Term Premia and Interest Rate Forecasts in Affine Models

GREGORY R. DUFFEE*

ABSTRACT

The standard class of affine models produces poor forecasts of future Treasury yields. Better forecasts are generated by assuming that yields follow random walks. The failure of these models is driven by one of their key features: Compensation for risk is a multiple of the variance of the risk. Thus risk compensation cannot vary independently of interest rate volatility. I also describe a broader class of models. These "essentially affine" models retain the tractability of standard models, but allow compensation for interest rate risk to vary independently of interest rate volatility. This additional flexibility proves useful in forecasting future yields.

CAN WE USE FINANCE THEORY to tell us something about the empirical behavior of Treasury yields that we do not already know? In particular, can we sharpen our ability to predict future yields? A long-established fact about Treasury yields is that the current term structure contains information about future term structures. For example, long-maturity bond yields tend to fall over time when the slope of the yield curve is steeper than usual. These predictive relations are based exclusively on the time-series behavior of yields. To rule out arbitrage, the cross-sectional and time-series characteristics of the term structure are linked in an internally consistent way. In principle, imposing these restrictions should allow us to exploit more of the information in the current term structure, and thus improve forecasts. But in practice, existing no-arbitrage models impose other restrictions for the sake of tractability; thus their value as forecasting tools is a priori unclear.

I examine the forecasting ability of the affine class of term structure models. By "affine," I refer to models where zero-coupon bond yields, their physical dynamics, and their equivalent martingale dynamics are all affine functions of an underlying state vector. A variety of nonaffine models have been developed, but the tractability and apparent richness of the affine class has led the finance profession to focus most of its attention on such models.

Although forecasting future yields is important in its own right, a model that is consistent with finance theory and produces accurate forecasts can make a deeper contribution to finance. It should allow us to address a key

issue: explaining the time variation in expected returns to assets. In the context of the term structure, explaining time-varying returns means explaining the failure of the expectations hypothesis of interest rates. Put differently, we would like to have an intuitive explanation for the positive correlation between the yield curve slope and expected excess returns to long bonds. If a model produces poor forecasts of future yields (and thus poor forecasts of future bond prices), it is unlikely that the model can shed light on the economics underlying the failure of the expectations hypothesis.

The first main conclusion reached here is that the class of affine models studied most extensively to date fails at forecasting. I refer to this class, which includes multifactor generalizations of both Vasicek (1977) and Cox, Ingersoll, and Ross (1985), and is analyzed in Dai and Singleton (2000), as "completely affine." I fit general three-factor completely affine models to the Treasury term structure over the period 1952 through 1994. Yield forecasts produced with these estimated models are typically worse than forecasts produced by simply assuming yields follow random walks. This conclusion holds for both in-sample forecasts and out-of-sample (1995 through 1998) forecasts.

Even more damning is the way in which the estimated models fail. They produce yield forecast errors that are strongly negatively correlated with the slope of the yield curve. In other words, the models fail to replicate the key empirical relation between expected returns and the slope of the yield curve; their underestimation of expected excess returns to long bonds tends to be largest when the slope of the term structure is steep.

This failure is a consequence of two features of the Treasury term structure, combined with a restriction built into completely affine models. The first feature is that the distribution of yields is not strongly skewed; yields vary widely over time around both sides of their sample means. The second feature is that, while the average excess return to Treasury bonds is not much greater than zero, the slope of the term structure predicts a relatively large amount of variation in excess returns to bonds. Fama and French (1993) note that this implies the sign of predicted excess returns to Treasury bonds changes over time.

Completely affine models do not simultaneously reproduce these two features of term structure behavior. The key restriction in these models is that compensation for risk is a fixed multiple of the variance of the risk. This structure ensures that the models satisfy a requirement of no-arbitrage: Risk compensation goes to zero as risk goes to zero. But because variances are nonnegative, this structure restricts the variability of the compensation that investors expect to receive for facing a given risk. The compensation is bounded by zero; therefore it cannot switch sign over time.

As will be made clear in the paper, the only way this framework can produce expected excess returns with low means and high volatilities is for some underlying factors driving the term structure to be highly positively skewed. But this strong positive skewness is inconsistent with the actual distribution of yields. Thus completely affine models can fit either of these features of Treasury yields, but not both simultaneously.

All is not lost, however. The second main conclusion of this paper is that the completely affine class can be extended to break the link between risk compensation and interest rate volatility. This extension from the completely affine class to the "essentially affine" class described here is costless, in the sense that the affine time-series and cross-sectional properties of bond prices are preserved in essentially affine models. The existence of extensions to the completely affine class is not new (Chacko (1997) constructs a general equilibrium example), but this paper is the first to describe and empirically investigate a general, very tractable extension to completely affine term structure models. I find that essentially affine models produce more accurate yield forecasts than completely affine models, both in sample and out of sample. However, there is a trade-off between flexibility in forecasting future yields and flexibility in fitting interest rate volatility.

The paper is organized as follows. The structure of affine models is discussed in detail in Section I. Section II explains intuitively why completely affine models work poorly. Section III describes the estimation technique. Section IV presents empirical results and Section V concludes.

I. Affine Models of the Term Structure

A. Affine Bond Pricing

The core of affine term structure models is the framework of Duffie and Kan (1996). Their model, summarized here, describes the evolution of bond prices under the equivalent martingale measure. Uncertainty is generated by n Brownian motions, $\widetilde{W}_t \equiv (\widetilde{W}_{t,1}, \ldots, \widetilde{W}_{t,n})'$. There are n state variables, denoted $X_t \equiv (X_{t,1}, \ldots, X_{t,n})'$. The instantaneous nominal interest rate r_t is affine in these state variables:

$$r_t = \delta_0 + \delta X_t,$$

where δ_0 is a scalar and δ is an n-vector. The evolution of the state variables under the equivalent martingale measure is

$$dX_t = [(K\theta)^Q - K^Q X_t]dt + \Sigma S_t\, d\widetilde{W}_t, \tag{1}$$

where K^Q and Σ are $n \times n$ matrices and $(K\theta)^Q$ is an n-vector. The Q superscript distinguishes parameters under the equivalent martingale measure from corresponding parameters under the physical measure. The matrix S_t is diagonal with elements

$$S_{t(ii)} \equiv \sqrt{\alpha_i + \beta_i' X_t}, \tag{2}$$

where β_i is an n-vector and α_i is scalar. It is convenient to stack the β_i vectors into the matrix β, where β_i' is row i of β. The scalars α_i are stacked in the n-vector α. The following discussion assumes that the dynamics of (1)

are well defined, which requires that $\alpha_i + \beta_i' X_t$ is nonnegative for all i and all possible X_t. Parameter restrictions that ensure these requirements are in Dai and Singleton (2000).

Denote the time-t price of a zero-coupon bond maturing at time $t + \tau$ as $P(X_t, \tau)$. Duffie and Kan (1996) show

$$P(X_t, \tau) = \exp[A(\tau) - B(\tau) X_t], \tag{3}$$

where $A(\tau)$ is a scalar function and $B(\tau)$ is an n-valued function. Thus, the bond's yield is affine in the state vector:

$$Y(X_t, \tau) = (1/\tau)[-A(\tau) + B(\tau)' X_t]. \tag{4}$$

The functions $A(\tau)$ and $B(\tau)$ can be calculated numerically by solving a series of ordinary differential equations (ODEs).

B. *The Price of Risk and Expected Returns to Bonds*

The model is completed by specifying the dynamics of X_t under the physical measure, which is equivalent to specifying the dynamics of the price of risk. Denote the state price deflator by π_t. The relative dynamics of π_t are

$$\frac{d\pi_t}{\pi_t} = -r_t dt - \Lambda_t' dW_t, \tag{5}$$

where W_t follows a Brownian motion under the physical measure. Element i of the n-vector Λ_t represents the price of risk associated with $W_{t,i}$. The dynamics of X_t under the physical measure can be written in terms of Λ_t and (1):

$$dX_t = ((K\theta)^Q - K^Q X_t) dt + \Sigma S_t \Lambda_t dt + \Sigma S_t dW_t. \tag{6}$$

Instantaneous bond-price dynamics can be written as

$$\frac{dP(X_t, \tau)}{P(X_t, \tau)} = (r_t + e_{\tau,t}) dt + v_{\tau,t} dW_t,$$

where $e_{\tau,t}$ denotes the instantaneous expected excess return to holding the bond. An application of Ito's lemma combined with the structure of the ODEs in Duffie and Kan (1996) reveals

$$e_{\tau,t} = -B(\tau)' \Sigma S_t \Lambda_t, \tag{7}$$

$$v_{\tau,t} = -B(\tau)' \Sigma S_t. \tag{8}$$

Equation (7) says that variations over time in expected excess returns are driven by variations in both the volatility matrix S_t and the price of risk vector Λ_t. A parametric model for bond-yield dynamics requires a functional form for Λ_t. This form should be sufficiently flexible to capture the empirically observed behavior of expected excess returns. Thus, to motivate the choice of functional form for Λ_t, we briefly review evidence on the behavior of bond returns.

A large literature documents that expected excess returns to Treasury bonds (over returns to short-term Treasury bills) are, on average, near zero, and vary systematically with the term structure.[1] When the slope of the term structure is steeper than usual, expected excess returns to bonds are high, while expected excess returns are low—often negative—when the slope is less steep. Thus the ratio of mean expected excess bond returns to the standard deviation of expected excess bond returns is low.

Earlier work has also shown that the shape of the term structure is related to the volatility of yields.[2] However, the slope-expected return relation is not simply proxying for a volatility-expected return relation. Supporting evidence is in Table I, which reports results from regressions of excess monthly bond returns on the lagged slope of the term structure and lagged yield volatility. Monthly returns to portfolios of Treasury bonds are from the Center for Research in Security Prices. Excess returns to these portfolios are produced by subtracting the contemporaneous return to a three-month Treasury bill. The slope of the term structure is measured by the difference between month-end five-year and three-month zero-coupon yields. The yields are interpolated from coupon bonds using the technique of McCulloch and Kwon (1993), as implemented by Bliss (1997).[3] Yield volatility is the standard deviation of the five-year zero-coupon yield, measured by the square root of the sum of squared daily changes in the yield during the month.

The sample period is July 1961 through December 1998. The results in Table I reveal that month t's volatility has no statistically significant predictive power for excess bond returns in month $t + 1$. By contrast, all of the estimated slope parameters are significant at the 10 percent level and half are significant at the 5 percent level. In addition, the variation in predicted excess returns is large relative to mean excess returns. Consider, for example, bonds with maturities between three and four years. The mean excess return is 7 basis points per month, while the standard deviation of predictable excess returns is roughly 25 basis points. Armed with this information about the empirical behavior of bond returns, we now discuss three different parameterizations of Λ_t.

---

[1] The literature is too large to cite in full here. Early research includes Fama and Bliss (1987). Two standard references are Fama and French (1989, 1993).

[2] This literature is also too large to cite in full. Chan et al. (1992) examine the sensitivity of volatility to the level of short-term interest rates. Andersen and Lund (1997) refine their work by decomposing the variation in interest-rate volatility into a component related to the level of short-term interest rates and a stochastic volatility component.

[3] I thank Rob Bliss for providing me with the yield data.

Table I

Regressions of Excess Returns to Treasury Bonds, July 1961 through December 1998

Monthly excess returns to portfolios of Treasury bonds are regressed on the previous month's term-structure slope and an estimate of the interest rate volatility during the previous month. The slope of the term structure is measured by the difference between five-year and three-month zero-coupon yields (interpolated from coupon bonds). Monthly volatility is measured by the square root of the sum of squared daily changes in the five-year zero-coupon bond yield. Asymptotic t-statistics, adjusted for generalized heteroskedasticity, are in parentheses. There are 449 monthly observations.

Maturity (years)	Mean Excess Return (%)	Coefficient on		Std. Dev. of Fitted Excess Rets
		Slope	Volatility	
$0 < m \leq 1$	0.011	0.027 (1.76)	0.116 (0.96)	0.036
$1 < m \leq 2$	0.045	0.085 (1.85)	0.413 (1.27)	0.119
$2 < m \leq 3$	0.064	0.132 (1.88)	0.582 (1.20)	0.179
$3 < m \leq 4$	0.074	0.187 (2.38)	0.706 (1.35)	0.241
$4 < m \leq 5$	0.063	0.214 (2.37)	0.692 (1.16)	0.265
$5 < m \leq 10$	0.094	0.296 (2.69)	0.804 (1.08)	0.354

C. Completely Affine Models

Fisher and Gilles (1996) and Dai and Singleton (2000) adopt the following form for Λ_t. Let λ_1 be an n-vector. Then the price of risk vector Λ_t is given by

$$\Lambda_t = S_t \lambda_1. \tag{9}$$

This class nests multifactor versions of Vasicek (1977) and Cox, Ingersoll, and Ross (1985; hereafter CIR). The main reason for the popularity of this form is that the vector $S_t \Lambda_t$ is affine in X_t. This implies affine dynamics for X_t under both the equivalent martingale and physical measures. Affine dynamics of X_t under the physical measure allow for closed-form calculation of various properties of conditional densities of discretely sampled yields. These properties are discussed in detail in Duffie, Pan, and Singleton (2000) and Singleton (2001). Of less importance is the fact that $\Lambda'_t \Lambda_t$, which is the instantaneous variance of the log state price deflator, is also affine in X_t. This latter property motivates the term "completely affine," as discussed in the next subsection.

This structure imposes two related limitations on Λ_t. First, variation in the price of risk vector is completely determined by the variation in S_t. Therefore, variations in expected excess returns to bonds are driven exclusively by the volatility of yields, an implication that appears inconsistent with the evidence in Table I. Second, the sign of element i of Λ_t is the same as that of element i of λ_1, because the diagonal elements of S_t are restricted to be nonnegative. The importance of this limitation will be clear in Section II.

D. Essentially Affine Models

The essentially affine class nests the completely affine class. We first define the elements of a diagonal matrix S_t^- as

$$S_{t(ii)}^- = \begin{cases} (\alpha_i + \beta_i' X_t)^{-1/2}, & \text{if } \inf(\alpha_i + \beta_i' X_t) > 0; \\ 0, & \text{otherwise.} \end{cases}$$

Thus, if diagonal element i of S_t is bounded away from zero, its reciprocal is diagonal element i of S_t^-. For any diagonal element of S_t with a lower bound of zero (whether or not it is accessible given the dynamics of X_t), the associated element of S_t^- is set to zero. Therefore the elements of S_t^- do not explode as the corresponding elements of S_t approach zero.

The form of Λ_t used in the essentially affine model is

$$\Lambda_t = S_t \lambda_1 + S_t^- \lambda_2 X_t, \tag{10}$$

where λ_2 is an $n \times n$ matrix. This form shares with (9) two important properties. First, if $S_{t(ii)}$ approaches zero, Λ_t does not go to infinity. Second, $S_t \Lambda_t$ is affine in X_t. Therefore, the physical dynamics of X_t are affine.

There are three important differences between (9) and (10). First, with $\lambda_2 \neq 0$, $\Lambda_t' \Lambda_t$ is not affine in X_t. Therefore, this model is not completely affine, but the variance of the state price deflator does not affect bond prices. This motivates the term "essentially affine." Second, the tight link between the price of risk vector and the volatility matrix is broken. The essentially affine setup allows for independent variation in prices of risk, which is the kind of flexibility needed to fit the empirical behavior of expected excess returns to bonds. Third, the sign restriction on the individual elements of Λ_t is removed.

For future reference, we require an expression for the physical dynamics of X_t. Substitute (10) into (6) and define I^- as the $n \times n$ diagonal matrix with $I_{ii}^- = 1$ if $S_{t(ii)}^- \neq 0$, $I_{ii}^- = 0$ if $S_{t(ii)}^- = 0$. Then the physical dynamics in the essentially affine model are

$$dX_t = ((K\theta)^Q - K^Q X_t)dt + \Sigma[S_t^2 \lambda_1 + I^- \lambda_2 X_t]dt + \Sigma S_t dW_t. \tag{11}$$

Combining terms and denoting element i of λ_1 by λ_{1i}, (11) can be written as

$$dX_t = (K\theta - KX_t)\,dt + \Sigma S_t\,dW_t, \tag{12a}$$

where

$$K = K^Q - \Sigma \begin{pmatrix} \lambda_{11}\beta_1' \\ \cdot \\ \cdot \\ \cdot \\ \lambda_{1n}\beta_n' \end{pmatrix} + \Sigma I^-\lambda_2 \tag{12b}$$

and

$$K\theta = (K\theta)^Q + \Sigma \begin{pmatrix} \alpha_1\lambda_{11} \\ \cdot \\ \cdot \\ \cdot \\ \alpha_n\lambda_{1n} \end{pmatrix}. \tag{12c}$$

E. An Essentially Affine Example

The following two-factor model illustrates the essentially affine framework. The instantaneous interest rate r_t follows a Gaussian process and there is some factor f_t that follows a square-root process. It is convenient to begin by modeling their dynamics under the physical measure. Under this measure, the processes are independent:

$$\begin{aligned}
d\begin{pmatrix} f_t \\ r_t \end{pmatrix} &= \begin{pmatrix} k_f & 0 \\ 0 & k_r \end{pmatrix}\left(\begin{pmatrix} \bar{f} \\ \bar{r} \end{pmatrix} - \begin{pmatrix} f_t \\ r_t \end{pmatrix} \right) dt \\
&\quad + \begin{pmatrix} \sigma_f & 0 \\ 0 & \sigma_r \end{pmatrix}\begin{pmatrix} \sqrt{f_t} & 0 \\ 0 & 1 \end{pmatrix} d\begin{pmatrix} W_{t,1} \\ W_{t,2} \end{pmatrix}.
\end{aligned} \tag{13}$$

The model is closed with a description of the dynamics of the market price of risk. If we adopt the completely affine version in (9), the result is the Vasicek (1977) model for r_t. In such a setup, the variable f_t is irrelevant for bond pricing, and we are left with a standard one-factor Gaussian model.

If, however, we use the essentially affine specification for the market price of risk, the factor f_t can affect bond prices, even though it cannot affect the path of r_t. The reason is that the compensation that investors require to face

the risk of $W_{t,2}$ can vary with f_t. The essentially affine model specifies the price of risk Λ_t as

$$
\Lambda_t = \begin{pmatrix} \lambda_{11} \sqrt{f_t} \\ \lambda_{12} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda_{2(11)} & \lambda_{2(12)} \\ \lambda_{2(21)} & \lambda_{2(22)} \end{pmatrix} \begin{pmatrix} f_t \\ r_t \end{pmatrix}
$$

$$
= \begin{pmatrix} \lambda_{11} \sqrt{f_t} \\ \lambda_{12} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \lambda_{2(21)} & \lambda_{2(22)} \end{pmatrix} \begin{pmatrix} f_t \\ r_t \end{pmatrix}.
$$

The dynamics of the state price deflator are therefore

$$
\frac{d\pi_t}{\pi_t} = -r_t \, dt - \begin{pmatrix} \lambda_{11} \sqrt{f_t} \\ \lambda_{12} + \lambda_{2(21)} f_t + \lambda_{2(22)} r_t \end{pmatrix}' d \begin{pmatrix} W_{t,1} \\ W_{t,2} \end{pmatrix}.
$$

The dynamics of r_t and f_t under the equivalent martingale measure are, from (12a) and (12b),

$$
\begin{aligned}
d \begin{pmatrix} f_t \\ r_t \end{pmatrix} &= \begin{pmatrix} k_f + \sigma_f \lambda_{11} & 0 \\ \sigma_r \lambda_{2(21)} & k_r + \sigma_r \lambda_{2(22)} \end{pmatrix} \left(\begin{pmatrix} \bar{f}^Q \\ \bar{r}^Q \end{pmatrix} - \begin{pmatrix} f_t \\ r_t \end{pmatrix} \right) dt \\
&\quad + \begin{pmatrix} \sigma_f & 0 \\ 0 & \sigma_r \end{pmatrix} \begin{pmatrix} \sqrt{f_t} & 0 \\ 0 & 1 \end{pmatrix} d \begin{pmatrix} \widetilde{W}_{t,1} \\ \widetilde{W}_{t,2} \end{pmatrix},
\end{aligned}
\tag{14}
$$

where $\bar{f}^Q$ and $\bar{r}^Q$ are the means of f_t and r_t under the equivalent martingale measure.

There are three important differences between this description of bond-price dynamics and the standard Vasicek model. First, the instantaneous interest rate r_t affects the price of interest rate risk, through the parameter $\lambda_{2(22)}$. In Vasicek, the price of interest rate risk is constant. Second, there is a source of uncertainty in bond prices that is independent of the physical dynamics of r_t. The factor f_t affects bond prices through the parameter $\lambda_{2(21)}$. Chacko (1997) builds an affine term structure model expressly designed to exhibit this second feature, and my example was inspired by his (substantially more complicated) model. We will see in Section IV that this kind of feature is critical to understanding the actual dynamics of Treasury bond yields. Third, the price of risk associated with innovations in $W_{t,2}$ can change sign, depending on the level of the factor f_t.

The essentially affine structure of Λ_t, although more flexible than the completely affine structure, nonetheless imposes limits on the possible dynamics of bond prices. Note that one element of K, which is the first matrix on the right-hand side of (13), is the same as the corresponding element of K^Q, which is the first matrix on the right-hand side of (14). Element (1,2)

must be zero under both the physical and equivalent martingale measures. Otherwise, the drift of f_t at $f_t = 0$ could be negative (because it would depend on r_t), which cannot be allowed because $\sqrt{f_t}$ enters into S_t.

To free up this element, and thus allow for a more flexible specification of the price of risk, we could model f_t as a Gaussian process. An example of such a model is Fisher (1998). By contrast, if both f_t and r_t were modeled as square-root diffusion processes, the essentially affine structure of Λ would be identical to the completely affine structure. This illustrates a more general point noted by Duffie and Kan (1996) and Dai and Singleton (2000), and that we will see in Section IV. With the affine setup, there is a trade-off between constructing a model that can capture complicated dynamics in volatilities and a model that can capture complicated dynamics in expected returns.

F. Semi-affine Models

Duarte (2000) chooses an alternative generalization of completely affine models. Let λ_0 be an n-vector. The price of risk vector is

$$\Lambda_t = \lambda_0 + S_t \lambda_1.$$

With this form, elements of Λ_t can switch sign over time, but they cannot move independently of S_t. As noted in Section I.C, this latter feature appears inconsistent with the empirical evidence. Thus, at first glance, it seems that the semi-affine setup allows for some, but not all, of the flexibility of the essentially affine setup. However, there are parameterizations of S_t for which the semi-affine model offers more flexibility than does the essentially affine model. One example is the multifactor CIR model, which is the focus of Duarte's empirical work. The semi-affine form of Λ_t implies nonaffine dynamics of X_t under the physical measure. Therefore, as noted by Duarte (2000), approximation or simulation techniques are typically necessary to reproduce the properties of discretely sampled yields.

G. A Canonical Form for Essentially Affine Models

There are a variety of normalizations that can be imposed on affine models. I follow the lead of Dai and Singleton's (2000) canonical completely affine model. They normalize Σ to the identity matrix. The form of the model is determined by the total number of state variables n and the number of state variables that affect the instantaneous variance of X_t, denoted m. The state vector is ordered so these are the first m elements of X_t. The resulting model is an $A_m(n)$ model. They set the first m elements of α to zero and the remaining $n - m$ elements to one. Their version of (2) is

$$S_{t(ii)} = \begin{cases} \sqrt{X_{t,i}}, & i = 1, \ldots, m; \\ \sqrt{1 + \beta_i' X_t} & i = m + 1, \ldots, n, \end{cases} \tag{15}$$

where for $i = m + 1, \ldots, n$,

$$\beta_i' = (\beta_{i1} \quad \ldots \quad \beta_{im} \quad 0 \quad \ldots \quad 0).$$

Using their framework, we can write the diagonal elements of S_t^- and I^- as

$$S_{t(ii)}^- = \begin{cases} 0, & i = 1, \ldots, m; \\ (1 + \beta_i' X_t)^{-1/2}, & i = m + 1, \ldots, n \end{cases}$$

$$I_{ii}^- = \begin{cases} 0, & i = 1, \ldots, m; \\ 1, & i = m + 1, \ldots, n. \end{cases} \tag{16}$$

Note that in (11), the matrix λ_2 shows up only in the term $I^-\lambda_2 X_t$. Therefore, we can normalize the first m rows of λ_2 to zero. Now reconsider (7), the instantaneous expected excess return to holding a bond with remaining maturity τ. From (10), (15), and (16), in the canonical form, this can be written as

$$e_{\tau,t} = -B(\tau)' \left[\begin{pmatrix} 0_m \\ \lambda_{1(m+1)} \\ . \\ . \\ . \\ \lambda_{1n} \end{pmatrix} + \left(\begin{pmatrix} M_{m\times m}^a & 0_{m\times(n-m)} \\ M_{(n-m)\times m}^b & 0_{(n-m)\times(n-m)} \end{pmatrix} + \begin{pmatrix} 0_{m\times n} \\ L_{(n-m)\times n} \end{pmatrix} \right) X_t \right]. \tag{17}$$

In (17), 0_m is an m-vector of zeros. The $0_{p\times q}$ matrices are defined similarly. The submatrix M^a is a diagonal matrix with the ith diagonal element equal to element i of λ_1. Row i of M^b is given by the first m elements of the vector $\lambda_{1(m+i)}\beta_{m+i}'$. The submatrix L consists of rows $(m + 1)$ through n of λ_2.

The additional flexibility of the essentially affine model in fitting time variation in expected excess returns to bonds is captured by the matrix L. In a completely affine setup, L is a zero matrix. Therefore, any elements of X_t that do not affect the instantaneous volatility of X_t (i.e., elements $m + 1, \ldots, n$) also cannot affect instantaneous expected excess returns to bonds. When L is nonzero, any such elements of X_t can affect expected excess returns. In addition, L provides a mechanism for all other elements in X_t to affect expected returns through a channel other than M^a or M^b.

If all elements of X_t affect the instantaneous volatility (i.e., a correlated multifactor CIR model, or what Dai and Singleton (2000) call an $A_n(N)$ model), there is no L matrix. Therefore, the essentially affine model generalizes the completely affine model only when there is at least one element in X_t that does not affect the instantaneous volatility of X_t.

II. The Intuition Behind the Failure of Completely Affine Models

A successful model of the term structure should be consistent with the variety of term-structure shapes observed in the data. In other words, given the model's parameters, for each observed shape, there should be a valid state vector X_t that can generate it. In addition, the model should reproduce the empirically observed patterns in expected returns to bonds, or equivalently, produce forecasts of future yields that subsume the forecasting information in the slope of the term structure. This section explains that completely affine models fit to the historical behavior of Treasury yields will not simultaneously achieve both of these goals.

For our purposes, the key features of the excess returns to bonds are that they are, on average, small, and exhibit substantial predictable variation. Recall from Section I that $e_{\tau,t}$ denotes the instantaneous expected excess return to a bond with maturity τ. Although we do not observe instantaneous returns, the evidence in Table I suggests that the ratio $E(e_{\tau,t})/\sqrt{\mathrm{Var}(e_{\tau,t})}$ is small—well below one—for all τ. (This ratio is the inverse of the coefficient of variation for $e_{\tau,t}$.)

We will see below that completely affine models can be parameterized to produce low values of $E(e_{\tau,t})/\sqrt{\mathrm{Var}(e_{\tau,t})}$ for all τ. However, completely affine models can fit this behavior only by giving up the ability to fit a wide range of term-structure shapes. Conversely, they can be parameterized to fit observed term-structure shapes, but not the behavior of expected excess returns. The intuition underlying this result is best seen in two steps. We first examine the behavior of one-factor completely affine models. Then we will see that the important properties of one-factor models carry over to multi-factor models.

A. One-factor Models

The intuition in a completely affine one-factor model is straightforward. Expected instantaneous excess bond returns, $e_{\tau,t}$, are proportional to the factor's variance; hence they are bounded by zero. For a random variable that is bounded by zero to have a standard deviation substantially larger than its mean, it must be highly skewed. This high skewness is a tight restriction on the admissible values of $e_{\tau,t}$, and thus a tight restriction on the admissible values of the factor.

To see this clearly, we work through the math. Our first goal is to reproduce the stylized fact that $E(e_{\tau,t})/\sqrt{\mathrm{Var}(e_{\tau,t})}$ is small. We restrict our attention to a non-Gaussian model, because in a completely affine Gaussian model $\mathrm{Var}(e_{\tau,t}) = 0$. The model is

$$r_t = \delta_0 + x_t,$$

$$dx_t = k(\theta - x_t)dt + \sigma\sqrt{x_t}\,dW_t,$$

$$\Lambda_t = \lambda_1\sqrt{x_t}.$$

From (7), the instantaneous expected excess return to a τ-maturity bond is

$$e_{\tau,t} = -B(\tau)\sigma\lambda_1 x_t.$$

Therefore, the inverse of the $e_{\tau,t}$'s coefficient of variation is

$$\frac{E(e_{\tau,t})}{\sqrt{\text{Var}(e_{\tau,t})}} = \frac{E(x_t)}{\sqrt{\text{Var}(x_t)}} = \frac{\theta}{\sqrt{\text{Var}(x_t)}}. \tag{18}$$

Equation (18) implicitly imposes $B(\tau)\sigma\lambda_1 < 0$, which is the condition that mean excess bond returns are positive. We set $E(e_{\tau,t})/\sqrt{\text{Var}(e_{\tau,t})} = 0.3$, which is a typical ratio for predictable excess returns in Table I. We set the unconditional mean and standard deviation of the instantaneous interest rate to 5.5 percent and 2.9 percent, respectively. These values correspond to the moments of the three-month bill yield over 1952 through 1998. In this model, $\text{Var}(r_t) = \text{Var}(x_t)$. Plugging the standard deviation into (18) produces $\theta = 0.87$ percent. Therefore $\delta_0 = 4.63$ percent.

The requirement that the mean of x_t is small relative to its standard deviation gives the model little flexibility in producing short-term interest rates that are below average. The instantaneous interest rate r_t cannot be less than δ_0, or 4.63 percent. But over the period 1952 through 1998, three-month yields have ranged from 0.6 to 16 percent. Put differently, the model's parameters and the observed variation in short-term interest rates over this period imply a range of x_t from -4.0 to 11.4; the implied x_t is negative in more than 40 percent of the monthly observations. If we parameterized the model to be consistent with the observed distribution of short-term interest rates (i.e., nonnegative implied x_t), we would require $\theta > 4.9$ percent. But then the ratio $E(e_{\tau,t})/\sqrt{\text{Var}(e_{\tau,t})}$ would exceed 1.6. We can parameterize the model to fit either the expected excess returns or observed term structures, but not both.

We can also think about this model's restriction on the behavior of interest rates in terms of skewness of expected excess returns. In order to produce a small $E(e_{\tau,t})/\sqrt{\text{Var}(e_{\tau,t})}$, the model will generate expected excess returns that are always positive, usually very close to zero, and occasionally well above zero. But as noted in Section I, observed expected excess returns are not so positively skewed; they range from positive to negative.

B. Multifactor Models

Multifactor models are better at fitting the behavior of expected excess bond returns. For example, it is simple to generate a near-zero value of $E(e_{\tau,t})/\sqrt{\text{Var}(e_{\tau,t})}$ for a specified maturity, while retaining substantial flexibility in fitting term structure shapes. All that is required is prices of risk (elements of Λ_t) with different signs. If one element of Λ_t is positive and another negative, then at some maturity τ, the factor loadings will weight these prices of risk such that $E(e_{\tau,t}) = 0$ and $\text{Var}(e_{\tau,t}) > 0$.

However, completely affine models will not produce near-zero values of $E(e_{\tau,t})/\sqrt{\text{Var}(e_{\tau,t})}$ for *all* maturities while also allowing for a wide variety of term-structure shapes. To slightly oversimplify, the intuition is that long-maturity bond yields are affected by a single factor—the factor with the greatest persistence under the equivalent martingale measure. Thus, we can use the earlier intuition developed for one-factor models to conclude that multifactor models cannot reproduce the observed behavior of long-maturity yields.

The reason why only a single factor will affect long-maturity bond yields is practical, not theoretical. There are a variety of types of shocks that affect the term structure (e.g., level, slope, twist). Multifactor models capture this variety through factors that die out at different rates under the equivalent martingale measure. In principle, we could construct, say, a two-factor model where both factors affected long-bond yields. The only requirement is that the factors have the same speed of mean reversion. But by doing so, we sacrifice the major advantage of multifactor models—the ability to fit different kinds of shocks to the term structure. Thus, such a model would produce a poor fit of term structure data relative to a model in which each factor had its own speed of mean reversion.

The failure of completely affine models to fit the empirical behavior of bonds can be seen in the parameter estimates of three-factor completely affine models in Dai and Singleton (2000). They use U.S. dollar interest rate swap yields to estimate the same general three-factor completely affine models that are estimated here. I use their data and the parameters of their preferred model to produce implied time series of the state vector and expected excess returns to bonds. The results of this exercise, which are not reported in any table, indicate the model captures the combination of low mean and high volatility for expected excess returns. However, in over one-quarter of the observations, the implied value of the state vector violates a nonnegativity constraint. The violations tend to occur when the long end of the term structure is well below its average. Thus the results in Dai and Singleton support the conclusion that completely affine models do not simultaneously fit the behavior of expected excess returns to bonds and the observed term structure shapes.

III. Estimation of Essentially Affine Models

A. *Three-factor Affine Models*

All of the affine models I estimate have three underlying factors ($n = 3$). Litterman and Scheinkman (1991) find that three factors explain the vast majority of Treasury bond price movements. This is fortunate, because general three-factor affine models are already computationally difficult to estimate owing to the number of parameters. Adding another factor would make this investigation impractical. Seven models are estimated: four completely affine models and three essentially affine models. A completely affine model

is estimated for each possible number of factors that do not affect the instantaneous volatility of X_t (from three to zero). Using the canonical form of Section I.G, the estimated models are $A_0(3)$ through $A_3(3)$. The other models that are estimated are the essentially affine generalizations of $A_0(3), A_1(3)$, and $A_2(3)$. (Recall $A_3(3)$ has no essentially affine generalization.)

The estimated models share the following expressions for the instantaneous interest rate, the physical dynamics of X_t, and the price of risk vector:

$$r_t = \delta_0 + \delta_1 X_{t,1} + \delta_2 X_{t,2} + \delta_3 X_{t,3}, \tag{19a}$$

$$d\begin{pmatrix} X_{t,1} \\ X_{t,2} \\ X_{t,3} \end{pmatrix} = \left[\begin{pmatrix} (K\theta)_1 \\ (K\theta)_2 \\ (K\theta)_3 \end{pmatrix} - \begin{pmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{pmatrix} \begin{pmatrix} X_{t,1} \\ X_{t,2} \\ X_{t,3} \end{pmatrix} \right] dt + S_t \, dW_t, \tag{19b}$$

$$S_{t(ii)} = \sqrt{\alpha_i + (\beta_{i1} \quad \beta_{i2} \quad \beta_{i3}) X_t}, \tag{19c}$$

$$\Lambda_t = S_t \begin{pmatrix} \lambda_{11} \\ \lambda_{12} \\ \lambda_{13} \end{pmatrix} + S_t^- \begin{pmatrix} \lambda_{2(11)} & \lambda_{2(12)} & \lambda_{2(13)} \\ \lambda_{2(21)} & \lambda_{2(22)} & \lambda_{2(23)} \\ \lambda_{2(31)} & \lambda_{2(32)} & \lambda_{2(33)} \end{pmatrix} X_t. \tag{19d}$$

Depending on the model, restrictions are placed on the parameters in (19a) through (19d).

B. The Data

I use month-end yields on zero-coupon Treasury bonds, interpolated from coupon bonds using the method in McCulloch and Kwon (1993). Their sample, which ends in February 1991, is extended by Bliss (1997). The entire data set covers the period January 1952 through December 1998.[4] The range of maturities is three months to ten years.

To perform both in-sample and out-of-sample tests, I estimate term-structure models using data from 1952 through 1994. The final four years of data are reserved for constructing out-of-sample forecast errors.

C. The Estimation Technique

I estimate these models using quasi maximum likelihood (QML), which is particularly easy to implement with completely and essentially affine models. Although QML does not use all of the information in the probability density of yields, it fully exploits the information in the first and second

---

[4] Bliss and McCulloch-Kwon (1993) use slightly different filtering procedures; thus the yields they report over periods of overlapping data do not match exactly. I use the yields in McCulloch and Kwon over their entire sample period and use the Bliss (1997) data after February 1991.

conditional moments of the term structure. Thus, QML will capture the tension in affine models between fitting conditional means and conditional variances.

Another advantage of QML (which it shares with maximum likelihood and related techniques) is that there is a positive probability that the estimated model could actually generate the observed time series of term structures. This is an important concern in estimating affine term structure models. As the discussion in Section II highlighted, there is a tradeoff between fitting the coefficients of variation for expected excess bond returns and fitting the term structure shapes in the data. A model estimated with QML will guarantee that the time-t state vector implied by time-t yields is in the state vector's admissible space (to avoid a likelihood of zero). By contrast, consider techniques such as Efficient Method of Moments (EMM) that compare sample moments from the data with population moments simulated from the model. These techniques do not require that the estimated term structure model be sufficiently flexible to reproduce the term structure shapes in the data. The parameters of the model in Dai and Singleton (2000), which were estimated with EMM, illustrate this point.

I implement QML following Fisher and Gilles (1996), which contains further details. I assume that at each month-end t, $t = 1,\ldots,T$, yields on n bonds are measured without error. These bonds have fixed times to maturity $\tau_1,\ldots,\tau_n$. Yields on k other bonds are assumed to be measured with serially uncorrelated, mean-zero measurement error.

Stack the perfectly observed yields in the vector Y_t and the imperfectly observed yields in the vector $\tilde{Y}_t$. Denote the parameter vector by Θ. Given Θ, Y_t can be inverted using (4) to form an implied state vector $\hat{X}_t$, as in (20).

$$\hat{X}_t = H_1^{-1}(Y_t - H_0).   \tag{20}$$

In (20), H_0 is an n-vector with element i given by $A(\tau_i)/\tau_i$, and H_1 is an $n \times n$ matrix with row i given by $B(\tau_i)/\tau_i$. The candidate parameter vector is required to be consistent with Y_t. This is enforced by requiring $\hat{X}_t$ to be in the admissible space for X_t, which is equivalent to requiring that the diagonal elements of S_t in (19c) be real.

Given $\hat{X}_t$, implied yields for the other k bonds can be calculated. Stack them in $\hat{\tilde{Y}}_t$. The measurement error is $\epsilon_t = \hat{\tilde{Y}}_t - \tilde{Y}_t$. The variance–covariance matrix of the measurement error is assumed to have the following time-invariant Cholesky decomposition:

$$E(\epsilon_t \epsilon_t') = CC'.   \tag{21}$$

To compute the quasi-likelihood value, assume that the one-period-ahead conditional distribution of the state variables is multivariate normal and equal to

$$f_X(X_{t+1}|X_t).$$

The mean and variance–covariance matrix of X_{t+1} are known (see the Appendix); thus, $f_X(X_{t+1}|X_t)$ is known. Then the distribution of Y_{t+1} conditional on Y_t is

$$f_Y(Y_{t+1}|Y_t) = \frac{1}{|\det(H_1)|}\, f_X(\hat{X}_{t+1}|\hat{X}_t).$$

Also, assume that the measurement error is jointly normally distributed with distribution $f_\epsilon(\epsilon_t)$. The log likelihood of observation t is then

$$l_t(\Theta) = \log f_Y(Y_t|Y_{t-1}) + \log f_\epsilon(\epsilon_t). \tag{22}$$

Stationarity is imposed by requiring that the eigenvalues of K are positive, allowing $f_Y(Y_1|Y_0)$ to be set equal to the unconditional distribution of Y_t. The estimated parameter vector Θ^* is chosen to solve

$$\max_{\Theta} L(\Theta) = \sum_{t=1}^{T} l_t(\Theta).$$

In the estimation that follows, I assume that the bonds with no measurement error are those with maturities of 6 months, 2 years, and 10 years. This choice was motivated by the desire to span as much of the term structure as possible without assuming that the 3-month yield, which exhibits some idiosyncratic behavior, is observed without error. The bonds with measurement error fill in the gaps in this term structure, with maturities of 3 months, 1 year, and 5 years.

D. The Maximization Technique

The QML functions for these models have a large number of local maxima. The most important reason for this is the lack of structure placed on the feedback matrix K. Similar QML values can be produced by very different interactions among the elements of the state vector. Another reason is that the feasible parameter space is not convex for any model with nonconstant volatilities. A feasible parameter vector satisfies the requirement that the diagonal elements of S_t are real for all t. Because I use the canonical form of Section I.G, this requirement is satisfied when $\hat{X}_{t,i} \geq 0$ for $i \leq m$. (Recall that m is the number of state variables that affect the instantaneous volatility of X_t.) Therefore, the requirement imposes $m \times T$ restrictions on the parameter vector. The restrictions are nonlinear functions of the parameters and the data. These problems led to the following maximization technique.

Step 1. Randomly generate parameters from a multivariate normal distribution with a diagonal variance-covariance matrix. The means and variances were set to 'plausible' values.

Step 2. Use (20) to calculate $\hat{X}_t$ for all t.

Step 3. If the parameter vector is not feasible, return to step 1; otherwise proceed.

Step 4. Use Simplex to determine the parameter vector that maximizes the QML value.

Step 5. Using the final parameter vector from Step 4 as a starting point, use NPSOL to make any final improvements in the QML value.

This procedure is repeated until Step 5 is reached 1,000 times. For most of the models, there was little improvement in the QML value after the first few hundred iterations.

E. Specification Tests

These specification tests use the fact that QML estimation can be viewed as a GMM estimator. The moments are the first derivatives of the quasi log likelihood function with respect to the parameter vector, resulting in an exactly identified model. By imposing overidentifying moment conditions we test the adequacy of the model.

E.1. Tests of Nested Models

For $m < n$, the completely affine model $A_m(n)$ is nested in a corresponding essentially affine model. The essentially affine version has an additional $n(n - m)$ free parameters corresponding to the bottom $n - m$ rows in the matrix λ_2. We can test the null hypothesis that these free parameters are all equal to zero, using the GMM version of a likelihood ratio test. A textbook discussion is in Greene (1997). Define the column vector $h_t(\Theta)$ as the derivative of (22) with respect to the parameter vector Θ, and define $h(\Theta)$ as the mean of these T vectors. Define the (inverse of) the weighting matrix W_t as

$$W_t^{-1}(\Theta) = (1/T) \sum_{t=1}^{T} h_t(\Theta) h_t(\Theta)'. \tag{23}$$

Denote the parameter vector for the essentially affine model estimated by QML as Θ_0^*. The parameter vector Θ_1 is an alternative vector that imposes the completely affine restriction on λ_2. Choose it to solve

$$q = \min_{\Theta_1} Th(\Theta_1)' W_t(\Theta_0^*) h(\Theta_1). \tag{24}$$

The results of Hansen (1982) imply that under the null hypothesis, q is distributed as $\chi^2((n - m)n)$. Similar tests can be used to evaluate other parametric restrictions on the estimated models. These other tests are discussed in more detail in Section IV.

E.2. Testing the Covariance Between Forecast Errors and the Term Structure Slope

This test asks whether the yield forecasts produced by the estimated models include the information in the slope of the term structure. Given a parameter vector Θ associated with a particular model, the implied state vector $\hat{X}_{t-\Delta}$ is given by inverting yields observed at time $t - \Delta$. The Δ-period-ahead conditional mean $E(X_t | \hat{X}_{t-\Delta})$ can then be constructed. Given this expected state vector, expected Δ-period-ahead bond yields and associated forecast errors can also be constructed. We need some notation for forecast errors. Denote by e_{t,Δ,τ_i} the forecast error realized at time t for a τ_i-maturity bond, where the forecast is made at time $t - \Delta$. The forecast errors for v bonds of different maturities are stacked in the vector $e_{t,\Delta}$.

$$e_{t,\Delta} \equiv (e_{t,\Delta,\tau_1} \quad e_{t,\Delta,\tau_2} \quad \cdots \quad e_{t,\Delta,\tau_v})'.$$

If an estimated term structure model does not make systematic forecast errors, forecasts of time-t yields made at time $t - \Delta$ should have forecast errors uncorrelated with any variable known at time $t - \Delta$. This motivates the specification test. Denote the slope of the yield curve at time $t - \Delta$ by $s_{t-\Delta}$. If the model is correctly specified,

$$E\left[(e_{t,\Delta} - \overline{e_{t,\Delta}})(s_{t-\Delta} - \overline{s_{t-\Delta}})\right] = 0. \tag{25}$$

Equation (25) contains v moments that can be used as overidentifying restrictions in GMM estimation of an affine model. The other moments are standard QML moments, which are the derivatives of (22) with respect to each element of the parameter vector. The weighting matrix is calculated at the QML parameter estimates, which are consistent under the null hypothesis that the model is correctly specified. Then an analogue to q in (24) is calculated. Again from the results of Hansen (1982), this value is distributed as $\chi^2(v)$ under the null hypothesis. The use of overlapping observations in these moment conditions produces sample moments that exhibit serial correlation. I therefore construct the weighting matrix following Newey and West (1987).

To implement this test, I set $\Delta = 1/2$, so that six-month-ahead forecasts are examined. This horizon was chosen arbitrarily. A cursory investigation of other forecast horizons indicated that the results were insensitive to this choice. I used eight lags in the Newey–West calculation of the weighting matrix; experimentation with similar lag lengths did not materially affect the results. I set $v = 3$, and formed forecasts for maturities of 6 months, 2 years, and 10 years. (These are the same maturities that are assumed to have no measurement error.) The slope of the term structure is measured by the difference between the 5-year bond yield and the 3-month bond yield. The first six observations are dropped to account for the length of the forecast horizon.

Statistical Comparison of Estimated Models

Three-factor affine models are estimated with quasi maximum likelihood (QML). The data are month-end yields on zero-coupon bonds with maturities between three months and 10 years, from January 1952 through December 1994. The models differ in the number of factors m that affect the instantaneous variance of yields and in the flexibility of the price of risk parameterization. Essentially affine models allow the price of risk to vary independently from the instantaneous variance of yields, while completely affine models do not. "Unrestricted" models impose no restrictions on the parameters other than those required by no-arbitrage. "Preferred" models drop parameters that contribute little to the models' QML values.

Two specification tests are reported. The first is of the null hypothesis that the model's parameter restrictions are true. For "unrestricted" models, the test compares completely affine models to their more general essentially affine counterparts. For "preferred" models, the test compares the preferred model to its unrestricted counterpart. The second tests the null hypothesis that the six-month-ahead yield forecast errors for bonds of three different maturities are uncorrelated with the slope of the term structure at the time the forecasts are made. Under the null, the test statistics are distributed as χ^2(number of param restrictions) and $\chi^2(3)$, respectively.

Model Type	m	Number of Free Params	QML value	First Test Stat (p-value)	Second Test Stat (p-value)
Unrestricted					
Completely	0	19	15,171.94	62.689 (0.000)	12.297 (0.006)
Completely	1	23	15,380.31	26.133 (0.000)	18.521 (0.000)
Completely	2	24	15,395.74	0.860 (0.835)	9.938 (0.019)
Completely	3	25	15,396.34		33.482 (0.000)
Essentially	0	28	15,196.45		2.385 (0.596)
Essentially	1	29	15,392.47		16.639 (0.001)
Essentially	2	27	15,396.04		11.381 (0.010)
Preferred					
Completely	2	19	15,393.55	1.238 (0.941)	10.406 (0.015)
Essentially	0	21	15,190.68	3.443 (0.841)	1.449 (0.694)
Essentially	1	22	15,387.91	17.882 (0.013)	14.361 (0.002)

IV. Results

A. Overview

Table II reports the QML values for each estimated model. Results for 10 models are shown. The first 7 model specifications are labeled "unrestricted." This means that the only parameter restrictions imposed are those im-

plied by the canonical form. These restrictions are either normalizations or requirements of no-arbitrage. To both limit the danger of overfitting and to aid in the interpretation of the parameter estimates, more parsimonious specifications are also estimated. They are discussed in Section IV.B.

There are three main points to take from Table II. First, models that are better able to produce time-varying volatilities have higher QML values. The $A_0(3)$ models (both completely and essentially affine), which have time-invariant yield volatilities, have the lowest QML values. As the number of factors that affect volatilities increases from zero through three, QML values monotonically rise.

Second, the additional flexibility offered by essentially affine models over completely affine models is important. The first specification test reveals that the completely affine $A_0(3)$ and $A_1(3)$ models are overwhelmingly rejected by their more general essentially affine counterparts. Recall that as m increases, the flexibility provided by essentially affine models diminishes. We see empirical evidence of this in decreasing values of the first test statistic as m increases, culminating in the lack of rejection of the completely affine $A_2(3)$ model relative to its essentially affine counterpart.

Third, only the essentially affine model with the greatest flexibility in producing time-varying risk premia can capture the forecasting power of the term-structure slope. According to the second specification test, the yield forecast errors produced by the purely Gaussian essentially affine model ($A_0(3)$) are uncorrelated with the slope of the term structure. For this model, all elements of λ_2 are free. When $m = 1$, the top row of λ_2 is set to zero. This additional restriction leads to a strong rejection by the second specification test. A similar rejection accompanies the $A_2(3)$ essentially affine model. None of the completely affine models pass this second specification test.

We can see from these results the trade-off between fitting conditional volatilities and producing accurate forecasts of future yields. The essentially affine model with the greatest forecasting power also has the least ability to fit conditional volatilities. An increase in m provides for greater flexibility in fitting conditional variances of yields but also provides for less flexibility (in an essentially affine model) in fitting expected excess returns to bonds. The QML values indicate that the overall goodness of fit of first and second moments is increased by giving up flexibility in forecasting to acquire flexibility in fitting conditional variances.

B. Parameter Estimates

To limit the size of the paper, I report more detailed information for only three of the models. They are the essentially affine $A_0(3)$, $A_1(3)$, and completely affine $A_2(3)$ models. The first is of particular interest because of its forecasting ability, the second illustrates the trade-off between forecasting ability and fitting conditional variances, while the third is the completely affine model that does the best at forecasting, as measured by the χ^2 statistic on the second specification test.

For each of these models, I estimate a more parsimonious specification by first computing the t-statistics for the unrestricted parameter estimates. I then set to zero all parameters for which the absolute t-statistics did not exceed one and reestimated the models.[5] This procedure eliminated five parameters from the completely affine $A_2(3)$ model and seven parameters from both the essentially affine $A_0(3)$ model and the essentially affine $A_1(3)$ model. For each preferred model, a joint test of the parameter restrictions is constructed using an analogue to (24). The test statistics and corresponding p-values are reported in the "First Test Stat" column.[6]

Parameter estimates for these preferred models are in Tables III through V. To conserve space, parameter estimates for the other models are not reported in the paper, and are available on request.

Table III reports the parameter estimates for the $A_0(3)$ essentially affine model. The canonical form imposes a lower triangular structure on K and imposes $\alpha = 1$, $\beta = 0$, $K\theta = 0$. Table IV reports the parameter estimates for the $A_1(3)$ essentially affine model. One feature of this table deserves highlighting. The parameter $(K\theta)_2$ is nonzero, but no standard error is reported. This is the result of two normalizations imposed on the model: $\theta_2 = \theta_3 = 0$. The normalizations are imposed by setting $(K\theta)_2$ and $(K\theta)_3$ to the necessary values given K. Other restrictions imposed in the canonical form are $\alpha_1 = k_{12} = k_{13} = 0, \alpha_2 = \alpha_3 = \beta_{11} = 1$, and $\beta_{ij} = 0$, $i \geq 1, j > 1$. Finally, Table V reports the parameter estimates for the $A_2(3)$ completely affine model. In the canonical form of the $A_2(3)$ model, $\alpha_1 = \alpha_2 = \beta_{33} = 0$, $\alpha_3 = \beta_{11} = \beta_{22} = 1$, $\beta_{ij} = 0$ for $i < 3$, $i \neq j$, and $\lambda_2 = 0$. The preferred specification sets $\beta_{31} = 0$ and $\beta_{32} = 1$, so that the second state variable drives the conditional volatilities of both the second and third state variables. The element $(K\theta)_3$ is nonzero with no standard error, because $\theta_3 = 0$ in the canonical model.

C. An Analysis of Forecast Errors

The estimated models, combined with month t bond yields, can be used to construct forecasts of month $t + i$ bond yields. Here we examine the accuracy of these forecasts, both in sample and out of sample. The in-sample period is January 1952 through December 1994. The out-of-sample period is January 1995 through December 1998. We focus on bonds with maturities of 6 months, 2 years, and 10 years, and forecast horizons of 3, 6, and 12 months. Forecast accuracy is measured by the root mean squared forecast error (RMSE). In-sample RMSEs are reported in Table VI and out-of-sample RMSEs are re-

---

[5] With the completely affine $A_2(3)$ model, the parameter β_{32} was set to one instead of zero.

[6] The test statistic for the essentially affine $A_1(3)$ model suggests a rejection of the preferred model in favor of the unrestricted model. However, the large test statistic appears to be a consequence of approximation errors in numerical computation of the derivative of the log-likelihood function with respect to k_{32}. The Numerical Recipes *dfridr* routine (a robust method for calculating derivatives and estimates of errors in the derivatives) reported large errors regardless of the initial stepsize. Because the estimate of parameter in the unrestricted model was nearly zero, and setting it to zero had a negligible effect on the QML likelihood function, I set it to zero in the preferred model.

Table III
Parameter Estimates for the Preferred Essentially Affine $A_0(3)$ Model

The model is defined in equation (19). With this version of the model, α is a vector of ones and both β and $K\theta$ are identically zero. The matrix C is the Cholesky decomposition $V = CC'$ of the variance–covariance matrix of the cross-sectional errors in fitting yields on bonds with maturities of three months, one year, and five years. Parameters are estimated with QML. Asymptotic standard errors are in parentheses.

	Constant Term		
δ_0	0.044		
	(0.025)		

	Index Number (i)		
Parameter	1	2	3
δ_i	0.01895	0.00790	0.00992
	(0.00223)	(0.00218)	(0.00051)
k_{1i}	0.564	0	0
	(0.047)		
k_{2i}	0	3.257	0
		(0.672)	
k_{3i}	−0.545	0	0.062
	(0.202)		(0.051)
λ_{1i}	−0.625	−0.235	−0.207
	(0.146)	(0.099)	(0.057)
$\lambda_{2(1i)}$	0	1.742	0
		(0.254)	
$\lambda_{2(2i)}$	0	−1.711	0
		(0.717)	
$\lambda_{2(3i)}$	0.648	0.297	−0.061
	(0.206)	(0.186)	(0.051)
C_{1i}	0.00227	0	0
	(0.00013)		
C_{2i}	−0.00050	0.00084	0
	(0.00007)	(0.00004)	
C_{3i}	0	−0.00017	0.00093
		(0.00006)	(0.00004)

ported in Table VIII. In Tables VII (in-sample) and IX (out-of-sample), forecast errors are regressed on the slope of the yield curve to determine whether the models' forecasts capture the forecasting power of the slope.

We need benchmarks to use in evaluating forecast accuracy. The simplest benchmark is a random walk. The month t yield on a τ-maturity bond is used as a forecast of the month $t + i$ yield on a τ-maturity bond. The RMSEs associated with this forecast method are reported in the "RW" columns of Tables VI and VIII. Note that the tables report different patterns in RMSEs across bonds. In the earlier period, yields were more volatile, with volatility declining with maturity. In the later period, yield volatility was higher at

The Journal of Finance

Parameter Estimates for the Preferred Essentially Affine $A_1(3)$ Model

The model is defined in equation (19). With this version of the model, $\alpha_1 = \beta_{12} = \beta_{13} = 0$, $\alpha_2 = \alpha_3 = \beta_{11} = 1$, and the first row of λ_2 is zero. The matrix C is the Cholesky decomposition $V = CC'$ of the variance–covariance matrix of the cross-sectional errors in fitting yields on bonds with maturities of three months, one year, and five years. Parameters are estimated with QML. Asymptotic standard errors are in parentheses.

	Constant Term		
δ_0	0.014		
	(0.005)		

	Index Number (i)		
Parameter	1	2	3
δ_i	0.00088	0.00118	0.00256
	(0.00021)	(0.00053)	(0.00124)
$(K\theta)_i$	0.155	−1.910	0
	(0.048)		
k_{1i}	0.031	0	0
	(0.020)		
k_{2i}	−0.383	0.594	5.340
	(0.235)	(0.053)	(3.833)
k_{3i}	0	0	2.832
			(0.490)
β_{2i}	10.269	0	0
	(9.96)		
β_{3i}	0.291	0	0
	(0.261)		
λ_{1i}	−0.042	−3.844	0
	(0.020)	(2.415)	
$\lambda_{2(2i)}$	39.334	0	5.259
	(53.816)		(3.647)
$\lambda_{2(3i)}$	0	0	−1.311
			(0.565)
C_{1i}	0.00227	0	0
	(0.00013)		
C_{2i}	−0.00049	0.00084	0
	(0.00007)	(0.00004)	
C_{3i}	0	−0.00016	0.00094
		(0.00006)	(0.00004)

long maturities than at short maturities. Thus, the out-of-sample period should provide a good test of the robustness of the estimated affine models.

A more sophisticated benchmark uses OLS regressions that predict future changes in yields with the current slope of the term structure. The regression is

$$Y_{\tau, t+i} - Y_{\tau, t} = b_0 + b_1(Y_{5yr, t} - Y_{3mo, t}) + e_{\tau, t+i}. \tag{26}$$

<div align="center">

Table V

**Parameter Estimates for the Preferred Completely
Affine $A_2(3)$ Model**

</div>

The model is defined in equation (19). With this version of the model, $\alpha_1 = \alpha_2 = 0$, $\alpha_3 = \beta_{11} = \beta_{22} = \beta_{32} = 1$, the remaining elements of the β matrix are zero, and λ_2 is a matrix of zeros. The matrix C is the Cholesky decomposition $V = CC'$ of the variance–covariance matrix of the cross-sectional errors in fitting yields on bonds with maturities of three months, one year, and five years. Parameters are estimated with QML. Asymptotic standard errors are in parentheses.

	Constant Term		
δ_0	0.018		
	(0.004)		

	Index Number (i)		
Parameter	1	2	3
δ_i	0.00066	0.00136	0.00598
	(0.00021)	(0.00050)	(0.00077)
$(K\theta)_i$	0	0.222	−2.299
		(0.103)	
k_{1i}	0.172	−0.295	0
	(0.064)	(0.056)	
k_{2i}	−0.197	0.406	0
	(0.066)	(0.059)	
k_{3i}	0.564	−1.669	1.721
	(0.279)	(0.234)	(0.176)
λ_{1i}	−0.042	0	−0.208
	(0.018)		(0.058)
C_{1i}	0.00227	0	0
	(0.00013)		
C_{2i}	−0.00049	0.00084	0
	(0.00007)	(0.00004)	
C_{3i}	0	−0.00017	0.00094
		(0.00006)	(0.00004)

The parameters of (26) are estimated using in-sample data. The equation is then used to construct forecasts and forecast errors for both the in-sample and out-of-sample periods. The resulting RMSEs are in the columns labeled "OLS" in Table VI and Table VIII. Although the in-sample RMSE for the regression is guaranteed to be no larger than the random walk RMSE, that is not true out of sample. Indeed, for eight of the nine combinations of maturity and forecast horizon, the out-of-sample OLS RMSE exceeds that of the random walk.

The in-sample parameter estimates from (26) are reported in Table VII in the column labeled "RW." This may seem like a misprint (why aren't they labeled "OLS"?), but recall that Table VII reports the parameter estimates of regressions of forecast errors on the month t slope of the yield curve. With the random-walk method of forecasting, the regression examined in Table VII is identical to the regression used to produce OLS forecasts. The results

<div align="center">

Table VI

Comparison of In-sample Forecasting Performance

</div>

This table reports root mean squared errors (RMSE) for month t forecasts of month $t + i$ bond yields. Eight different forecast methods are compared. The column labeled "RW" (random walk) uses month t yields as forecasts of future yields. The column labeled "OLS" uses a univariate OLS regression to form forecasts, where the dependent variable is change in the yield from t to $t + i$ and the regressor is the month t slope of the yield curve. The final six columns use either completely affine (C.A.) or essentially affine (E.A.) three-factor models to form forecasts. Preferred models are restricted versions of unrestricted models. The models differ in the number of factors j that are allowed to affect conditional volatility ($A_j(3)$).

The regression and affine models are estimated using data from January 1952 through December 1994, and the forecasts are produced over the same period (in-sample forecasts). The slope of the yield curve is the five-year zero-coupon yield less the three-month zero-coupon yield. Bond yields are measured in decimal form (i.e., 0.04 corresponds to four percent/year).

Bond Maturity	Forecast Horizon	RW	OLS	Unrestricted			Preferred		
				C. A.	E. A.		C. A.	E. A.	
				$A_2(3)$	$A_0(3)$	$A_1(3)$	$A_2(3)$	$A_0(3)$	$A_1(3)$
6 months	3	1.023	1.020	1.045	1.009	1.019	1.048	1.009	1.019
2 years	3	0.871	0.869	0.880	0.837	0.847	0.883	0.837	0.853
10 years	3	0.549	0.532	0.554	0.526	0.543	0.554	0.528	0.547
6 months	6	1.376	1.370	1.418	1.342	1.367	1.427	1.345	1.368
2 years	6	1.154	1.149	1.173	1.091	1.121	1.181	1.089	1.133
10 years	6	0.760	0.722	0.774	0.711	0.756	0.772	0.713	0.764
6 months	12	1.803	1.797	1.843	1.731	1.798	1.868	1.742	1.798
2 years	12	1.541	1.529	1.566	1.450	1.527	1.583	1.445	1.544
10 years	12	1.109	1.018	1.137	1.011	1.121	1.131	1.009	1.133

document that short-maturity yields tend to rise and long-maturity yields tend to fall when the slope is steeper than average, although the statistical evidence at the short end is weak. These results correspond to the standard violations of the expectations hypothesis of interest rates.

This violation is also apparent in the behavior of bond yields in the out-of-sample period. The "RW" column in Table IX reports the results of estimating (26) from January 1995 through December 1998. The point estimates are typically more negative than their counterparts in Table VII, although the t-statistics are smaller owing to fewer observations.

The final six columns in Table VI through Table IX examine the forecasting ability of various affine models. The results document that the completely affine $A_2(3)$ model is a failure at forecasting future interest rates. Table VI reports that in sample, both the unrestricted and preferred specifications produce forecasts that are worse than those produced by the assumption that yields follow random walks. This unimpressive performance is mirrored by the performance of the other completely affine models examined in this paper. For every estimated model, the assumption that yields follow a random walk produces superior in-sample forecasts for each of these

The Relation Between In-sample Forecast Errors and the Yield-curve Slope

Various models are used to produce month t forecasts of month $t + i$ bond yields and the corresponding forecast errors are constructed. This table reports parameter estimates from regressions of forecast errors on the month t slope of the yield curve. Six forecast methods are compared. The column labeled "RW" (random walk) uses month t yields as forecasts of future yields. Therefore, the forecast error regression is simply a regression of changes in bond yields from t to $t + i$ on the month t yield-curve slope. The final six columns use either completely affine (C.A.) or essentially affine (E.A.) three-factor models to form forecasts. Preferred models are restricted versions of unrestricted models. The models differ in the number of factors j that are allowed to affect conditional volatility ($A_j(3)$).

The regression and affine models are estimated using data from January 1952 through December 1994, and the forecasts are produced over the same period (in-sample forecasts). The slope of the yield curve is the five-year zero-coupon yield less the three-month zero-coupon yield. Asymptotic t-statistics, in parentheses, are adjusted for generalized heteroskedasticity and moving average residuals.

| Bond Maturity | Forecast Horizon | RW | Unrestricted | | | Preferred | | |
| | | | C. A. | E. A. | | C. A. | E. A. | |
			$A_2(3)$	$A_0(3)$	$A_1(3)$	$A_2(3)$	$A_0(3)$	$A_1(3)$
6 months	3	0.072	−0.182	−0.041	−0.135	−0.182	0.019	−0.124
		(0.73)	(−1.84)	(−0.42)	(−1.39)	(−1.83)	(0.19)	(−1.27)
2 years	3	−0.043	−0.182	−0.043	−0.134	−0.183	0.013	−0.129
		(−0.52)	(−2.22)	(−0.54)	(−1.69)	(−2.22)	(0.16)	(−1.63)
10 years	3	−0.125	−0.159	−0.027	−0.141	−0.158	−0.018	−0.140
		(−2.72)	(−3.50)	(−0.61)	(−3.19)	(−3.49)	(−0.39)	(−3.15)
6 months	6	0.118	−0.324	−0.085	−0.252	−0.326	0.015	−0.233
		(0.91)	(−2.55)	(−0.69)	(−2.03)	(−2.53)	(0.12)	(−1.88)
2 years	6	−0.082	−0.326	−0.091	−0.261	−0.330	−0.003	−0.249
		(−0.76)	(−3.17)	(−0.91)	(−2.60)	(−3.17)	(−0.03)	(−2.51)
10 years	6	−0.220	−0.280	−0.049	−0.255	−0.280	−0.031	−0.252
		(−3.45)	(−4.48)	(−0.78)	(−4.14)	(−4.46)	(−0.50)	(−4.09)
6 months	12	0.129	−0.567	−0.208	−0.484	−0.575	−0.058	−0.453
		(0.70)	(−3.30)	(−1.21)	(−2.74)	(−3.30)	(−0.33)	(−2.60)
2 years	12	−0.158	−0.551	−0.191	−0.486	−0.560	−0.069	−0.462
		(−1.06)	(−3.86)	(−1.32)	(−3.26)	(−3.86)	(−0.48)	(−3.15)
10 years	12	−0.410	−0.506	−0.135	−0.480	−0.507	−0.101	−0.472
		(−3.62)	(−4.53)	(−1.22)	(−4.24)	(−4.52)	(−0.92)	(−4.19)

maturities and forecast horizons. (These additional results are not reported in any table.)

The regressions reported in Table VII show that the forecast errors of the completely affine $A_2(3)$ model are strongly negatively correlated with the slope of the term structure. The parameter estimates are more negative than are the corresponding parameter estimates in the random walk case. The model completely misses the forecasting information in the slope of the term structure. When the term structure is more steeply sloped than usual, the

<div align="center">

Table VIII

Comparison of Out-of-sample Forecasting Performance

</div>

This table reports root mean squared errors (RMSE) for month t forecasts of month $t + i$ bond yields. Eight different forecast methods are compared. The column labeled "RW" (random walk) uses month t yields as forecasts of future yields. The column labeled "OLS" uses a univariate OLS regression to form forecasts, where the dependent variable is change in the yield from t to $t + i$ and the regressor is the month t slope of the yield curve. The final six columns use either completely affine (C.A.) or essentially affine (E.A.) three-factor models to form forecasts. Preferred models are restricted versions of unrestricted models. The models differ in the number of factors j that are allowed to affect conditional volatility ($A_j(3)$).

The regression and affine models are estimated using data from January 1952 through December 1994, while the forecasts are produced over January 1995 through December 1998 (out-of-sample forecasts). For each bond there are $48 - i$ forecasts and associated errors. The slope of the yield curve is the five-year zero-coupon yield less the three-month zero-coupon yield. Bond yields are measured in decimal form (i.e., 0.04 corresponds to four percent/year).

| | | | | Unrestricted | | | Preferred | | |
| | | | | C. A. | E. A. | | C. A. | E. A. | |
Bond Maturity	Forecast Horizon	RW	OLS	$A_2(3)$	$A_0(3)$	$A_1(3)$	$A_2(3)$	$A_0(3)$	$A_1(3)$
6 months	3	0.298	0.298	0.325	0.281	0.288	0.350	0.281	0.284
2 years	3	0.499	0.511	0.501	0.454	0.458	0.523	0.457	0.450
10 years	3	0.484	0.498	0.476	0.460	0.457	0.485	0.469	0.453
6 months	6	0.400	0.413	0.483	0.373	0.399	0.548	0.365	0.385
2 years	6	0.652	0.675	0.656	0.565	0.576	0.711	0.566	0.560
10 years	6	0.669	0.693	0.647	0.623	0.616	0.669	0.636	0.606
6 months	12	0.484	0.523	0.621	0.434	0.488	0.778	0.421	0.455
2 years	12	0.762	0.787	0.759	0.608	0.635	0.879	0.600	0.606
10 years	12	0.815	0.829	0.764	0.724	0.719	0.811	0.738	0.698

OLS forecast is that long-maturity yields will fall, but the model forecasts that the yields will rise. Put differently, the model is consistent with the expectations hypothesis, and the observed bond yields are not.

This poor forecasting performance carries over to the out-of-sample period. Table VIII documents that the unrestricted specification produces forecasts that are inferior to random-walk forecasts in five of the nine combinations of maturity and forecast horizon. The preferred specification does even worse, producing inferior forecasts for seven of the nine combinations. The point estimates in Table IX confirm that the model's forecasts get the wrong sign of the relationship between the slope of the term structure and future changes in yields.

The essentially affine models produce dramatically better forecasts. The most successful forecasting model, both in sample and out of sample, is the essentially affine, completely Gaussian model. Table VI documents that within the sample, both the unrestricted and preferred $A_0(3)$ models outforecast the OLS regressions (and therefore also outforecast the random-walk assumption) for each combination of maturity and forecast horizon. Table VIII makes the same point out of sample. Moreover, these forecasts capture the predic-

The Relation Between Out-of-sample Forecast Errors and the Yield-curve Slope

Various models are used to produce month t forecasts of month $t + i$ bond yields and the corresponding forecast errors are constructed. This table reports parameter estimates from regressions of forecast errors on the month t slope of the yield curve. Six forecast methods are compared. The column labeled "RW" (random walk) uses month t yields as forecasts of future yields. Therefore, the forecast error regression is simply a regression of changes in bond yields from t to $t + i$ on the month t yield-curve slope. The final six columns use either completely affine (C.A.) or essentially affine (E.A.) three-factor models to form forecasts. Preferred models are restricted versions of unrestricted models; model parameters that add little to the model's QML value are set to zero. The models differ in the number of factors j that are allowed to affect conditional volatility ($A_j(3)$).

The regression and affine models are estimated using data from January 1952 through December 1994, while the forecasts are produced over January 1995 through December 1998 (out-of-sample forecasts). For each bond, there are $48 - i$ forecasts and associated errors. The slope of the yield curve is the five-year zero-coupon yield less the three-month zero-coupon yield. Asymptotic t-statistics, in parentheses, are adjusted for generalized heteroskedasticity and moving average residuals.

| Bond Maturity | Forecast Horizon | RW | Unrestricted | | | Preferred | | |
			C. A. $A_2(3)$	E. A. $A_0(3)$	E. A. $A_1(3)$	C. A. $A_2(3)$	E. A. $A_0(3)$	E. A. $A_1(3)$
6 months	3	0.121	−0.206	0.039	−0.074	−0.220	0.113	−0.064
		(1.04)	(−1.71)	(0.35)	(−0.66)	(−1.82)	(1.01)	(−0.57)
2 years	3	−0.151	−0.268	−0.002	−0.108	−0.284	0.056	−0.104
		(−0.76)	(−1.37)	(−0.01)	(−0.59)	(−1.45)	(0.30)	(−0.57)
10 years	3	−0.265	−0.280	−0.107	−0.220	−0.286	−0.112	−0.219
		(−1.42)	(−1.51)	(−0.59)	(−1.23)	(−1.55)	(−0.61)	(−1.22)
6 months	6	0.034	−0.497	−0.094	−0.272	−0.525	0.014	−0.258
		(0.20)	(−2.48)	(−0.56)	(−1.65)	(−2.59)	(0.09)	(−1.56)
2 years	6	−0.380	−0.560	−0.153	−0.324	−0.588	−0.071	−0.318
		(−1.11)	(−1.68)	(−0.52)	(−1.10)	(−1.75)	(−0.24)	(−1.08)
10 years	6	−0.552	−0.571	−0.305	−0.485	−0.583	−0.307	−0.483
		(−1.60)	(−1.68)	(−0.92)	(−1.49)	(−1.71)	(−0.91)	(−1.49)
6 months	12	−0.086	−0.825	−0.245	−0.482	−0.882	−0.116	−0.468
		(−0.35)	(−3.22)	(−1.09)	(−2.28)	(−3.39)	(−0.52)	(−2.19)
2 years	12	−0.844	−1.037	−0.500	−0.734	−1.088	−0.405	−0.727
		(−2.27)	(−3.00)	(−1.55)	(−2.39)	(−3.13)	(−1.24)	(−2.35)
10 years	12	−1.085	−1.083	−0.737	−0.977	−1.105	−0.730	−0.975
		(−3.75)	(−3.89)	(−2.58)	(−3.68)	(−3.95)	(−2.52)	(−3.66)

tive power of the term-structure slope. In Tables VII and IX, the only evidence for predictability of forecast errors is in the out-of-sample forecast errors for 10-year bonds at the 12-month horizon.

The essentially affine $A_1(3)$ model is not quite as successful as the Gaussian model at forecasting. From Table VI, we see that in-sample forecasts from both the unrestricted and preferred specifications are typically supe-

rior to random-walk forecasts, but outforecast OLS regressions for only half of the maturity/horizon combinations. Moreover, from Table VIII, the forecast errors are negatively correlated with the slope of the yield curve. The statistical strength of this negative correlation rises as both the bond's maturity and the forecasting horizon lengthen.

An examination of Table VIII indicates that this essentially affine model performs somewhat better out of sample. Forecasts from the preferred specification are superior to random-walk and OLS forecasts at all maturities and forecast horizons. Nonetheless, Table IX indicates that the model's out-of-sample forecast errors are negatively correlated with the slope of the yield curve. Thus, the model misses some of the explanatory power of the term-structure slope.

Overall, these results indicate that for the purposes of forecasting, completely affine models are largely useless. Even the simplest, most naive rule—a random walk—dominates the explanatory power of completely affine models. A corollary is that we should not use completely affine models to attempt to understand why the expectations hypothesis fails, because the models cannot reproduce this failure.[7] By contrast, forecasts from a purely Gaussian essentially affine model dominate naive forecasts.

D. *The Predictability of Excess Returns and Volatilities*

A few diagrams help shed light on the behavior of these competing models. Figure 1 is a graphical summary of the behavior of the preferred essentially affine $A_0(3)$ model. Panel A displays instantaneous effects that one-standard-deviation shocks to each factor have on the term structure of yields. The three shocks can be interpreted as a level shock (the long dashes), a slope shock (the solid line), and a twist (the short dashes). Panel B displays the (nonexistent) instantaneous effect of these shocks on yield variances.

Panel C displays the effect that these shocks have on bonds' instantaneous expected excess returns (over r_t). There are two distinct types of shocks to expected returns. The short dashes correspond to the twist shock in Panel A. This shock has a strong effect on instantaneous expected returns, but it is also very short-lived. (This latter fact cannot be seen in the panel.) Thus, this shock is responsible for high-frequency fluctuations in expected excess returns.

The other type of shock to expected excess returns corresponds to the slope shock in Panel A. It is more persistent (this also cannot be seen in the panel), and thus accounts for more persistent fluctuations in expected returns. The combined effects of these shocks on expected excess returns to 2-year bonds are displayed in Panel E. Panel F is the same plot for 10-year bonds. These latter panels show that expected excess returns fluctuate sharply and widely around zero. For example, the expected instantaneous excess return in Panel E has a mean of 1.25 percent and a standard deviation of 3.09 percent.

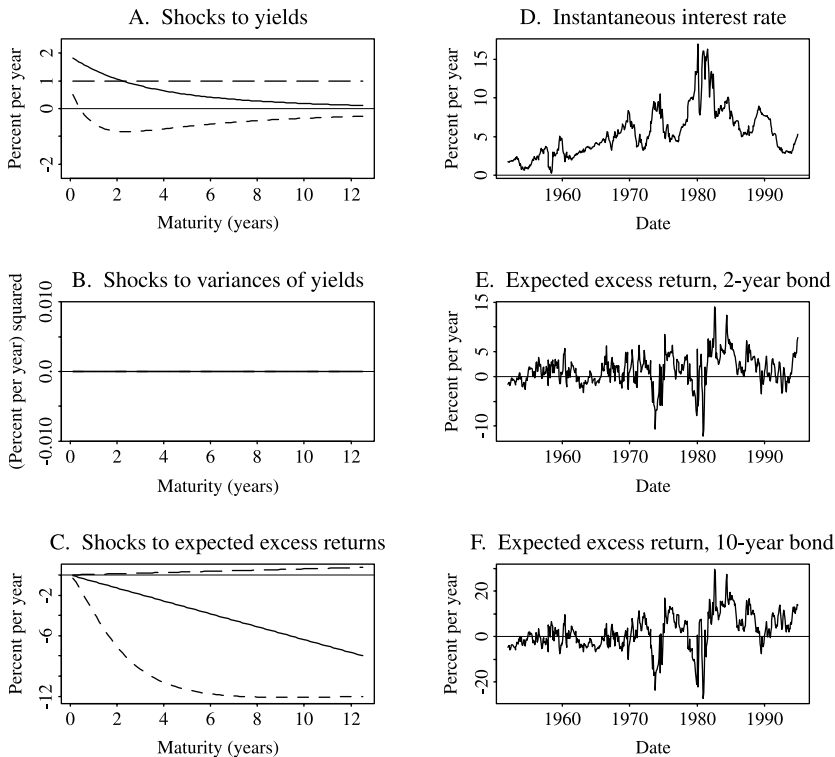[7] See Dai and Singleton (2001) for a related perspective.

Figure 1. Summary of the estimated essentially affine $A_0(3)$ model. Panels A through C display the instantaneous responses of yields, variances, and expected excess returns (over r_t) to one-standard-deviation shocks to each of the three factors. Panels D through F display fitted expected instantaneous returns over the sample period January 1952 through December 1994. Panel D is the instantaneous interest rate. Panels E and F are the instantaneous expected excess returns to the 2-year and 10-year bonds.

Because this model is so successful at forecasting future yields, it is worth a more careful examination. An intuitive way to interpret shocks to bond yields is to decompose the shocks into shocks to expected future short-term interest rates and shocks to expected excess returns. This decomposition is straightforward; thus I will not discuss it in detail here. Instead, I will simply summarize the results.

A positive level shock corresponds to an immediate, near-permanent increase in short-term interest rates. The half-life of the shock to short-term interest rates is more than 11 years. Because the shock does not substantially alter investors' required excess returns to bonds, short-maturity and long-maturity bond yields respond in the same way to this shock.

A positive slope shock corresponds to an immediate increase in short-term interest rates that lasts about as long as a business cycle. The half-life of the shock is four years. Because short-term interest rates are expected to decline over time, the shock lowers the slope of the term structure. The shock

also lowers expected excess returns to bonds by affecting the price of risk vector. We can see this in the parameters of λ_2 in Table III. An increase in the first factor (the slope factor) affects the price of risk of the third factor (the level factor) through element (3,1) of λ_2. This decrease in expected returns further decreases the slope of the term strucure because longer-maturity bond returns are more sensitive than shorter-maturity bond returns to level shocks, and thus to the price of risk of level shocks.

Twists are very similar to the "f_t" factor in the two-factor example discussed in Section II. A twist shock has basically no effect on current or future short-term interest rates. Instead, the shock changes investors' required excess returns to bonds by affecting the price of risk associated with the level and slope factors. The half-life of such a shock is less than three months. We can call this a "flight to quality" shock. Investors experience short-lived periods of unwillingness to hold risky Treasury instruments, thus driving expected excess bond returns higher.

Figure 2 contains information about the preferred essentially affine $A_1(3)$ model. Panel A displays a level shock, a slope shock, and a twist shock. The solid line is the level shock, and it affects the conditional variance of yields, as shown in Panel B. The long-dashed line is the twist shock, and in Panel C, we see its strong effect on expected excess returns. However, Panel C also indicates that the other two shocks have little effect on expected excess returns. The net effect is that in Panels E and F, the fluctuations in expected excess returns are less volatile than the fluctuations in the corresponding panels in Figure 1. For example, the expected instantaneous excess return in Panel E has a mean of 1.90 percent and a standard deviation of 1.85 percent.

Why does a shock to the slope affect expected excess returns in Figure 1 but not in Figure 2? The answer is that the channel that operates in the model underlying Figure 1 is unavailable in the model underlying Figure 2. Panel C in Figure 1 reflects a relationship between shocks to the slope and shocks to the price of risk of level shocks. These cross-factor relationships are more limited in the essentially affine $A_1(3)$ model. In the canonical form, the first factor drives conditional volatilities; thus its price of risk cannot be affected by any other factors. Figure 2 indicates that this first factor is the level factor; shocks to the slope cannot affect its price of risk. Therefore, this model produces poorer forecasts of future bond yields than does the essentially affine $A_0(3)$ model.

Figure 3 displays the same panels for the preferred completely affine $A_2(3)$ model. The model generates a richer pattern of time variation in volatilities than do the other two models. The cost of these more accurate measures of volatility is an inability to fit expected excess returns. Expected excess returns in Panels E and F are always positive, never large, and not volatile. For example, the expected instantaneous excess return in Panel E has a mean of 0.79 percent and a standard deviation of 0.41 percent. Moreover, these expected excess returns roughly track the instantaneous interest rate displayed in Panel D. Because higher short-term rates typically correspond to lower slopes, the figure indicates that expected excess returns move inversely with the slope of the yield curve, but this is counterfactual.
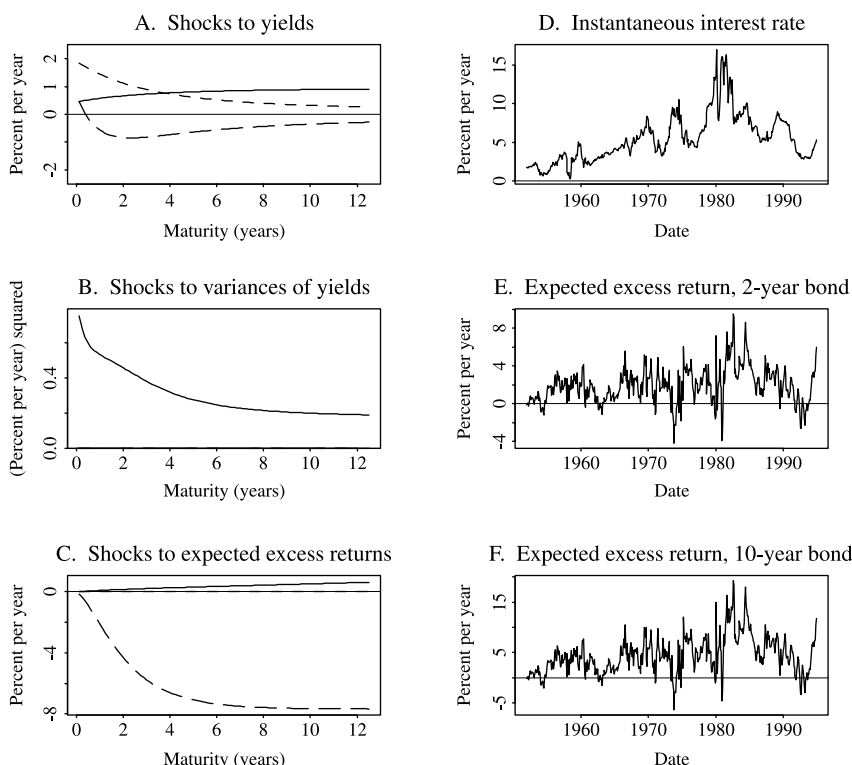
A. Shocks to yields

D. Instantaneous interest rate

B. Shocks to variances of yields

E. Expected excess return, 2-year bond

C. Shocks to expected excess returns

F. Expected excess return, 10-year bond

Figure 2. Summary of the estimated essentially affine $A_1(3)$ model. Panels A through C display the instantaneous responses of yields, variances, and expected excess returns (over r_t) to one-standard-deviation shocks to each of the three factors. Panels D through F display fitted expected instantaneous returns over the sample period January 1952 through December 1994. Panel D is the instantaneous interest rate. Panels E and F are the instantaneous expected excess returns to the 2-year and 10-year bonds.

The results discussed in this section indicate that the completely affine $A_2(3)$ model fails to reproduce the behavior of expected excess returns to Treasury bonds. The same conclusion holds for the other completely affine models estimated in this paper that are not discussed in detail here. The models systematically fail to capture the large fluctuations in expected excess returns to bonds. Essentially affine models do a better job of reproducing the behavior of expected excess returns, although the magnitude of the improvement is inversely related to the ability of the models to fit the time variation in conditional variances of yields.

V. Concluding Comments

Recent term structure research has concentrated on what I call completely affine models. This paper documents that completely affine models do not forecast future yields well over the nearly 50-year period examined here.
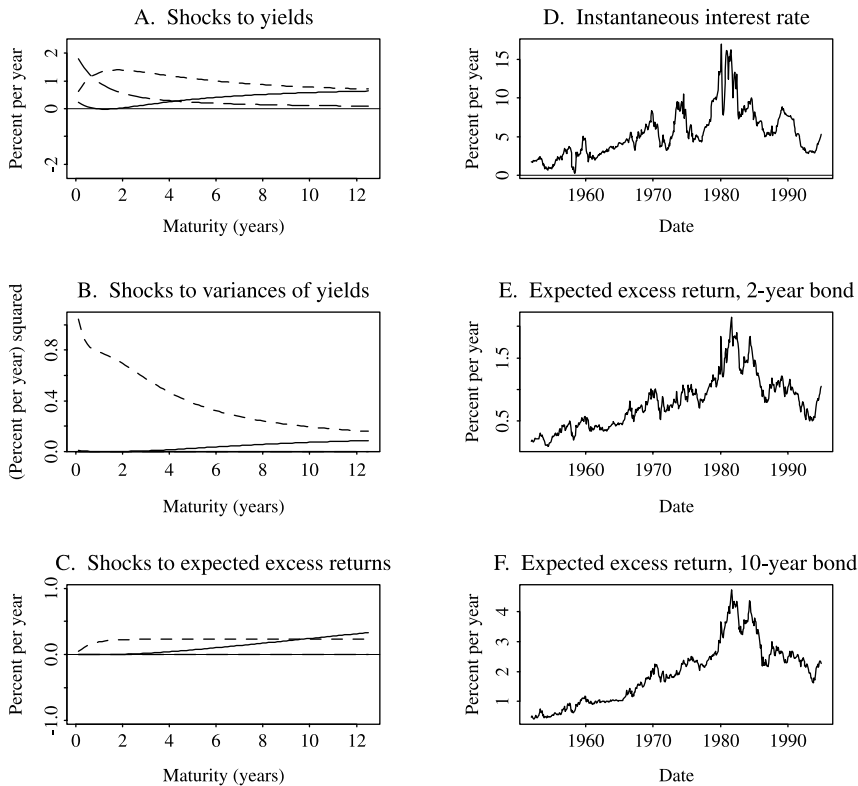
Figure 3. Summary of the estimated completely affine $A_2(3)$ model. Panels A through C display the instantaneous responses of yields, variances, and expected excess returns (over r_t) to one-standard-deviation shocks to each of the three factors. Panels D through F display fitted expected instantaneous returns over the sample period January 1952 through December 1994. Panel D is the instantaneous interest rate. Panels E and F are the instantaneous expected excess returns to the 2-year and 10-year bonds.

They consistently underestimate future returns to bonds when the term structure is more steeply sloped than usual; put differently, these models do not reproduce the well-known failure of the expectations hypothesis.

Essentially affine models generalize completely affine models. They allow greater flexibility in fitting variations in the price of interest rate risk over time, while retaining the affine time-series and cross-sectional properties of bond prices. One of the essentially affine models investigated in this paper—the pure Gaussian model—generates reasonable forecasts of future yields, in the sense that the predictive power of the term structure is subsumed within the model's forecasts.

The forecast accuracy of this Gaussian model allows us to properly interpret the usual level, slope, and twist yield-curve factors in terms of their predictions for future short-term interest rates and excess returns to longer-term bonds. Level shocks correspond to near-permanent changes in interest

rates and only minimal changes in expected excess returns. Slope shocks correspond to business-cycle-length fluctuations in both interest rates and expected excess returns to bonds, while twist shocks correspond to short-lived "flight to quality" variations in expected excess returns. In other words, twist shocks do not affect current or expected future short-term interest rates; they are pure shocks to risk premia.

Essentially affine models are not magic bullets. The models cannot capture time variation in conditional variances without giving up part of their flexibility in fitting time variation in the price of interest rate risk. It remains to be seen whether an essentially affine model can be constructed that reproduces the time variation observed in both the conditional variances of yields and expected returns to bonds.

Appendix

This appendix gives closed-form representations for first and second conditional moments of a state vector that follow the affine process of (12a) and (2). The results are an application (and a specialization) of the results in Fisher and Gilles (1996).

Assume that K can be diagonalized, or

$$K = NDN^{-1}, \quad D \text{ diagonal.} \tag{A1}$$

The diagonal elements of D are denoted $d_i, \ldots, d_n$. A discussion of computing moments when K cannot be diagonalized is in Fisher and Gilles (1996).

The approach taken here is to compute the first and second conditional moments of a linear transformation of X_t. The transformation is chosen so that the feedback matrix K is diagonal under the transformation. The linear transformation is then reversed to calculate the conditional moments of X_t. Define

$$X_t^* \equiv N^{-1} X_t. \tag{A2}$$

Then the dynamics of X_t^* are, from (12a), (2), (A1), and (A2),

$$dX_t^* = D(\theta^* - X_t^*) + \Sigma^* S_t^* dW_t, \tag{A3}$$

where

$$S_{t(i,i)}^* = \sqrt{\alpha_i + \beta_i^{*\prime} X_t^*},$$

$$\theta^* = N^{-1}\theta,$$

$$\Sigma^* = N^{-1}\Sigma,$$

$$\beta^* = \beta N.$$

We now calculate the first and second moments of X_t^*. Some notation is helpful. If Z is an n-vector, the $n \times n$ diagonal matrix in which element (i,i) equals Z_i is denoted $diag(Z)$. If Z is a diagonal matrix, the diagonal matrix in which element (i,i) equals $e^{Z_{ii}}$ is denoted e^Z. Finally, the n-vector $\beta_{\bullet i}$ is column i of β.

A. Conditional Mean

The expectation of X_T^* conditional on X_t^* is given by

$$E[X_T^*|X_t^*] = \theta^* + e^{-D(T-t)}(X_t^* - \theta^*). \tag{A4}$$

Because $e^{-D(T-t)}$ is diagonal, this expectation can also be simply expressed element-by-element:

$$E[X_T^*|X_t^*] = \theta_i^* + e^{-d_i(T-t)}(X_{t,i}^* - \theta_i^*). \tag{A4'}$$

Another useful way to express (A4) is by separating the terms that depend on X_t^* from the terms that do not:

$$E[X_T^*|X_t^*] = (I - e^{-D(T-t)})\theta^* + e^{-D(T-t)}X_t^*. \tag{A4''}$$

Given this conditional mean of X_T^*, we reverse the transformation to express the conditional mean of X_T:

$$E[X_T|X_t] = NE[X_T^*|X_t] = N(I - e^{-D(T-t)})\theta^* + Ne^{-D(T-t)}N^{-1}X_t.$$

Note that the conditional mean of X_T could be expressed directly in terms of the parameters of (12a); no transformation into X_t^* is required, because the above expression is equivalent to

$$E[X_T|X_t] = (I - e^{-K(T-t)})\theta + e^{-K(T-t)}X_t,$$

where $e^{-K(T-t)}$ is the fundamental matrix associated with $-K(T-t)$. The value of the approach taken here is that (A4') is used in determining the conditional variance–covariance matrix of X_t.

B. Conditional Variance

The matrix $\Sigma^* S_t^* S_t^{*\prime} \Sigma^{*\prime}$ is the instantaneous variance–covariance matrix of the transformed state vector. We can write this as

$$\Sigma^* S_t^* S_t^{*\prime} \Sigma^{*\prime} = \Sigma^* diag(\alpha^*) \Sigma^{*\prime} + \sum_{i=1}^n \Sigma^* diag(\beta_{\bullet i}^*) \Sigma^{*\prime} X_{t,i}^*$$

$$\equiv G_0 + \sum_{i=1}^n G_i X_{t,i}^*, \tag{A5}$$

where $G_0 \equiv \Sigma^* diag(\alpha^*) \Sigma^{*\prime}$ and the $n \times n$ matrices G_i are defined as $[\Sigma^* diag(\beta^*_{\bullet i}) \Sigma^{*\prime}]$. Define the $n \times n$ matrix $F(t,s)$ as

$$F(t,s) \equiv G_0 + \sum_{i=1}^{n} G_i [E(X_s^* | X_t^*)]_i.$$

This matrix is the instantaneous variance–covariance matrix of X_s^*, but evaluated at the expectation of X_s^* (conditional on time-t information) instead of at the true value of X_s^*. Using (A4′), this matrix can be expressed as

$$F(t,s) = G_0 + \sum_{i=1}^{n} G_i [\theta_i^* + e^{-d_i(s-t)} (X_{t,i}^* - \theta_i^*)]. \tag{A6}$$

Fisher and Gilles (1996) show the conditional variance of X_T^* can be written as

$$\mathrm{Var}[X_T^* | X_t^*] = \int_t^T e^{-D(T-s)} F(t,s) e^{-D(T-s)} \, ds. \tag{A7}$$

Substituting (A6) into (A7) produces (A8):

$$\mathrm{Var}[X_T^* | X_t^*] = \int_t^T e^{-D(T-s)} G_0 e^{-D(T-s)} \, ds$$

$$+ \sum_{i=1}^{n} \left[\theta_i^* \int_t^T e^{-D(T-s)} G_i e^{-D(T-s)} \, ds \right]$$

$$+ \sum_{i=1}^{n} \left[(X_{t,i}^* - \theta_i^*) \int_t^T e^{-D(T-s)} G_i e^{-D(T-s)} e^{-d_i(s-t)} \, ds \right]. \tag{A8}$$

If $f(j,k)$ maps (j,k) into the scalar value f, the notation $\{f(j,k)\}$ denotes the matrix with element (j,k) given by $f(j,k)$. The conditional variance can then be written as

$$\mathrm{Var}[X_T^* | X_t^*] = \int_t^T \{[G_0]_{j,k} e^{(s-T)(d_j+d_k)}\} ds$$

$$+ \sum_{i=1}^{n} \left[\theta_i^* \int_t^T \{[G_i]_{j,k} e^{(s-T)(d_j+d_k)}\} ds \right]$$

$$+ \sum_{i=1}^{n} \left[(X_{t,i}^* - \theta_i^*) \int_t^T \{[G_i]_{j,k} e^{(s-T)(d_j+d_k)-d_i(s-t)}\} ds \right]. \tag{A9}$$

Integrating (A9) produces (A10):

$$\mathrm{Var}[X_T^*|X_t^*] = \{(d_j + d_k)^{-1}[G_0]_{j,k}(1 - e^{-(T-t)(d_j+d_k)})\}$$

$$\times \sum_{i=1}^{n}[\theta_i^*\{(d_j + d_k)^{-1}[G_i]_{j,k}(1 - e^{-(T-t)(d_j+d_k)})\}]$$

$$+ \sum_{i=1}^{n}[(X_{t,i}^* - \theta_i^*)\{(d_j + d_k - d_i)^{-1}[G_i]_{j,k}$$

$$\times (e^{-d_i(T-t)} - e^{-(d_j+d_k)(T-t)})\}]. \tag{A10}$$

Note that by collecting terms, the variance–covariance matrix in (A10) can be rewritten in terms of the individual elements of X_t^* as in (A11):

$$\mathrm{Var}[X_T^*|X_t^*] = b_0 + \sum_{i=1}^{n} b_i X_{t,i}^*. \tag{A11}$$

The $n \times n$ matrices $b_i, i = 0,\ldots,n$ depend on the horizon $T - t$. We now calculate the conditional variance of X_T using the notation of (A11). Since

$$\mathrm{Var}(X_T|\Omega) = N\,\mathrm{Var}(X_T^*|\Omega)N',$$

we have

$$\mathrm{Var}(X_T|X_t) = Nb_0N' + \sum_{i=1}^{n}\left(\sum_{j=1}^{n} Nb_jN'N_{j,i}^{-1}\right)X_{t,i}.$$

REFERENCES

Andersen, Torben G., and Jesper Lund, 1997, Estimating continuous-time stochastic volatility models of the short-term interest rate, *Journal of Econometrics* 77, 343–377.

Bliss, Robert R., 1997, Testing term structure estimation methods, *Advances in Futures and Options Research* 9, 197–231.

Chacko, George, 1997, Multifactor interest rate dynamics and their implications for bond pricing, Working paper, Harvard University.

Chan, K. C., G. Andrew Karolyi, Francis A. Longstaff, and Anthony B. Sanders, 1992, An empirical comparison of alternative models of the short-term interest rate, *Journal of Finance* 47, 1209–1227.

Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53, 385–407.

Dai, Qiang, and Kenneth J. Singleton, 2000, Specification analysis of affine term structure models, *Journal of Finance* 55, 1943–1978.

Dai, Qiang, and Kenneth J. Singleton, 2001, Expectation puzzles, time-varying risk premia, and dynamic models of the term structure, Working paper, Stanford University; forthcoming, *Journal of Financial Economics*.

Duarte, Jefferson, 2000, The relevance of the price of risk in affine term structure models, Working paper, University of Chicago.

Duffie, Darrell, and Rui Kan, 1996, A yield-factor model of interest rates, *Mathematical Finance* 6, 379–406.

Duffie, Darrell, Jun Pan, and Kenneth Singleton, 2000, Transform analysis and asset pricing for affine jump-diffusions, *Econometrica* 68, 1343–1376.

Fama, Eugene F., and Robert R. Bliss, 1987, The information in long-maturity forward rates, *American Economic Review* 77, 680–692.

Fama, Eugene F., and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fisher, Mark, 1998, A simple model of the failure of the expectations hypothesis, Working paper, Federal Reserve Board.

Fisher, Mark, and Christian Gilles, 1996, Estimating exponential-affine models of the term structure, Working paper, Federal Reserve Board.

Greene, William H., 1997, *Econometric Analysis*, 3rd ed. (Prentice-Hall, Upper Saddle River, NJ).

Hansen, Lars Peter, 1982, Large sample properties of generalized method of moment estimators, *Econometrica* 50, 1029–1053.

Litterman, Robert, and Jose Scheinkman, 1991, Common factors affecting bond returns, *Journal of Fixed Income* 1, 54–61.

McCulloch, J. Huston, and Heon-Chul Kwon, 1993, U.S. term structure data, 1947–1991, Working paper 93-6, Ohio State University.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Singleton, Kenneth J., 2001, Estimation of affine asset pricing models using the empirical characteristic function, *Journal of Econometrics* 102, 111–141.

Vasicek, Oldrich, 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* 5, 177–188.