine family, of extending the risk premium specification, allowing for time-varying volatility, and allowing for non-zero factor correlations in matching *LPEH*.

We also show that quadratic-Gaussian models are inherently capable of (qualitatively) matching *LPEH*, because their market prices of risk are richer than those in standard affine

[11] See, for example, Chen and Scott (1993), Pearson and Sun (1994), and Duffie and Singleton (1997).

[12] We are grateful to Greg Duffee for bringing this unpublished manuscript to our attention. While Fisher shows that his model qualitatively resolves *LPEH*, he does not compare the model-implied and historical projection coefficients d_n^y, as is done subsequently here in Section 5, to assess whether extended Gaussian models quantitatively resolve *LPEH*.

[13] See Dai and Singleton (2000)for a canonical representation of these families of affine *DTSM*s.

[14] Duffee's "essentially" affine models are "affine" as this term was originally used by Duffie and Kan (1996), so we will use the shorter "affine" term when referring to the models he studied.

models.[15] Indeed the basic structure of the market prices of risk in quadratic- and extended-Gaussian *DTSM*s is the same.

Backus et al. (1997) also present empirical evidence against the related *EH* null hypothesis of $d_n^f = 1$, particularly at the shorter maturities– observation *LPF*; see the lower half of Table 1. They argue that this pattern can be matched by a "negative CIR" process. Our resolution of *LPEH* (and hence *LPF*) 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 *LPF*. In addition, we provide a link to, and reinterpretation of, the modeling implications of the forward-rate regressions in Fama and Bliss (1987).

To complement these qualitative assessments, we explore the quantitative fit of affine *DTSM*s to *LPEH* in two ways. First, for the case of one-factor Gaussian and Quadratic-Gaussian models, we illustrate the central points of this paper by showing that the parameters can be calibrated so that the model-implied projection coefficients match, remarkably closely, the historical coefficients in Table 1, while at the same time matching *UMY*. Additionally, we find empirically that (for the purpose of matching *LPEH*) the one-factor quadratic- and extended-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 *LPEH* the same as showing that the same parameters match *LPEH* and other aspects of the distributions of yields, say those summarized by the likelihood function of the data. Both of these concerns are addressed in an extensive exploration of *LPEH* within the families of two-factor extended Gaussian ($A_0(2)$) and mixed Gaussian-square-root ($A_1(2)$) models. We fit a variety of models in these families by the method of full-information maximum likelihood (*ML*) and then compare the model-implied and historical versions of d_n^y. In this manner we are able to assess more formally whether *ML* estimates of two-factor affine models (extended and standard) match *LPEH*.

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 *LPEH*. Additionally, we provide two "structural" interpretations of our parameterizations of risk premiums, one based on the representative agent, stochastic habit formation model in Dai (2000) and the other on the monetary-based explanation of *LPY* in McCallum (1994). Section 4 shows empirically that our illustrative one-factor Gaussian *DTSM*s match *LPEH* at admissible parameter values. Section 4.3 presents a similar calibration exercise for the one-factor quadratic-Gaussian model. A more formal and extensive empirical assessment of the fit of two-factor affine *DTSM*s to *LPEH* is presented in Section 5. Concluding remarks are presented in Section 6. Technical details are collected in an appendix.

[15] See Ahn et al. (2000) and Section 4.3.

2 Risk-Premium Adjusted Projections

If the failure of the *EH* 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 *LPEH* 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 subsequent empirical analysis.

2.1 Yield Projections

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

$$e^n_t = -(n-1)E_t\left[R^{n-1}_{t+1} - R^n_t\right] + (R^n_t - r_t), \tag{1}$$

where E_t denotes expectation conditioned on date t information. Rearranging (1) gives the fundamental relation[16]

$$E_t\left[R^{n-1}_{t+1} - R^n_t + \frac{1}{n-1}D^n_{t+1}\right] = \frac{1}{n-1}(R^n_t - r_t). \tag{2}$$

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

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

and the forward term premium

$$p^n_t \equiv f^n_t - E_t[r_{t+n}]. \tag{4}$$

Since $R^n_t \equiv \frac{1}{n}\sum_{i=0}^{n-1} f^i_t$, the term premiums p^n_t and c^n_t are linked by the simple relation:

$$c^n_t \equiv \frac{1}{n}\sum_{i=0}^{n-1} p^i_t. \tag{5}$$

[16]Expression (2) is formally equivalent to equation (11) of Fama and Bliss (1987), which, in our notation, is:

$$E_t\left[R^{n-1}_{t+1} - R^{n-1}_t + \frac{1}{n-1}D^n_{t+1}\right] = \frac{1}{n-1}(f^{n-1}_t - r_t).$$

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^n_{t+1} can be decomposed into a pure "premium" part (D^{*n}_{t+1}) and an "expectations" part:[17]

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

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

Since the $(E_t r_{t+i} - E_{t+1} r_{t+i})$ have zero date-t conditional means, e^n_t depends only on the premium term D^{*n}_{t+1}:[18]

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

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

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

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

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

onto the (scaled) slope of the yield curve, $(R^n_t - r_t)/(n-1)$, has a coefficient of one.[19] The *EH* hypothesis is obtained by setting the risk premiums in (10) to constants.

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

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

[18] 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 *EH*. 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.

[19] 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 *LPEH*

The challenges set forth by Fama and the studies of *LPY* in the *EH* literature are statements about correlations among yields and, as such, are naturally studied using linear projections. Therefore, in attempting to generate *LPEH* within *DTSM*s we start our inquiry with models in which conditional expectations are linear in known functions of the state vector, a trait 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 the "physical" or "actual" measure, the affine diffusion

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

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 $\sqrt{S(t)}$ (see Dai and Singleton (2000) and the references therein):

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

where ℓ_0 is an $N \times 1$ vector of constants.

Focusing narrowly on the objective of matching *LPEH*, we know from previous studies that this specification imposes potentially severe limitations on a model's ability to match first-moment properties of the yield data. In *CIR*-style models, the α_i are zero and the β_i are normalized to vectors with unity in the i^{th} component and zero elsewhere, so $\Lambda_i(t) = \ell_{0i} \sqrt{Y_i(t)}$. Consistent with the observations in Backus et al. (1997), our interpretation of the difficulty in previous studies of matching *LPEH* with *CIR*-style models is that these risk premiums affect the rate of mean reversion, but not the long-run mean of Y under the risk neutral measure. So matching *LPY* and *UMY* simultaneously is challenging, if not infeasible. Models in $A_N(N)$ are not easily "fixed up" with the particular modifications of the market prices of risk proposed here without introducing arbitrage opportunities (Cox et al. (1985)).

If the state Y is Gaussian, then $S(t)$ is the identity matrix, $\Lambda(t)$ is a vector of constants, and the *EH* null hypothesis is true. Accordingly, we extend the basic Gaussian *DTSM* by letting the market price of risk be defined by the relation[20]

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

where λ^0 is an $N \times 1$ vector and λ^Y is an $N \times N$ matrix of constants. The tension between *UMY* and *LPF* highlighted by Backus et al. (1997) is relaxed by this specification of $\Lambda(t)$

[20] Appendix C shows that the conditions for Girsanov's theorem to apply are satisfied by the measure change associated with the market price of risk (14) without imposing any restrictions on model parameters (except for the usual stationarity assumption).

(while preserving no arbitrage): λ^0 controls the shape of the mean yield curve (to make the mean yield curve upward sloping, make λ^0 as negative as needed) and λ^Y controls the time-varying behavior of the excess returns on bonds.

More flexibility is afforded by standard affine models within the sub-families $A_m(N)$ with $m < N$. To be concrete, let $N = 2$ and $m = 1$, so that $Y' = (Y_1, Y_2)$ with factor variances

$$[S(t)]_{11} = Y_1(t), \qquad [S(t)]_{22} = \alpha_2 + \beta_2 Y_1(t); \tag{15}$$

the volatilities of both state variables are driven by Y_1, which follows a square-root process. Substituting (15) into (13), we obtain

$$\Lambda(t)' = \left(\sqrt{Y_1(t)}\ell_{01}, \sqrt{\alpha_2 + \beta_2 Y_1(t)}\ell_{02}\right). \tag{16}$$

As long as $\alpha_2 \neq 0$, then the market prices of risk affect both the mean and persistence of Y under the risk-neutral measure. This same logic extends more generally to the case of models in $A_m(N)$ with $0 < m < N$ and $\alpha_i \neq 0$ for some $m < i \leq N$.

Additional flexibility in matching *LPEH* is obtained by, following Duffee (1999), extending the specification of $\Lambda(t)$ in $A_m(N)$ models to satisfy

$$\Sigma\sqrt{S(t)}\Lambda(t) = \begin{pmatrix} 0_{m\times 1} \\ \lambda_{02} \end{pmatrix} + \begin{pmatrix} \lambda_{11}^Y & 0 \\ \lambda_{21}^Y & \lambda_{22}^Y \end{pmatrix} Y(t), \tag{17}$$

where λ_{02} is an $(N-m)\times 1$ vector, λ_{11}^Y is a $m\times m$ diagonal matrix, λ_{21}^Y and λ_{22}^Y are $m\times(N-m)$ and $(N-m)\times(N-m)$ matrices (all constants), and it is presumed that $\inf(\alpha_i + \beta_i' Y_1(t)) > 0$ for $i = m+1, \ldots, N$. Both formulations (16) and (17) will be examined empirically for our illustrative two-factor $(N = 2)$ models in Section 5.

Another family of *DTSM*s with the potential to resolve *LPEH* 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 et al. (2000) show that the market price of risk in their canonical N-factor quadratic-Gaussian model takes exactly the same form as (14). Thus, qualitatively, this model inherently gives the requisite flexibility to resolve *LPEH*.

A natural question at this juncture is: What are the economic underpinnings of our parameterization of Λ_t in (14)? Following are two possible structural underpinnings of this affine parameterization within a one-factor Gaussian setting. First, it turns out that McCallum (1994)'s resolution of the *EH* puzzle based on the behavior of a monetary authority is substantively equivalent to our affine parameterization of Λ_t. 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 a 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 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 AR(1) 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 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.[21]

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 correlation between the stochastic interest rate and Sharpe ratio of the risky production technology is necessarily negative under very general economic assumptions. This negative correlation resolves the *LPEH* puzzles. The models with affine, state-dependent market price of risk studied here can be interpreted as zeroth-order approximations to the interest rate dynamics implied by Dai's model.

4 *LPEH* and One-Factor Models

To illustrate the potential of certain affine and quadratic-Gaussian models to resolve the *LPEH* puzzle, we proceed in this section to calibrate one-factor models and verify that *LPY* is in fact matched. The case of the one-factor Gaussian model is particularly revealing of the importance of our state-dependent formulation of the risk premium, since the *EH* restrictions are true under the standard Gaussian model with constant premium.

4.1 Econometric Strategy

Our strategy for assessing whether a specific one-factor model can match *LPEH* is to estimate the model parameters using the moment equations (9) and (21); compute the model-implied p_t^n and c_t^n, evaluated at the estimated parameters; and, finally, to examine whether the sample counterparts to the term structures of projection coefficients

$$\tilde{d}_n^y \equiv \frac{\text{cov}(R_{t+1}^{n-1} - R_t + D_{t+1}^{*n}/(n-1), (R_t^n - r_t)/(n-1))}{\text{var}((R_t^n - r_t)/(n-1))} \tag{18}$$

are statistically different from a horizontal line at 1. Finding a model for which the fitted $\tilde{d}_n^y$ do not differ significantly from one resolves the *LPY* puzzle, because it is the model-implied risk premiums that determine the D_{t+1}^{n*}.

More concretely, to explore the links between *DTSM*s and *LPY*, we focus on forward term premiums. This is equivalent (see footnote 18) 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). Moreover, from (2) and (8), we can write

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

[21] The special case of $\lambda = 2$ corresponds to the case of $\alpha_n \equiv 0$ in our model of risk premia, or to the case of $\lambda_Y = 0$ (constant market price of risk) in our dynamic model. In this special case, the monetary authority induces mean reversion in the short rate, but does not induce a differential in the speeds of mean reversion under the physical and the risk neutral measures. Consequently, the risk premia is constant.

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 models imply p_t^n that are special cases of

$$p_t^n = \delta_n + \alpha_n(f_t^n - r_t) + \beta_n r_t, \tag{20}$$

where the $(\delta_n, \alpha_n, \beta_n)$ 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 19)

$$E_t[f_{t+1}^{n-1} - r_t] - (f_t^n - r_t) + (E_t[p_{t+1}^{n-1}] - p_t^n) \equiv E_t[u_{t+1}] = 0. \tag{21}$$

For these models, the risk premium parameters α_n and β_n turn 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 (20) and the assumption of stationarity imply

$$\delta_n = (1 - \alpha_n)E[f_t^n - r_t] - \beta_n E[r_t]. \tag{22}$$

Thus, if the model is correctly specified, δ_n can be inferred from (α_n, β_n) and the sample means of the forward-spot spread and one-period short rate.

We estimate κ and λ^Y using the moment conditions

$$E[u_{t+1}^n z_t] = 0, \text{ with } z_t = (f_t^n - r_t, r_t)', \ n = 6, 12, 24, 60, 84, 120. \tag{23}$$

Only a subset of the maturities between 1 and 120 months are used because the smoothed Fama-Bliss dataset is interpolated. As a goodness-of-fit statistic we use the minimized value of the *GMM* objective function (Hansen (1982)).

This procedure essentially forces our one-factor models to match *UMY*, while ignoring the restrictions implicit in the dependence of δ_n on the primitive parameters of the model, say ς. Two motivations for starting with this approach are: (1) it highlights the roles of the model parameters λ^Y and κ in resolving *LPY*, and (2) we know *a priori* that all of the parameters in ς other than κ and λ^Y are available for matching *UMY* (as well as other features of the distributions of bond yields). These parameters include λ^0 and the long-run mean θ of Y. After showing that our extended Gaussian models match *LPY* with the δ_n concentrated out, we then show that λ^0 and θ (and hence the model-implied δ_n) can be chosen so that these models also match the mean yield curve (*UMY*) almost perfectly, except for the very shortest maturities.

4.2 Calibration of One-Factor Gaussian Models

In the one-factor Gaussian *DTSM* the instantaneous short rate is given by $r_{0t} = a_0 + b_0 Y_t$, and Y_t follows a one-dimensional Gaussian process (11) with $N = 1$, $S(t) = 1$, and $\Sigma = 1$.[22] The linearity of our one-factor Gaussian model implies that the yield and forward risk premiums

[22] The latter is a normalization, imposed without loss of generality (Dai and Singleton (2000)).

are both affine in Y; equivalently, they are affine in any yield or yield spread that is itself affine in Y. To facilitate interpretation and comparison with Fama's analysis, we represent p_t^n as in (20) with $\beta_n = 0$ and

$$\alpha_n = \frac{e^{-\kappa n\Delta} - e^{-\tilde{\kappa} n\Delta}}{1 - e^{-\tilde{\kappa} n\Delta}}, \quad (24)$$

$$\delta_n = (1-\alpha_n)(A_n^\Delta - a_1) + (1-\alpha_n)(B_n^\Delta - b_1)\theta, \quad (25)$$

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, $E[D_{t+1}^n | f_t^n - r_t] = E[D_{t+1}^{*n} | f_t^n - r_t]$ is linear in $f_t^n - r_t$. Of course Fama does not impose the dynamic restrictions (24) and (25), because (20) (with $\beta_n = 0$) is essentially the starting point of his empirical analysis.

The estimates $(\hat{\kappa}, \hat{\lambda}^Y)$ are $(0.0012, 0.0008)$ with asymptotic standard errors $(0.029, 0.019)$, respectively. Though both κ and λ^Y are statistically insignificant individually, the ratio $\frac{\lambda^Y}{\kappa}$ is significant: the point estimate of the ratio is 0.6577 with standard error 0.2967. This finding seems to be a consequence of the fact that, when both κ and λ^Y are very small, α_n is almost independent of n and is well approximated by $\frac{\lambda^Y/\kappa}{1+\lambda^Y/\kappa}$. Thus, in the region around the converged estimates of κ and λ^Y, the *GMM* objective function is essentially flat along the hyperplane defined by $\lambda^Y = 0.6577\kappa$. The *GMM* test statistic is $\chi^2(10) = 2.52$ so the overidentifying restrictions are not rejected.

Figure 1 displays the estimated slope coefficients $\tilde{d}_n^y$ implied by the one-factor model. For comparison, we have also plotted the estimated d_n^y from Table 1, obtained under the null hypothesis that p_t^n is constant for all n. For all but the shortest maturities, $\tilde{d}_n^y$ lie within one sample standard error of one.[23] We conclude that a one-factor Gaussian *DTSM* with affine dependence of the market prices of risk on the state variable Y potentially generates a risk premium specification that is consistent with the historical patterns *LPY*. Put differently, our (calibrated) adjustments for risk premiums move the *EH* projection coefficients d_n^y from values significantly below one to fitted values for $\tilde{d}_n^y$ that are largely insignificantly different from one.

It remains to explore the abilities of the model to match the average slope of the yield curve (*UMY*), conditional on the values of κ and λ^Y that match *LPY*. 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 *LPY*.

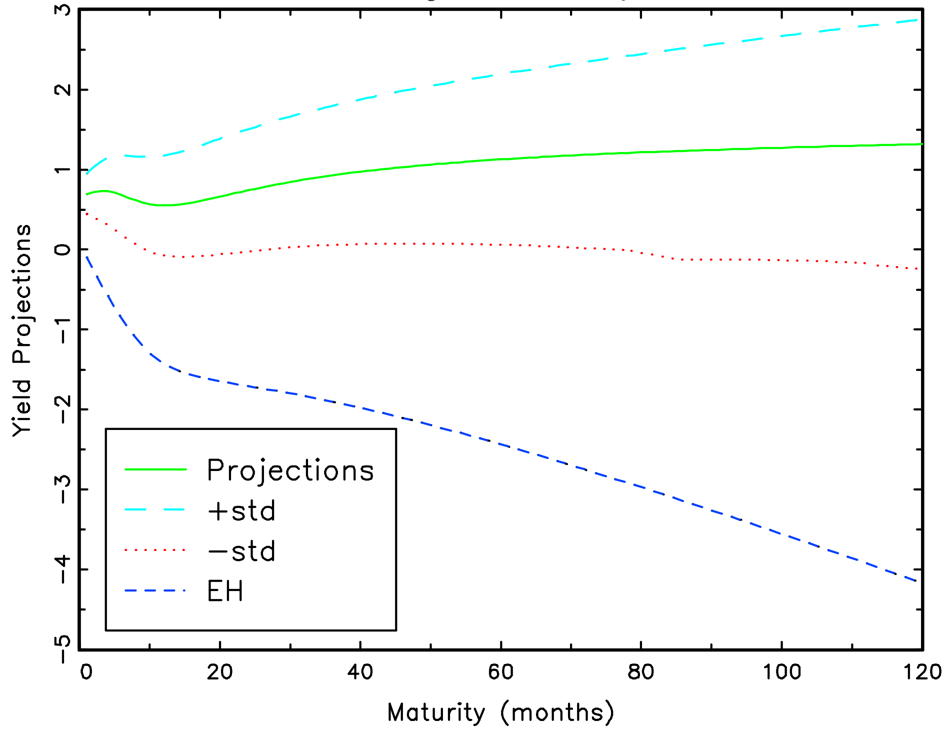
In the one-factor Gaussian model, the population mean of a zero coupon yield with maturity n months is given by (see Appendix A.2)

$$\bar{R}^n = a_n + b_n\theta, \quad (26)$$

[23] These standard error bands reflect the sampling variation of the parameter estimates, but not of the sample moments used in estimating $\tilde{d}_n^y$. Accounting for the latter would most likely widen these bands.

Figure 1: EH Projections Under One-Factor Gaussian Model

Plot of yield projection coefficients $\tilde{d}_n^y$ implied by the one-factor Gaussian model against maturity n.



where, with $\Delta = \frac{1}{12}$, $\tilde{\kappa} = \kappa + \lambda^Y$, $\tilde{\theta} = \tilde{\kappa}^{-1}(\kappa\theta - \lambda^0)$, and

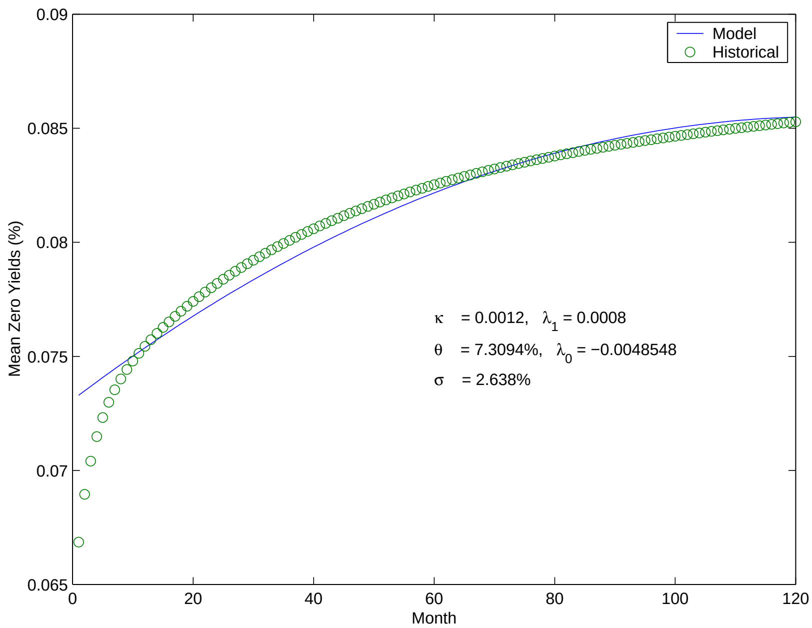
$$a_n = (1 - b_n)\left(\tilde{\theta} - \frac{\sigma^2}{2\tilde{\kappa}^2}\right) + \frac{\sigma^2}{4\tilde{\kappa}} b_n^2 n\Delta, \qquad b_n = \frac{1 - e^{-\tilde{\kappa}n\Delta}}{\tilde{\kappa}n\Delta}. \tag{27}$$

We proceed with the values of κ and λ^Y used to match LPY. We estimate σ from the sample volatility of the monthly changes of the one-month yield: $\sigma = \sqrt{12}std(r_t - r_{t+1})$, thereby imposing the discipline of matching the sample volatility of the short-term rate. We then calibrate λ^0 and θ jointly so that the $\bar{R}_t^n$ match their sample counterparts as closely as possible in the sense of mean-squared errors. This results in the parameter values $\kappa = 0.0012$, $\lambda^Y = 0.0008$, $\sigma = 2.638\%$, $\theta = 7.309\%$, and $\lambda^0 = -0.0049$.

Figure 2 plots the historical sample mean of the zero yield curve (circles) and its model-implied counterpart (solid line). We see that the mean curve generated from the model is almost on top of the sample mean curve. The most notable discrepancies are at the short end: the average 1-month yield is about 6.9%, whereas the model implies a mean of 7.3%.

Figure 2: Average Yield Curve

Model-implied and sample average yield curves for U.S. Treasury zero-coupon bonds.



4.3 Calibration of One-Factor Quadratic-Gaussian Models

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

$$-\log P(t,\tau) = A(\tau) + Y'B(\tau) + Y'C(\tau)Y \tag{28}$$

where

$$B(\tau) = \left(\frac{\tilde{\kappa}}{\Gamma}\frac{e^{\Gamma\tau}-1}{e^{\Gamma\tau}+1}+1\right)Q(\tau)\,b_0 + \left(\frac{2\tilde{\kappa}\tilde{\theta}}{\Gamma}\frac{e^{\Gamma\tau}-1}{e^{\Gamma\tau}+1}\right)Q(\tau)\,c_0 \tag{29}$$

$$C(\tau) = Q(\tau)\,c_0 \tag{30}$$

$$Q(\tau) = \frac{e^{2\Gamma\tau}-1}{(\Gamma+\tilde{\kappa})(e^{2\Gamma\tau}-1)+2\Gamma}, \tag{31}$$

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

$$E_t[r_{t+n}] = \mu_n + \nu_n Y_t + \omega_n Y_t^2, \tag{32}$$

with the coefficients expressed as functions of the primitive parameters in Appendix B. Letting $a_1 = A(\Delta)/\Delta$, $b_1 = B(\Delta)/\Delta$, $c_1 = C(\Delta)/\Delta$, and $H_n^\Delta \equiv [H((n+1)\Delta) - H(n\Delta)]/\Delta$ for any coefficient H, we show in Appendix B that the forward risk premiums in this model can be expressed as in (20) with coefficients

$$\alpha_n = 1 - \frac{\nu_n/b_1 - \omega_n/c_1}{B_n^\Delta/b_1 - C_n^\Delta/c_1} \tag{33}$$

$$\beta_n = (B_n^\Delta/b_1 - \nu_n/b_1) - (B_n^\Delta/b_1 - 1)\alpha_n. \tag{34}$$

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, the forward risk premiums in the two-factor, extended Gaussian and one-factor quadratic-Gaussian models have the same structure (see Section 5), but they are not identical because the *dynamic restrictions* imposed on the parameters α_n and β_n are different.

Figure 3: Linear vs Quadratic Gaussian Models

Model-implied estimates of forward projections coefficients $\tilde{d}_n^f$ from the one-factor Gaussian and quadratic-Gaussian *DTSM*s.

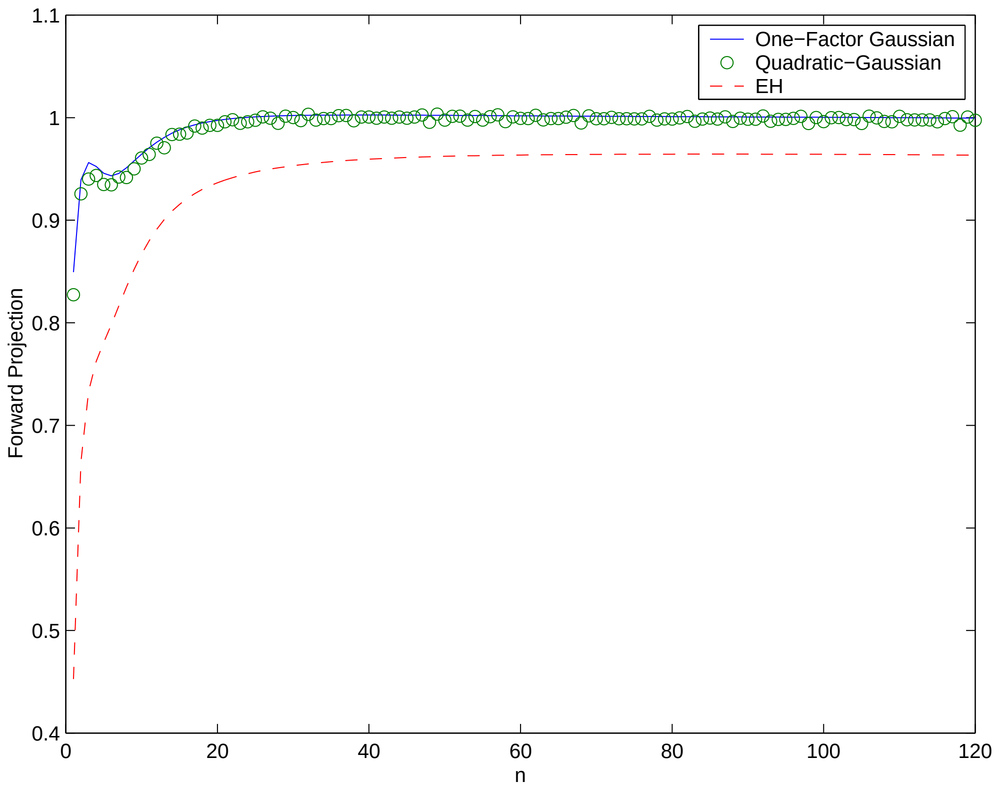


Figure 3 displays the forward-rate projection coefficients $\tilde{d}_n^f$ implied by the one-factor quadratic-Gaussian model estimated with the same moments used in estimating the one- and two-factor Gaussian models. We see that the quadratic-Gaussian model also does a good job of matching the pattern *LPF* at all but the shortest maturities. What is perhaps most striking about Figure 3 is that the estimated $\tilde{d}_n^f$ 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 *LPEH*, these two one-factor models perform equally well. Evidently, for the purpose of matching *LPF*, the dynamic restrictions implicit in (20) for the one-factor quadratic-Gaussian model are more restrictive than those in the two-factor Gaussian model.

The reason that the one-factor extended Gaussian and quadratic-Gaussian models give virtually the same results is that the estimated mean reversion coefficients and the quadratic constant c_0 are small.[24]

5 *LPEH* and Two-Factor Models

The preceding calibration of one-factor models, while demonstrating that *LPEH* can be matched by choice of admissible parameters in certain *DTSM*s, leaves open the question of whether we can *simultaneously* match the patterns of *LPEH* and other, higher-order moments of yield distributions. We turn next to a more demanding assessment of two-factor affine *DTSM*s by computing maximum likelihood (*ML*) estimates of models within the families $A_0(2)$ and $A_1(2)$ and examining whether the implied risk premiums, computed at the *ML* estimates, resolve the *LPEH* puzzles. Our shift to two-factor models is in recognition of the widely documented observation that more than one risk factor is necessary to describe yield curve dynamics. Extending our analysis to more than two factors would, of course, only increase the flexibility we would have in matching *LPEH* (and other features of yield distributions) at *ML* estimates of affine models.

All of the models examined have $r(t) = Y_1(t) + Y_2(t)$ and the state-vector Y following special cases of the bivariate diffusion:

$$dY_1(t) = \kappa_{11}(\theta_1 - Y_1(t))\,dt + \sigma_{11}\sqrt{\alpha_1 + \beta_1 Y_1(t)}dB_1(t), \tag{35}$$

$$dY_2(t) = [-\kappa_{21}Y_1(t) + \kappa_{22}(\theta_2 - Y_2(t))]\,dt + \sigma_{22}\,dB_2(t) \tag{36}$$

with $cov(dB_1(t), dB_2(t)) = 0$ and market prices of risk satisfying

$$\Sigma\sqrt{S(t)}\Lambda(t) = \begin{pmatrix} 0 \\ \lambda_2^0 \end{pmatrix} + \begin{pmatrix} \lambda_{11}^Y & 0 \\ \lambda_{21}^Y & \lambda_{22}^Y \end{pmatrix} Y(t), \tag{37}$$

with all of the non-zero elements of λ^0 and λ^Y being scalars. The specific models and the constraints they impose on (35) - (36) are given in Table 2.[25] The first two are two-factor Gaussian models, with the suffix "SG" denoting state-dependent risk premiums on

[24] 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.

[25] There are several dimensions along which these models are not "maximal" within their families ($A_0(2)$ and $A_1(2)$, respectively). In the case of the Gaussian models, we are free to relax the constraints $\lambda_{12}^Y = 0$ and $\lambda_1^0 = 0$. Upon freeing up these constraints, we found little change in the value of the log-likelihood function and virtually no improvement in matching *LPEH* within the family $A_0(2)$.

In the case of the $A_1(2)$ models, we could let the conditional variance of Y_2 have affine dependence on Y_1. However, this presents potential numerical identification problems and the likelihood function is already quite flat with the chosen set of free parameters (see Table 3), so we proceed with constant volatility σ_{22}.

the Gaussian factors and the final suffixes "U" and "C"denoting $\kappa_{21} \neq 0$ and $\kappa_{21} = 0$, respectively (uncorrelated and correlated risk factors due to feedback in the drift). The last four models have Y_1 following a square-root process and Y_2 following a Gaussian process. The suffix "CG" means that the second, Gaussian factor has a constant (state-independent) risk premium.

Table 2: Two-Factor Affine Models and Their Constraints

Model	Constraints
$A_0(2)$SG-U	$\alpha_1 = 1,\ \beta_1 = 0,\ \kappa_{21} = 0$
$A_0(2)$SG-C	$\alpha_1 = 1,\ \beta_1 = 0$
$A_1(2)$CG-U	$\alpha_1 = 0,\ \beta_1 = 1,\ \kappa_{21} = 0,\ \lambda^Y_{21} = 0,\ \lambda^Y_{22} = 0$
$A_1(2)$CG-C	$\alpha_1 = 0,\ \beta_1 = 1,\ \lambda^Y_{21} = 0,\ \lambda^Y_{22} = 0$
$A_1(2)$SG-U	$\alpha_1 = 0,\ \beta_1 = 1,\ \kappa_{21} = 0$
$A_1(2)$SG-C	$\alpha_1 = 0,\ \beta_1 = 1$

The *ML* estimates, their estimated standard errors, and the values of the log-likelihood functions are displayed in Table 3.[26] All of the models have one factor (the first) mean reverting faster than the other (second) factor, consistent with previous empirical studies of two-factor models. For the most flexible "SG-C" models with each of the families $A_0(2)$ and $A_1(2)$, both risk factors show substantial mean reversion under the physical measure. However, the rates of mean reversion of the second factor under the risk-neutral measure $(\kappa_{22} + \lambda^Y_{22})$ are relatively much slower (0.0025 and 0.0025, respectively).

Another notable feature of these estimates is that in all cases where κ_{21} is a free parameter, $\hat{\kappa}_{21}$ is estimated to be substantially negative implying that feedback through the drift matrix κ induces *negative* correlation among the two factors.[27] Importantly, within the family $A_N(N)$, and in particular $A_2(2)$, with all N factors driving volatilities (CIR-style models), it is theoretically impossible to accommodate this negative feedback (Dai and Singleton (2000)). Therefore, the fact that the data calls for negatively correlated factors rules out *a priori* consideration of this family of models. This limitation of the $A_2(2)$ family may partially explain why Roberds and Whiteman (1999) were unable to match *LPEH* with a two-factor CIR-style model.

Feedback in the drift is not the only source of factor correlations under the risk-neutral measures, however. With $\lambda^Y_{21} \neq 0$, the state-dependence of the risk premiums for the second Gaussian factor is a second source of factor correlation. Interestingly, in the $A_0(2)$SG-U

[26]Estimation of the $A_0(2)$ models is a standard likelihood problem. Full information, maximum likelihood estimates of the $A_1(2)$ models were obtained using the methods proposed by Pedersen and Singleton (1999). 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.

[27]In our analysis of three-factor models for swap rates (Dai and Singleton (2000)) we also found that allowing for negative correlations among the risk factors substantially improved the fits of standard affine models to the conditional distributions of yields, including the conditional means (yield forecastability).

model with $\kappa_{21} = 0$, $\lambda_{21}^Y < 0$ so the state dependence of the second risk premium induces negative correlation under the risk-neutral measure even though it is absent under the physical measure. On the other hand, in the $A_0(2)$SG-C model with $\kappa_{21} \neq 0$, $\lambda_{21}^Y > 0$ so the latter state dependence tends to mitigate the negative correlation under the physical measure. Finally, in the case of model $A_1(2)$SG-C, the large negative value of λ_{21}^Y induces substantially more negative correlation under the risk-neutral measure than (the negative correlation) under the physical measure.

Table 3: ML Estimates of the Two-Factor Affine Models

(Standard errors are in parenthesis.)

Parameter	$A_0(2)$SG-U	$A_0(2)$SG-C	$A_1(2)$CG-U	$A_1(2)$CG-C	$A_1(2)$SG-U	$A_1(2)$SG-C
θ_1	0.00001	0.00001	0.0425	0.0447	0.0425	0.0452
	(.001)	(.001)	(.005)	(.006)	(.010)	(.007)
θ_2	0.0702	0.0672	-0.0021	0.0114	0.0212	0.0224
	(.014)	(.016)	(1.05)	(.851)	(.021)	(.004)
κ_{11}	0.6489	0.6500	0.9373	0.7513	0.9372	0.4887
	(.211)	(.209)	(.110)	(.097)	(.114)	(.132)
κ_{21}	*	-0.2353	*	-0.2591	*	-0.1666
		(.078)		(.029)		(.040)
κ_{22}	0.1217	0.1219	0.0020	0.0024	0.1231	0.3434
	(.038)	(.001)	(.001)	(.001)	(.086)	(.067)
σ_{11}	0.0236	0.0236	0.1092	0.1032	0.1092	0.0995
	(.0005)	(.0005)	(.004)	(.003)	(.009)	(.003)
σ_{22}	0.0101	0.0099	0.0116	0.0084	0.0116	0.0053
	(.001)	(.0007)	(.0003)	(.0003)	(.0004)	(.001)
λ_{11}^Y	0.0900	0.0877	-0.2173	-0.0446	-0.2169	0.2207
	(.211)	(.210)	(.106)	(.091)	(.106)	(.040)
λ_{21}^Y	-0.1080	0.1126	*	*	*	-0.6776
	(.079)	(.121)				(.272)
λ_{22}^Y	-0.1192	-0.1194	*	*	-0.1211	-0.3409
	(.037)	(.050)			(.009)	(.067)
λ_{02}	0.0072	0.0069	-0.0014	-0.0125	0.0001	-0.0005
	(.003)	(.003)	(.002)	(.002)	(.0008)	(.003)
Likelihood	36.547	36.557	36.668	36.712	36.670	36.733

The model-implied $\tilde{d}_n^y$ from the $A_0(2)$ and $A_1(2)$ models are displayed in Figures 4 and 5, respectively, along with the historical results from Table 1 ("Historical Campbell-Shiller"). Recall that if the model-implied risk premiums are well matched to those of the data-generating process, then the term structure of estimated $\tilde{d}_n^y$ should be a horizontal line at unity. Within the Gaussian family $A_0(2)$, allowing for state-dependent risk premiums substantially improves the model's ability to match *LPEH*: compare the "Historical Campbell-Shiller" result (the Gaussian model with constant term premiums) to the result for the Gaussian "SG-U" model. Allowing for non-zero factor correlations (SG-C) further improves the fit, but even this model seems to not fully resolve *LPEH*.

More success at matching *LPEH* is obtained within the family $A_1(2)$ that accommodates stochastic volatility. Again, there is a substantial improvement of fit to *LPEH* by allowing for negatively correlated factors. Moreover, when an extended, state-dependent risk premium is introduced for the second, Gaussian factor and the factors are correlated, the $A_1(2)$SG-C model matches the maturity structure of LPY virtually perfectly for maturities longer than two to three years.[28]

We conjecture that a three-factor model within the families $A_1(3)$ or $A_2(3)$ would have the requisite additional flexibility to match LPY at maturities under two years as well, at least down to six months or so. At the very short end of the treasury yield curve, Duffee (1996) shows that there are significant insitutional/liquidity effects that could affect yield projections. Longstaff (2000) finds that, for very short-term generic repo rates, which are less encumbered by liquidity effects, there is little evidence against the *EH*. The third factor in empirical affine three-factor models typically has the fastest rate of mean reversion leading its influence on yield curve movements to die out relatively quickly (see, e.g., Chen and Scott (1993), Dai and Singleton (2000), and Duffee (1999)) so we do not believe that our basic findings on the maturity patterns of *LPEH* for longer-term rates will be materially affected by the inclusion of a third factor. If anything, a third factor should only improve a model's matching ability.

Figure 5 for the $A_1(2)$ family also sheds light on the relative importance of a state-dependent risk premium for $Y_2(t)$ and allowing for (negatively) correlated factors in matching *LPEH*. Starting with model $A_1(2)CG-U$, we see that having the standard state-dependence of the risk premium for the CIR-style factor $Y_1(t)$ gives virtually no improvement in fit to LPY over a model with constant risk premiums in which *EH* is true. Model $A_1(2)SG-U$ introduces state-dependence of the risk premium for $Y_2(t)$ by having $\lambda_{22}^Y \neq 0$, but $\lambda_{21}^Y = 0$ so this model completely rules out factor correlation (through both the drift matrix κ and risk premium matrix λ_Y). We see that own state-dependence of the risk premium for $Y_2(t)$ leads to a modest improvement in fit to LPY. On the other hand, model $A_1(2)$CG-C allows for factor correlation induced only by $\kappa_{21} \neq 0$, while forcing state-independence of the risk premium for $Y_2(t)$. Interestingly, this parameterization largely eliminates the puzzling negative projection coefficients. Thus, for matching LPY within this family of *DTSM*s, negative factor correlation through the drift seems more important than own state-dependence of the second risk premium. This observation is also likely to be relevant to Duffee's analysis of the forecasting performances of alternative affine models. His reference "completely" affine model is the $A_3(3)$ model with risk premium parameterization (13). Since this model rules out negative factor correlations, his assessment of the improvement in forecasting power of affine models with risk premium specification (13) over those with specification (17) is likely to be exaggerated.

Negative factor correlation through κ alone is not sufficient to match LPY, however. The results for model $A_1(2)$SG-C show that allowing the extra flexibility and amplified negative correlation that comes from having $\lambda_{21}^Y < 0$ is necessary to match LPY.

[28]This matching does not, of course, presume that there is no small-sample bias in these projection coefficients. In fact, from previous studies of both yield regressions and dynamic factor models we know that estimates of conditional means tend to be biased in small samples. We do not expect the bias in one of the estimates of $\tilde{d}_n^y$ to be notably different from the others. Therefore, we view the differences in these figures as arising largely from differences in the dynamics of yields implied by different affine models.

Figure 4: Model Implied Estimates of $\tilde{d}_n^y$ from the $A_0(2)$ models.

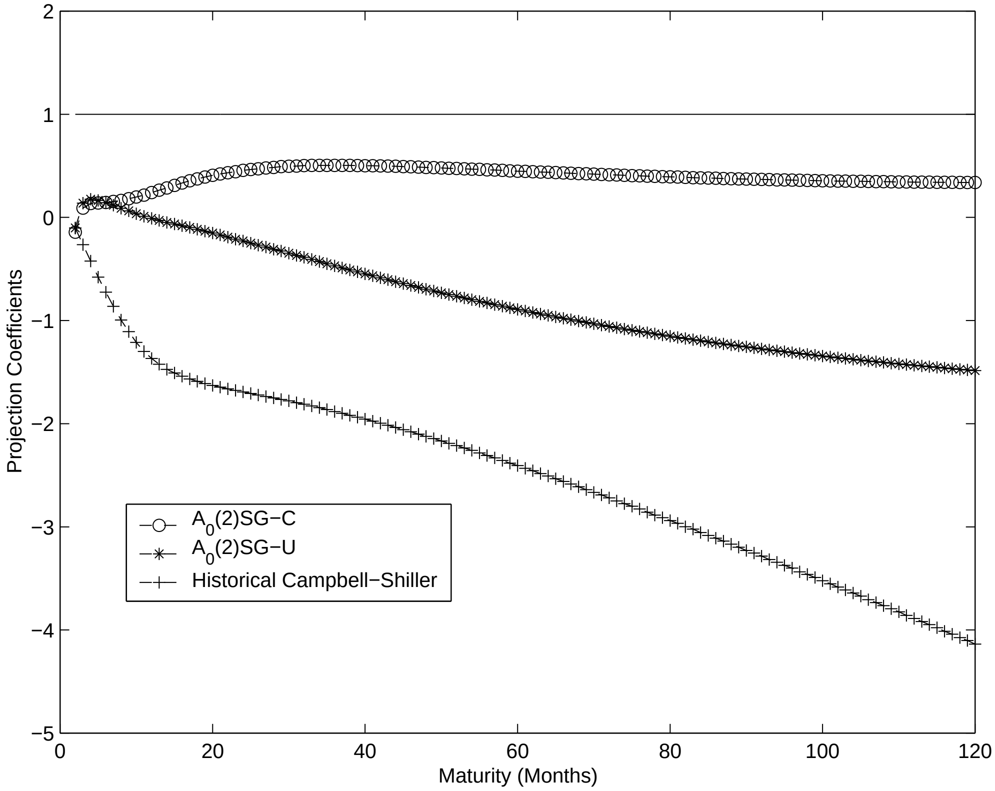
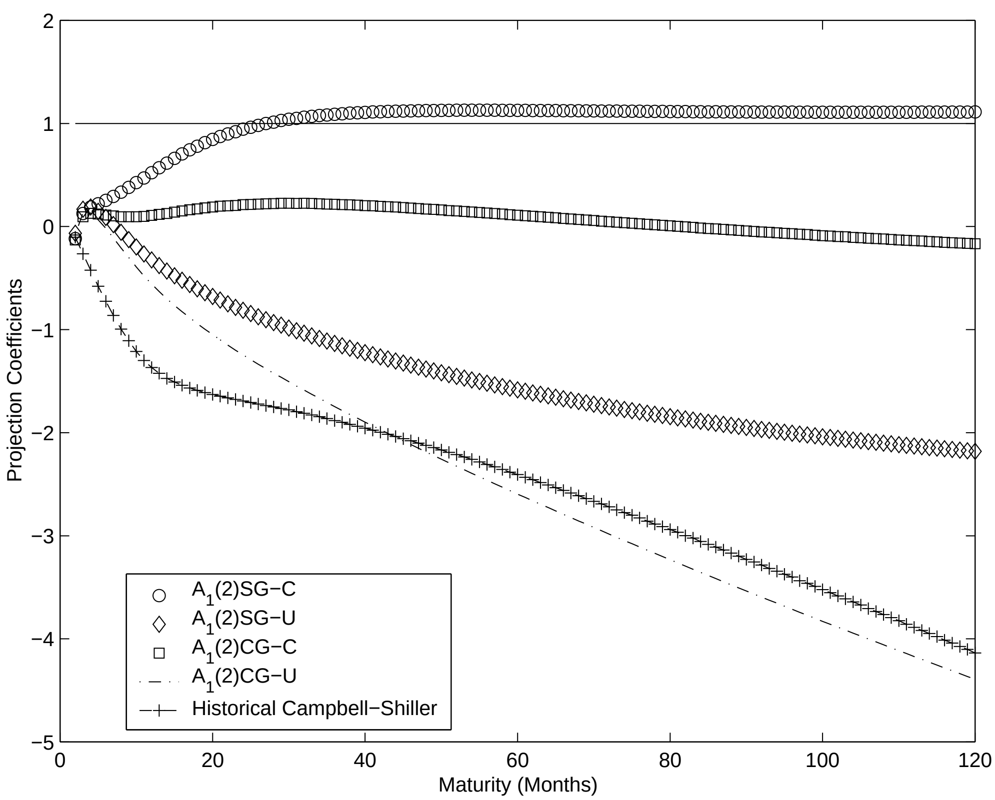


Figure 5: Model Implied Estimates of $\tilde{d}_n^y$ from the $A_1(2)$ models.



6 Conclusion

We began this exploration of expectations puzzles with the conjecture that richer parameterizations of risk premiums – parameterizations that have the risk premiums affecting both the drifts and rates of mean reversion of the state variables– will give the requisite flexibility for matching *LPEH*. 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 *LPEH* quite closely. We then took up the more demanding challenge of formulating models that match *LPEH* 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 two-factor affine models, we showed that a "mixed" Gaussian-square-root diffusion model with correlated factors and state-dependence of both 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 *LPEH*. In our calibration exercise, we found that one-factor versions of these models perform equally well in this regard. However, in our maximum likelihood analysis of two-factor affine models, we found that the Gaussian models with state-dependent risk premiums did not match *LPEH* as well as an affine model with stochastic volatility introduced through a CIR-style factor. Seemingly central to the goodness-of-fit of the latter $A_1(2)$ model was the negative correlation among the two factors. In fact, simply allowing for negative correlation in a standard $A_1(2)$ model with constant risk premium for the Gaussian factor took us a long way (though not all the way) toward resolving *LPEH*.

We do not presume that all of the important features of the conditional distributions of bond yields are captured by a two-factor model. The widely cited analysis of three factors by Litterman and Scheinkman (1991), and our failure to match the maturity structure of LPY at the very short end of the maturity spectrum, suggest otherwise. Extending our analysis to three (or more factors) is an interesting topic for future research. Such extensions would bring more flexibility in matching *LPEH*, because of their richer correlation structures through their drifts and through state-dependence of the risk premiums.[29]

Our focus on affine and quadratic-Gaussian models is motivated by the historical importance of these models. We conjecture that there is a much larger class of *DTSM*s with the key features of the risk premiums outlined above that will also match *LPEH*. One such family is obtained by combining a standard affine model with a Markov switching process as in Bansal and Zhou (2000).[30] From Naik and Lee (1997), for the case of CIR-style models with shifts only in the long-run mean of a square-root diffusion, we know that Markov switching introduces additional free parameters (only) into the "intercept" weights $A(\tau)$ in the pricing relation $\log P(t,\tau) = A(\tau) + B(\tau)'Y(t)$ (see Evans (2000) for the analogous result for

[29] The findings in Duffee (1999) that three-factor models with state-dependent risk premiums on the (conditionally) Gaussian factors out-forecast their affine counterparts with state-independent premiums suggests that the latter source of flexibility will remain important in matching *LPEH*.

[30] Bekaert et al. (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.

discrete-time CIR-style models). Thus, among other things, introducing regime switching into an otherwise standard affine model has the effect of giving the model more flexibility to match both the long-run mean of yields (influenced by the $A(\tau)$) and their correlation structure (determined by the $B(\tau)$). Another family of *DTSM*s that might resolve *LPEH* 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. We also defer to future research the question of whether these models generate the maturity pattern of LPY, while matching other features of the conditional distributions of bond yields.

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 $Y(t)$:

$$r_0(t) = a_0 + Y(t)'b_0, \tag{38}$$

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

$$dY(t) = \kappa(\theta - Y(t))dt + \sigma dW(t), \tag{39}$$

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

The market price of risk[31] is given by

$$\Lambda(t) = \sigma^{-1}(\lambda^0 + \lambda^Y Y(t)), \tag{40}$$

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,[32] the risk neutral dynamics of the state vector is given by

$$dY(t) = \tilde{\kappa}(\tilde{\theta} - Y(t))dt + \sigma d\tilde{W}(t), \tag{41}$$

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, $1 \leq i \leq N$, X is a non-singular real matrix, with diagonal elements normalized to 1.[33] 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

$$E\left[Y(t+\tau)|Y(t)\right] = e^{-\kappa\tau}Y(t) + (I - e^{-\kappa\tau})\theta. \tag{42}$$

The conditional variance is given by

$$\mathrm{Var}(Y(t+\tau)|Y(t)) = X^{-1}\Omega(\tau)X'^{-1}, \tag{43}$$

[31] The pricing kernel is given by

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

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

[33] 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

$$\Omega_{ij}(\tau) = \Sigma_{ij} \frac{1 - e^{-(\kappa_i + \kappa_j)\tau}}{\kappa_i + \kappa_j}, \tag{44}$$

$$\Sigma = X\sigma\sigma' X'. \tag{45}$$

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

$$P(t,\tau) = e^{-A(\tau) - B(\tau)' Y(t)}, \tag{46}$$
$$R(t,\tau) = a(\tau) + b(\tau)' Y(t), \tag{47}$$

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

$$b(\tau) = (I - e^{-\tilde{\kappa}'\tau})(\tilde{\kappa}'\tau)^{-1} b_0, \tag{48}$$

$$a(\tau) = a_0 + (b_0 - b(\tau))'\tilde{\theta} - \frac{1}{2}\mathrm{Tr}\left[\Xi(\tau)\tilde{X}'^{-1}\tilde{\kappa}'^{-1} b_0 b_0' \tilde{\kappa}^{-1}\tilde{X}^{-1}\right], \tag{49}$$

$$\Xi_{ij}(\tau) = \tilde{\Sigma}_{ij}\left[1 - \frac{1 - e^{-\tilde{\kappa}_i\tau}}{\tilde{\kappa}_i\tau} - \frac{1 - e^{-\tilde{\kappa}_j\tau}}{\tilde{\kappa}_j\tau} + \frac{1 - e^{-(\tilde{\kappa}_i + \tilde{\kappa}_j)\tau}}{(\tilde{\kappa}_i + \tilde{\kappa}_j)\tau}\right], \tag{50}$$

$$\tilde{\Sigma} = \tilde{X}\sigma\sigma'\tilde{X}'. \tag{51}$$

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

$$E_t\left[r_{t+n}\right] = \mu_n + Y(t)'\nu_n, \text{ where} \tag{52}$$
$$\mu_n = a_1 + \theta'(I - e^{-\kappa' n})b_1 \tag{53}$$
$$\nu_n = e^{-\kappa' n} b_1 \tag{54}$$

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

$$f_t^n \equiv -\frac{1}{\Delta}\ln\frac{P(t,(n+1)\Delta)}{P(t,n\Delta)} = A_n^{\Delta} + B_n^{\Delta'} Y(t), \tag{55}$$

where

$$A_n^{\Delta} \equiv \frac{A_{n+1} - A_n}{\Delta} \text{ and } B_n^{\Delta} \equiv \frac{B_{n+1} - B_n}{\Delta}. \tag{56}$$

Thus, the forward risk premium is given by

$$p_t^n \equiv f_t^n - E_t[r_{t+n}] = (A_n^\Delta - \mu_n) + Y(t)'(B_n^\Delta - \nu_n), \tag{57}$$

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

$$\begin{aligned} A(\tau) &= a_0\tau + (\tau - B(\tau))(\tilde{\theta} - \frac{\sigma^2}{2\tilde{\kappa}^2}) + \frac{\sigma^2}{4\tilde{\kappa}}B(\tau)^2, && (58) \\ B(\tau) &= \frac{1-e^{-\tilde{\kappa}\tau}}{\tilde{\kappa}} b_0. && (59) \end{aligned}$$

The forward-spot spread is given by

$$f_t^n - r_t = (A_n^\Delta - a_1) + (B_n^\Delta - b_1)Y(t). \tag{60}$$

Substituting (60) into (57), we have

$$f_t^n - E_t[r_{t+n}] = \delta_n + \alpha_n(f_t^n - r_t), \tag{61}$$

where

$$\begin{aligned} \delta_n &= \frac{A_n^\Delta - \mu_n}{B_n^\Delta - b_1}, && (62) \\ \alpha_n &= \frac{B_n^\Delta - e^{-\kappa' n\Delta} b_1}{B_n^\Delta - b_1} = \frac{e^{-\tilde{\kappa}n} - e^{-\kappa n}}{e^{-\tilde{\kappa}n} - 1} && (63) \end{aligned}$$

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

A.3 Two-factor Case

In the two-factor model, we use the forward-spot spread and the one-period rate to back out the state vector. Since

$$\begin{aligned} f_t^n - r_t &= (A_n^\Delta - a_1) + (B_n^\Delta - b_1)'Y(t), && (64) \\ r_t &= a_1 + b_1'Y(t), && (65) \end{aligned}$$

we have

$$Y(t) = J'^{-1}\begin{pmatrix} f_t^n - r_t - (A_n^\Delta - a_1) \\ r_t - a_1 \end{pmatrix}, \tag{66}$$

where J is a 2×2 matrix formed by stacking $(B_n^\Delta - b_1)$ and b_1. It follows that

$$p_t^n = \delta_n + \alpha_n(f_t^n - r_t) + \beta_n r_t, \tag{67}$$

where

$$\begin{pmatrix} \alpha_n \\ \beta_n \end{pmatrix} = J^{-1}(B_n^\Delta - e^{-\kappa' n\Delta} b_1), \tag{68}$$

$$\delta_n = (1 - \alpha_n)(A_n^\Delta - a_1) - \beta_n a_1. \tag{69}$$

Since $E[p_t^n] = E[f_t^n - r_t]$, δ_n can be related to the sample means of the forward spread and the short rate:

$$\delta_n = (1 - \alpha_n)E(f_t^n - r_t) - \beta_n E(r_t). \tag{70}$$

When the state variables are independent, we obtain

$$\alpha_n = \frac{e^{-\tilde{\kappa}_1 n} - e^{-\kappa_1 n}}{e^{-\tilde{\kappa}_1 n} - e^{-\tilde{\kappa}_2 n}} - \frac{e^{-\tilde{\kappa}_2 n} - e^{-\kappa_2 n}}{e^{-\tilde{\kappa}_1 n} - e^{-\tilde{\kappa}_2 n}}, \tag{71}$$

$$\beta_n = -\left(e^{-\tilde{\kappa}_2 n} - 1\right)\alpha_n + \left(e^{-\tilde{\kappa}_2 n} - e^{-\kappa_2 n}\right). \tag{72}$$

Similar relationship can be derived for any two-factor affine models. We refer interested readers to Dai and Singleton (2000) for details.

B Quadratic-Gaussian Model

Starting with the specification of the one-factor quadratic-Gaussian model in Section 4.3, 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

$$r_t = a_1 + b_1 Y_t + c_1 Y_t^2, \tag{73}$$

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

$$f_t^n \equiv -\frac{1}{\Delta}\log\frac{P_t^{n+1}}{P_t^n} = A_n^\Delta + B_n^\Delta Y_t + C_n^\Delta Y_t^2. \tag{74}$$

The expected short rate is given by

$$E_t[r_{t+n}] = \mu_n + \nu_n Y_t + \omega_n Y_t^2, \tag{75}$$

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:

$$f_t^n - E_t[r_{t+n}] = \delta_n + \alpha_n(f_t^n - r_t) + \beta_n r_t, \tag{76}$$

where

$$\begin{pmatrix} B_n^\Delta - b_1 & b_1 \\ C_n^\Delta/c_1 - 1 & 1 \end{pmatrix}\begin{pmatrix} \alpha_n \\ \beta_n \end{pmatrix} = \begin{pmatrix} B_n^\Delta - \nu_n \\ C_n^\Delta/c_1 - \omega_n/c_1 \end{pmatrix}, \tag{77}$$

or

$$\alpha_n = 1 - \frac{\nu_n/b_1 - \omega_n/c_1}{B_n^\Delta/b_1 - C_n^\Delta/c_1} \tag{78}$$

$$\beta_n = (B_n^\Delta/b_1 - \nu_n/b_1) - (B_n^\Delta/b_1 - 1)\alpha_n \tag{79}$$

and

$$\delta_n = (1-\alpha_n)E[f_t^n - r_t] - \beta_n E_t[r_t]. \tag{80}$$

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.[34]

C Conditions for Girsanov's Theorem

The goal is to show that

$$Z(t) = e^{\int_0^t \Lambda_s' dW_s - \frac{1}{2}\int_0^t \Lambda_s'\Lambda_s ds}, \tag{81}$$

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

$$|\Lambda_t| \leq K_T(1 + W^*(t)),\ 0 \leq t \leq T, \tag{82}$$

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

[34]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.

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 Λ_t is an affine function of $Y(t)$, it is obvious that (82) holds.

References

Ahn, D.-H., Dittmar, R. F., and Gallant, A. R. (2000). Quadratic gaussian models: Theory and evidence. working paper, University of North Carolina.

Backus, D., Foresi, S., Mozumdar, A., and Wu, L. (1997). Predictable changes in yields and forward rates. Working Paper, New York University.

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

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

Bekaert, G., Hodrick, R., and Marshall, D. (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., Hodrick, R., and Marshall, D. (1997b). 'Peso problem' explanations for term structure anomalies. Working paper, Columbia University.

Campbell, J. Y. and Shiller, R. J. (1991). Yield spreads and interest rate movements: A bird's eye view. *Review of Economic Studies*, 58:495–514.

Chen, R. and Scott, L. (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., Ingersoll, J. E., and Ross, S. A. (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 Singleton, K. (2000). Specification analysis of affine term structure models. *Journal of Finance*, forthcoming.

Duarte, J. (1999). The relevance of parameterization of market price of risk in affine term structure models. Working paper, University of Chicago.

Duffee, G. (1996). Idiosyncratic variation in treasury bill yields. *Journal of Finance*, 51:527–552.

Duffee, G. R. (1999). Forecasting future interest rates: Are affine models failures? Working paper, University of California, Berkeley.

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

Duffie, D. and Singleton, K. (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 Bliss, R. R. (1987). The information in long-maturity forward rates. *American Economic Review*, 77(4):680–692.

Fisher, M. (1998). A simple model ofthe 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 Shreve, S. (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.

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

Longstaff, F. (2000). The term structure of very short term rates: New evidence for the expectations hypothesis. *Journal of Financial Economics*, forthcoming.

McCallum, B. T. (1994). Monetary policy and the term structure of interest rates. NBER Working Paper No. 4938.

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

Pearson, N. D. and Sun, T. (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.

Pedersen, D. D. L. and Singleton, K. (1999). Modeling credit spreads on sovereign debt. working paper, Stanford University.

Roberds, W. and Whiteman, C. (1999). Endogenous term premia and anamolies in the term structure of interest rates: Explaining the predictability smile. *Journal of Monetary Economics*, 44:555–580.