Discrete-Time Models of Bond Pricing*

David Backus,[†] Silverio Foresi,[‡] and Chris Telmer[§]

October 23, 1998

Abstract

We explore a variety of models and approaches to bond pricing, including those associated with Vasicek, Cox-Ingersoll-Ross, Ho and Lee, and Heath-Jarrow-Morton, as well as models with jumps, multiple factors, and stochastic volatility. We describe each model in a common theoretical framework and explain the reasoning underlying the choice of parameter values. Our framework has continuous state variables but discrete time, which we regard as a convenient middle ground between the stochastic calculus of high theory and the binomial models of classroom fame. In this setting, most of the models we examine are easily implemented on a spreadsheet.

JEL Classification Codes: E43, G12.

Keywords: bond yields; pricing kernels; forward rates; fixed income derivatives.

*We thank Frank Diebold, Kai Li, and especially Liuren Wu for helpful comments on early versions of this paper. We also acknowledge a special debt to Stan Zin, whose reliable insight and good humor we have come to rely on. The most recent version of this paper, as well as associated Matlab programs, is available at http://www.stern.nyu.edu/~dbackus.

[†]Stern School of Business, New York University and NBER; dbackus@stern.nyu.edu.

[‡]Salomon Smith Barney; sforesi@sbi.com.

[§]Graduate School of Industrial Administration, Carnegie Mellon University; telmerc@andrew.cmu.edu.

1 Introduction

A newcomer to the theory of bond pricing would be struck by the enormous variety of models in use and by the variety of methods used to study them. We provide a selective review and synthesis of this diverse body of work, with the goal of clarifying differences in models and approaches. Our focus is less on theory for theory's sake and more on the properties of bond yields implied by a model's structure and parameter values. We think of this as engineering, in contrast to the physics of more theoretical papers. We are guided in this effort by four principles:

Principle 1: Common theoretical language. In the literature on bond pricing, models are variously described as "arbitrage-free," "equilibrium," and so on. They are developed using "state prices," "risk-neutral probabilities," or "pricing kernels." In fact, all of these approaches share a common intellectual foundation and can be expressed in a common way. We express each in terms of a pricing kernel, but explain differences in language and approach along the way.

Principle 2: Discrete time. Although discrete time is occasionally less elegant than continuous time, it makes fewer technical demands on users. As a result, we can focus our attention on the properties of a model rather than the technical issues raised by the method used to apply it.

Principle 3: Continuous state variables. We think it's important to let interest rates assume a continuous range of real values, and not the discrete set of possibilities familiar to users of binomial models. We regard this combination of discrete time and continuous states as a convenient middle ground between the stochastic calculus of high theory and the binomial models of classroom fame.

Principle 4: Parameter values are as important as models. Duffie (1992, pp xiii-xiv) writes: "The decade spanning roughly 1969-79 seems like a golden age of dynamic asset pricing theory. ... The decade or so since 1979 has, with relatively minor exceptions, been a mopping-up operation." While this may be true of theory, we think finance professionals continue to make significant progress in understanding the ability of models to explain the prices we observe in markets. In practice, the choice of parameters is critical to a model's performance and deserves as serious study as the model itself. In this respect, we think we can add something to theoretical books like Duffie's, and thus help to bridge the gap between theorists and practitioners.

With these principles in mind, we review the basic theory of asset pricing and its application to bonds. We express various models and approaches in a common theoretical framework in which time is discrete and state variables are (for the most part) continuous. The emphasis is on the use of these models by practitioners:

which model to use, how to solve it, and how to choose its parameter values. With one exception, we do not explore the application of these models to the pricing of derivatives. Although this is clearly the major use of fixed income models, both the theory and related data issues would make this a much longer paper. Nevertheless, the foundations laid here are a necessary first step in that direction.

After summarizing the salient features of US bond yields in Section 2, we review the theory underlying the modern approach to asset pricing (Section 3). In Section 4, we describe the popular Vasicek and Cox-Ingersoll-Ross models and explain how their parameters might be chosen to approximate some of the observed properties of bond yields. Discrepancies between these models and observed bond yields motivate more complex models. In Section 5, we discuss arbitrage-free models, in which time-dependent parameters are introduced to allow models to reproduce current market conditions, including the current yield or forward rate curve. Examples include a Ho-and-Lee-inspired version of the Vasicek model, which can be calibrated to the current yield or forward rate curve, and a linear application of the approach of Heath, Jarrow, and Morton. In Section 6, we consider the popular binomial framework and derive its implicit pricing kernel. In Section 7, we describe Das's jump model, which allows innovations to interest rates to follow non-Gaussian distributions. Section 8 is devoted to multifactor affine models, including two-factor versions of Vasicek and Cox-Ingersoll-Ross, the Longstaff-Schwartz model, and models with stochastic volatility and central tendency factors. In Section 9, we provide a short discussion of options on zeros, emphasizing the term structure of volatility in log-normal environments in which the Black-Scholes formula holds exactly.

2 Notation and Evidence

The obvious starting point for any modeling exercise is a description of what the models are intended to explain. The models we examine are designed to explain prices of fixed income securities of all kinds. In principle, this includes not only the prices of bonds, but of interest-rate derivatives: swaps and swaptions, interest rate futures, caps and floors, and the like. We limit ourselves, however, to US treasury bonds, whose properties are described below.

We summarize bond prices in terms of yields. One of the (small) issues that arises when we do this in discrete time is that the time interval of the model need not correspond to the time interval over which yields are reported. By convention, we report yields and other interest rates as annual percentages. Our modeling time interval, however, is one month: that is, from here on we take one period to be one month. As a result, we must at times include translation factors of 12 or 1200, converting monthly yields to annual yields or annual percentages, respectively.

With this detail out of the way, we denote the continuously-compounded *yield* or *spot rate* on an n-period discount bond at date t by y_t^n, defined by

$$y_t^n \;=\; -n^{-1} \log b_t^n, \tag{1}$$

where b_t^n is the dollar *price* at date t of a claim to one dollar at $t+n$. One-period *forward rates* are defined by

$$f_t^n \;=\; \log(b_t^n / b_t^{n+1}), \tag{2}$$

so that yields are averages of forward rates:

$$y_t^n \;=\; n^{-1} \sum_{i=0}^{n-1} f_t^i. \tag{3}$$

The *short rate* is $r_t = y_t^1 = f_t^0$.

In practice, yields and forward rates are estimated rather than observed. From prices of bonds for a variety of maturities, the *discount function* b_t^n (viewed as a function of n at each date t) is interpolated between missing maturities n and smoothed to reduce the impact of noise (nonsynchronous price quotes, bid/ask spreads, and so on). We use data constructed by McCulloch and Kwon (1993) from quoted prices of US treasury securities.

The properties of US treasury yields between January 1952 and February 1991 are summarized in Tables 1 and 2. We will return to these tables later, when we use them to choose values for model parameters, but for now it's worth noting their basic features. One is the shape of the average yield curve: average yields rise with maturity between one month and ten years, but the rate of increase falls with maturity. This concave shape is familiar to observers of bond markets, but masks a great deal of variety in the shape of the yield curve at specific times. Another feature is persistence: autocorrelations of yields are well above 0.9 (monthly) for all maturities. Short-long spreads exhibit substantially less persistence, suggesting that some of the persistence in yields stems from something that is common to both short and long rates. A third feature is volatility, which is apparent in all three variables: yields, spreads, and monthly changes in yields. Note that the maturity patterns of these variables differ substantially. A final feature is kurtosis: both yield spreads and yield changes exhibit substantial excess kurtosis, particularly at short maturities.

3 Arbitrage and Pricing Kernels

Although the form varies, modern asset pricing theory is based on a single theoretical result: that in any arbitrage-free environment, there exists a positive random variable m that satisfies

$$1 = E_t(m_{t+1} R_{t+1}), \tag{4}$$

for (one-period) returns R on all traded assets at all dates t. We refer to m as a *pricing kernel*, since prices of assets grow from it. A model consists, then, of a description of m. The same content is sometimes expressed in terms of *state prices* or *risk-neutral probabilities*, which we discuss in Section 6.

One of the many nice things about the pricing relation (4) is that it applies to everything: once we have a model that values bonds, we can (in principle) use it to value bond-related derivatives of all kinds. The catch is the qualifier, "in principle." The theory works fine; the challenge in practice is to approximate the m that theory says must exist.

Bond pricing is an elegant, if straightforward, application of (4). If an arbitrary claim to next-period cash flows c_{t+1} costs p_t now, then the return is $R_{t+1} = c_{t+1}/p_t$ and (4) implies

$$p_t = E_t(m_{t+1} c_{t+1}). \tag{5}$$

This takes a particularly simple and useful form with bonds. The one-period return on an $n+1$-period bond is $R_{t+1} = b_{t+1}^n / b_t^{n+1}$, so the prices satisfy

$$b_t^{n+1} = E_t(m_{t+1} b_{t+1}^n). \tag{6}$$

From this we can compute bond prices recursively, starting with the initial condition $b_t^0 = 1$ (a dollar today costs one dollar).

4 One-Factor Models

Two of the most popular bond pricing models are those constructed by Vasicek (1977) and Cox, Ingersoll, and Ross (1985). We describe both in this section and explain how their parameter values might be chosen to correspond to properties of bond yields. Each of these models has a single factor, by which we mean that prices depend on a single state variable z (say), typically associated with the short rate r. The models are similar, too, in having four parameters: three governing the dynamic behavior of the state variable and one controlling the market's valuation of risk. With these ingredients, theory then tells us how long rates are connected to the short rate.

4.1 Vasicek

The archetype of bond pricing models is Vasicek (1977). In discrete time, the single state variable z follows a first-order autoregression (meaning we regress z on its own lag):

$$\begin{aligned} z_{t+1} &= \varphi z_t + (1-\varphi)\theta + \sigma\varepsilon_{t+1} \\ &= z_t + (1-\varphi)(\theta - z_t) + \sigma\varepsilon_{t+1}, \end{aligned} \tag{7}$$

with $\{\varepsilon_{t+1}\}$ distributed normally and independently with mean zero and variance one. The mean of z is θ. The conditional variance is σ^2 and the unconditional variance $\sigma^2/(1-\varphi^2)$, formulas that should be familiar to those acquainted with linear time series methods (for example, Harvey, 1993). The parameter φ controls mean reversion: If $\varphi = 1$, z is a random walk and shows no tendency to return to any specific value — hence the term "random walk." But if $0 < \varphi < 1$, z is expected to return to its mean value of θ at rate $1-\varphi$, as in the second line of equation (7). We complete the model with the pricing kernel m, which satisfies

$$-\log m_{t+1} = \delta + z_t + \lambda\varepsilon_{t+1}. \tag{8}$$

We refer to λ as the *price of risk*, since it determines the covariance between shocks to m and z, and thus the risk characteristics of bonds and related assets. We set $\delta = \lambda^2/2$ for reasons that will be apparent momentarily.

We compute bond prices recursively using the theory outlined in Section 3. The pricing relation (6) and initial condition $b_t^0 = 1$ tell us that the price of a one-period bond is the conditional mean of the pricing kernel: $b_t^1 = E_t m_{t+1}$. Since the kernel is conditionally log-normal, we need the following property of log-normal random variables: If $\log x$ is normal with mean μ and variance σ^2, then $\log E(x) = \mu + \sigma^2/2$. From equation (8) we see that $\log m_{t+1}$ has conditional mean $-(\delta + z_t)$ and conditional variance λ^2, so the one-period bond price satisfies

$$\log b_t^1 = -\delta - z_t + \lambda^2/2 = -z_t.$$

The short rate is therefore

$$r_t = -\log b_t^1 = z_t,$$

as claimed earlier. Since z is the short rate, we can base the value of the parameters of (7) on the properties of the short rate, such as those reported in Table 1.

Prices of long bonds follow by induction. Let us guess that the price of an n-period bond can be expressed

$$-\log b_t^n = A_n + B_n z_t \tag{9}$$

for some choice of coefficients $\{A_n, B_n\}$. Since $b_t^0 = 1$ we know $A_0 = B_0 = 0$, so we can certainly start this process up. The expression for a one-period bond implies $A_1 = 0$ and $B_1 = 1$. Given the coefficients for maturity n, we use (6) to evaluate the price of an $n+1$-period bond. The right side involves

$$\begin{aligned}\log m_{t+1} + \log b_{t+1}^n &= -\delta - z_t - \lambda\varepsilon_{t+1} - A_n - B_n z_{t+1} \\ &= -[A_n + \delta + B_n(1-\varphi)\theta] - (1 + B_n\varphi)z_t - (\lambda + B_n\sigma)\varepsilon_{t+1},\end{aligned}$$

which has conditional moments

$$E_t(\log m_{t+1} + \log b_{t+1}^n) = -[A_n + \delta + B_n(1-\varphi)\theta] - (1 + B_n\varphi)z_t$$

and

$$Var_t(\log m_{t+1} + \log b_t^n) \;=\; (\lambda + B_n\sigma)^2.$$

The implied bond price is therefore

$$-\log b_t^{n+1} \;=\; A_n + \delta + B_n(1-\varphi)\theta - (\lambda + B_n\sigma)^2/2 + (1 + B_n\varphi)z_t.$$

Lining up coefficients with (9) gives us the recursions

$$A_{n+1} \;=\; A_n + \delta + B_n(1-\varphi)\theta - (\lambda + B_n\sigma)^2/2 \tag{10}$$

$$B_{n+1} \;=\; 1 + B_n\varphi. \tag{11}$$

These equations look complicated, but given values for $(\theta, \varphi, \sigma, \lambda)$, we can easily evaluate them on a spreadsheet. They are a closed-form solution to the model, in the sense of being computable with a finite number of elementary operations.

Forward rates in this model take a particularly simple form, which we note for future reference:

$$f_t^n \;=\; (1-\varphi^n)\theta + \frac{1}{2}\left[\lambda^2 - \left(\lambda + \frac{1-\varphi^n}{1-\varphi}\sigma\right)^2\right] + \varphi^n z_t. \tag{12}$$

This expression illustrates the impact of the short rate on long forwards (the impact declines with n) and the form of the risk premium (ugly, but governed by λ). This expression gives us insight into the relation between forward rates and expected future short rates. If $\lambda = \sigma = 0$, $f_t^n = E_t r_{t+n}$, a common version of the expectations hypothesis. When $\lambda = 0$ and $\sigma \neq 0$, the nonlinearity of the pricing relation results in a downward sloping average forward rate curve (often referred to as the effects of "Jensen's inequality"). If λ is sufficiently positive, this effect is reversed and average forward rates (hence yields) increase with maturity.

The recursions tell us how to compute bond prices given values for the parameters. In practice, however, we are often interested in the reverse question: What parameter values are indicated by observed bond prices? We choose parameters to approximate some of the salient features of bond yields reported in Table 1. θ is the unconditional mean of the short rate, so we set it equal to the sample mean of the one-month yield in Table 1:

$$\theta = \frac{5.314}{1200} = 0.004428.$$

(1200 converts an annual percentage rate to a monthly rate.) The mean reversion parameter φ is the first autocorrelation of the short rate. In Table 1, the autocorrelation of y^1 is 0.976, so we set φ equal to this value. The volatility parameter σ is the standard deviation of innovations to the short rate. We choose it to equate the unconditional variance of the short rate equal to its value in the data:

$$\frac{\sigma^2}{1-\varphi^2} = \left(\frac{3.064}{1200}\right)^2.$$

With $\varphi = 0.976$, the implied value is $\sigma = 0.0005560$. Thus the values of $(\theta, \sigma, \varphi)$ are chosen to match the mean, standard deviation, and autocorrelation of the short rate.

We choose the final parameter, the price of risk λ, to approximate the slope of the yield curve. A little experimentation tells us that λ governs the average slope of the yield curve, with negative values required to reproduce the upward slope we see in the data. Mean bond yields in the model are

$$E\left(y^n\right) = n^{-1}\left(A_n + B_n\theta\right). \tag{13}$$

The value $\lambda = -0.0824$ reproduces the mean 10-year bond yield, as we see in Figure 1. With more negative values the mean yield curve is steeper, and with less negative (or positive) values the yield curve is flatter (or downward sloping).

We see in Figure 1 that the model generates a mean yield curve with much less curvature than we see in the data. The problem is φ: the time series of the short rate indicates a value of φ close to one, but we need a smaller value to generate the required concavity of the yield curve. There is no choice of this parameter (or the others) that does both. We will see shortly that the Cox-Ingersoll-Ross model suffers from the same deficiency.

4.2 Cox-Ingersoll-Ross

The Cox-Ingersoll-Ross (CIR) model has a similar structure. The difference lies in the behavior of the state variable z: In the Vasicek model the conditional variance is constant, while in CIR it varies with the state. Our version follows Sun (1992, eq 6): z obeys the "square root process"

$$z_{t+1} = (1-\varphi)\theta + \varphi z_t + \sigma z_t^{1/2}\varepsilon_{t+1}, \tag{14}$$

with $0 < \varphi < 1$, $\theta > 0$, and $\{\varepsilon_t\}$ is distributed normally and independently with mean zero and variance one. Despite the unusual form of the innovation, (14) is a first-order autoregression. The unconditional mean of z is θ. (To show this, it's useful to compute the conditional expectation first, then the unconditional expectation.) The first autocorrelation is φ and higher-order autocorrelations are powers of φ. The conditional variance is

$$Var_t(z_{t+1}) = z_t\sigma^2,$$

which has a mean of $\theta\sigma^2$. The unconditional variance is $Var(z) = \theta\sigma^2/(1-\varphi^2)$.

The most interesting feature of (14) is that it guarantees nonnegative z if the time interval is small: With the square-root process the conditional variance gets smaller as z approaches zero, which reduces the chance of getting a negative value. With normal ε's there is still a positive probability that z_{t+1} is negative, but the probability falls to zero as the time interval shrinks. In continuous time, z is strictly positive under the stated conditions. This is a useful feature in a bond pricing model, since the existence of currency places a lower bound of zero on nominal interest rates.

The pricing kernel for a discrete time version of CIR is

$$-\log m_{t+1} = (1+\lambda^2/2)z_t + \lambda z_t^{1/2}\varepsilon_{t+1}, \tag{15}$$

so again the kernel is conditionally log-normal. We will see shortly that the coefficient of z is a fortuitous choice, intended to make z the short rate.

Bond pricing in this setting is similar to Vasicek: We apply the pricing relation (6) to the pricing kernel (15) and compute bond prices recursively. The key to making this work is that both the conditional mean and the conditional variance are linear functions of z. As a result, bond prices are log-linear functions of z as in equation (9). Using the same methods we applied to the Vasicek model, we find that the coefficients of the log-linear bond price formulas satisfy the recursions,

$$\begin{aligned} A_{n+1} &= A_n + B_n(1-\varphi)\theta \\ B_{n+1} &= 1 + \lambda^2/2 + B_n\varphi - (\lambda + B_n\sigma)^2/2, \end{aligned}$$

starting with $A_0 = B_0 = 0$. Since $A_1 = 0$ and $B_1 = 1$, z is the short rate.

We choose values for parameters much as we did for the Vasicek model. We set the autocorrelation parameter φ equal to the autocorrelation of the short rate:

$\varphi = 0.976$. We set θ equal to the mean short rate, which again implies $\theta = 0.004428$. We choose σ to reproduce the variance of the short rate:

$$\frac{\theta\sigma^2}{1-\varphi^2} = \left(\frac{3.064}{1200}\right)^2,$$

which implies $\sigma = 0.008356$. Average yields follow (13). The average 10-year bond yield implies $\lambda = -1.07$. The results of this exercise are pictured in Figure 2, which is virtually identical to Figure 1.

4.3 Technical Issues

With the warning that many readers should skip this section, we review a number of technical issues.

1. Unit roots. Some have suggested that the short rate process should have a unit root: that φ should be one. The evidence from short rates isn't wildly at odds with this idea, perhaps not at odds at all given the low power of unit root tests against nearby alternatives. What is left out, however, is the implication of a unit root for the spread between long and short rates: the mean spread gets increasingly negative at long enough maturities, approaching minus infinity in the limit. In practice, this might become apparent only at maturities beyond the interest of any practitioner. Another issue is the lower bound of zero imposed on the short-rate by the CIR model. This is inconsistent, in this setting, with a unit value for φ.

2. Normalizations. In the pricing kernels, we chose coefficients of z equal to 1 (Vasicek) and $1+\lambda^2/2$ (CIR), both with the purpose of equating z to the short rate. There's nothing sacred about this. We could have easily chosen normalizations that equated z to the 5-year forward rate, the 10-year yield, minus the short rate, or the spread between the 10-year yield and the short rate. Since all are linear in z, they result in linear transformations of the model. In the Vasicek model, all such versions are observationally equivalent. In the CIR model, versions are equivalent for all rates or spreads that are increasing in the state variable. (The square root means that z has a sign as well as a magnitude.)

3. Intercepts in pricing kernels. If we allow δ in (8) to be an additional free parameter in the pricing kernel of the Vasicek model (earlier we set it equal to $\lambda^2/2$), we find that bond prices depend only on the sum $\delta + \theta$. Neither parameter can be identified separately. We chose (effectively) to drop δ from the model. An equivalent choice is to drop θ by setting it equal to zero. The two versions of the model imply identical asset prices. In the CIR model, this isn't the case; see Pearson and Sun (1994).

4.4 Assessment

These one-factor models are a good place to start, but they are not a good place to stop: there are simply too many discrepancies between them and the world around us. These discrepancies make greater demands on modelers, and make this a longer paper.

One discrepancy in the Vasicek and CIR models is the shape of the mean yield curve: If φ is chosen to reproduce the autocorrelation of the short rate, the mean yield curve is substantially less concave in the models than it is in the data. This anomaly was pointed out by Gibbons and Ramaswamy (1993) and remains, in our view, one of the obvious signs that one-factor models are inadequate.

Another discrepancy is the pattern of autocorrelations. Both of these models are linear: yields of all maturities, yield spreads, and, indeed, all linear combinations of yields are linear functions of z. As a consequence, they share with z its autocorrelation. In the data, however, long yields and yield spreads exhibit, respectively, higher and lower autocorrelations than the short rate. A related issue is the decline in volatility with maturity, such as the standard deviations of one-month changes reported in Table 2B. These models imply less variability of long rates than short rates, but the rate of decline is greater in both models than we see in the data.

Yet another discrepancy is that innovations in z are conditionally normal. The evidence suggests, to the contrary, that interest-rate innovations have substantial excess kurtosis. Note, for example, the kurtosis of one-month changes in the short rate (Table 2B).

All of these discrepancies point toward the more complex models to come.

5 Arbitrage-Free Models

Ho and Lee (1986) started a revolution in industry practice that has been carried on by Black, Derman, and Toy (1990); Heath, Jarrow, and Morton (1992); Hull and White (1990, 1993); and many others. The logic of most academic work, of which Section 4 is typical, is to choose parameter values that approximate average behavior of bond yields. For practical use this kind of approximation is inadequate. The four parameters of the Vasicek and CIR models can be chosen to match five points on the yield curve (four parameters plus the short rate), but cannot approximate the complete yield curve to the degree of accuracy required by market participants. Ho and Lee suggested that such models might include additional time-dependent adjustment factors that could be used to "tune" them to observed asset prices. In the most common applications, adjustment factors are used to allow the model to match the current yield curve exactly. Such models are generally referred to (with some violence to the language) as "arbitrage-free."

5.1 The Approach of Ho and Lee

Although Ho and Lee (1986) used a binomial model, we can illustrate their insight in the Vasicek model. The result bears more than a passing resemblance to Hull and White's (1990, 1993) "extended Vasicek model."

Consider a variant of the Vasicek model with state equation

$$z_{t+1} = \varphi z_t + \sigma \varepsilon_{t+1} \tag{16}$$

and pricing kernel

$$-\log m_{t+1} = \lambda^2/2 + \delta_t + z_t + \lambda \varepsilon_{t+1}.$$

One change from the model in Section 4.1 is the elimination of θ. This comes without loss of generality, for reasons outlined in Section 4.3 (θ and δ perform the same function). The critical change is the presence of the time-dependent intercept δ in the pricing kernel. We will see shortly that we can choose $(\delta_t, \delta_{t+1}, \ldots)$ to reproduce any observed yield curve.

Once again prices are log-linear functions of z, but the functions depend on time:

$$-\log b_t^n = A_{nt} + B_{nt} z_t.$$

The pricing relation implies

$$\begin{aligned} A_{n+1,t} &= A_{n,t+1} + \lambda^2/2 + \delta_t - (\lambda + B_{nt}\sigma)^2/2 \\ B_{n+1,t} &= 1 + B_{n,t+1}\varphi. \end{aligned}$$

The boundary conditions, $B_{0t} = 0$ for all t, imply

$$B_{nt} = 1 + \varphi + \cdots + \varphi^{n-1} = \frac{1-\varphi^n}{1-\varphi}$$

for all t. Forward rates are therefore

$$\begin{aligned} f_t^n &= A_{n+1,t} - A_{nt} + (B_{n+1,t} - B_{nt})\, z_t \\ &= \delta_{t+n} + \frac{1}{2}\left[\lambda^2 - \left(\lambda + \frac{1-\varphi^n}{1-\varphi}\sigma\right)^2\right] + \varphi^n z_t. \end{aligned} \tag{17}$$

It's immediately apparent that we can choose the δ's to make forward rates — and hence yields — anything we like.

Ho and Lee developed this approach in a binomial model, but the idea is more general: to add time-dependent parameters that allow users to match observed bond yields. Since then, others have noted that other parameters might also be allowed to vary with time. The most important of these is the volatility parameter σ, which Black, Derman, and Toy (1990) showed could be chosen to reproduce the volatility of different parts of the yield curve. This extension was critical to the pricing of interest-rate related options, for which volatility is a key parameter. Moreover, there is overwhelming evidence in these markets that volatility varies with maturity as well as time. Hull and White (1990, 1993) further refined the approach by allowing analogs of θ and φ to vary with time.

These additional parameters are clearly needed in applied work, where a model that fails to reproduce the current yield curve can hardly be trusted to price more complex securities. At the same time, they are no panacea: even a bad model can be tuned to reproduce the current yield curve with enough extra parameters. What's needed is a balance between the fundamental parameters of the model $(\varphi, \sigma, \lambda)$ and the time-dependent adjustment factors (δ_t in our example). The modeling efforts in this paper contribute primarily to constructing a model that is good on average. Once this is done, practical application will almost certainly call for adjustment factors along the lines described here.

5.2 The Approach of Heath, Jarrow, and Morton

Heath, Jarrow, and Morton (1992) owe a debt to Ho and Lee in using time-dependent parameters, but they approach them from a novel and interesting direction: They focus on forward rates and the movement of the entire forward rate curve from one date to the next. Their approach exploits simplifications that stem from modeling forward rates directly and sheds new light on the role of volatility parameters in pricing models. We describe the implications of their approach for the pricing kernel, and describe how it can be used to calibrate models to both the forward rate curve and the term structure of volatility.

We illustrate the Heath, Jarrow, and Morton (HJM) approach with a linear one-factor example. Suppose the forward rate curve evolves according to

$$f_{t+1}^{n-1} \;=\; f_t^n + \alpha_{nt} + \sigma_{nt}\varepsilon_{t+1} \tag{18}$$

for all $n \geq 0$, where $\{\varepsilon_t\}$ is (as usual) iid normal with mean zero and variance one. HJM pose the question: What restrictions are placed on the parameters $\{\alpha_{nt}, \sigma_{nt}\}$ by the assumption that movements in forward rates are free from arbitrage opportunities? They attack this question by focusing directly on forward rates. If forward rates follow (18), the return on an $n+1$-period bond can be expressed

$$\begin{aligned}\log R_{t+1} &= r_t - \sum_{j=1}^{n} (f_{t+1}^{j-1} - f_t^j) \\ &= r_t - \sum_{j=1}^{n} \alpha_{jt} - \sum_{j=1}^{n} \sigma_{jt} \varepsilon_{t+1} \\ &= r_t - A_{nt} - S_{nt} \varepsilon_{t+1}, \end{aligned} \tag{19}$$

with the obvious definitions of the partial sums A_{nt} and S_{nt}.

At this point we take two different paths, one followed by HJM, the other more in keeping with our focus on the pricing kernel. HJM's path starts with the moments of the bond return, which include

$$\begin{aligned} Var_t (\log R_{t+1}) &= S_{nt}^2, \\ \log E_t R_{t+1} &= r_t - A_{nt} + S_{nt}^2/2. \end{aligned}$$

(This trickery with logs is less troublesome in continuous time.) HJM assume that for some specific maturity τ, the expected excess return is proportional to its standard deviation:

$$- A_{\tau t} + S_{\tau t}^2/2 \;=\; -\gamma_t S_{\tau t}. \tag{20}$$

They refer to the proportionality factor γ_t as the market price of risk. Absence of arbitrage opportunities then places restrictions on the parameters.

A second path is based on a pricing kernel. We replace HJM's price of risk relation (20) with a pricing kernel of the form

$$- \log m_{t+1} \;=\; \delta_t + \lambda_t \varepsilon_{t+1}. \tag{21}$$

Applying the pricing relation (4) to the return (19) gives us

$$r_t \;=\; \delta_t + A_{nt} - (\lambda_t + S_{nt})^2/2$$

for all $n \geq 0$, with the convention $A_{0t} = S_{0t} = 0$. The difference between this equation for $n = 0$ and $n = \tau$ implies the restrictions

$$A_{\tau t} - \lambda_t S_{\tau t} - S_{\tau t}^2/2 \;=\; 0 \tag{22}$$

for all $\tau \geq 0$.

The two solution paths are readily shown to lead to the same destination. Equation (20) implies

$$A_{\tau t} - \gamma_t S_{\tau t} - S_{\tau t}^2/2 \;=\; 0,$$

which is equivalent to (22) if $\gamma_t = \lambda_t$. This equivalence should come as no surprise. We know from Section 3 that absence of arbitrage opportunities implies the existence of a pricing kernel satisfying (4). Given the linearity of the forward rate equations (18), it should be no surprise either that the pricing kernel takes the log-linear form of (21).

The benefits of working directly with forward rates are apparent if we compare the HJM characterization of the solution, equations (18) and (22), to that implied by our treatment in the previous section. Our time-dependent version of Vasicek, summarized by equations (17,16), implies

$$f_{t+1}^{n-1} = f_t^n + \frac{1}{2}\left[\left(\lambda + \frac{1-\varphi^n}{1-\varphi}\sigma\right)^2 - \left(\lambda + \frac{1-\varphi^{n-1}}{1-\varphi}\sigma\right)^2\right] + \varphi^{n-1}\sigma\varepsilon_{t+1}.$$

This is a special case of our HJM example with

$$\begin{aligned}
\sigma_{jt} &= \varphi^{j-1}\sigma \\
\alpha_{jt} &= \frac{1}{2}\left[\left(\lambda + \frac{1-\varphi^j}{1-\varphi}\sigma\right)^2 - \left(\lambda + \frac{1-\varphi^{j-1}}{1-\varphi}\sigma\right)^2\right].
\end{aligned}$$

These choices satisfy (22), but the computations are substantially more complex than (22), even in this linear setting.

One of the most useful features of the HJM approach is the ability to specify arbitrary volatilities. As we just saw, the Vasicek model implies geometrically declining volatility, the direct result of mean reversion in the state variable. With HJM, we can choose any volatilities we like, including those implied by option prices. This suggests a sequential choice of parameter values. First, we choose volatilities σ_{jt} to match the term structure of volatilities implied by (say) options on interest rate futures. Given these choices (and a value for λ_t), we choose "drift" parameters α_{jt} to satisfy (22). The resulting prices are arbitrage-free by construction.

6 Binomial Models

Binomial models are easily explained and implemented, which undoubtedly accounts for their widespread use in teaching and industry. They are based, obviously, on a discrete state variable: the short rate either rises or falls each period by a preset amount. They are also invariably described using state prices or risk-neutral probabilities, rather than a pricing kernel. Our purpose here is to explain the connections between the various languages used to describe models.

6.1 Alternatives to the Pricing Kernel

In binomial models, the state can either go "up" or "down" over any unit of time. These movements have been described in a variety of ways, but the Ho and Lee model serves as a prototype for fixed income. In this model, as in the Vasicek and Cox-Ingersoll-Ross models, the state variable is generally taken to be the short rate. Between any two consecutive dates t and $t+1$, changes in the short rate follow

$$r_{t+1} \;=\; r_t + \alpha_t + \sigma\varepsilon_{t+1}, \tag{23}$$

with

$$\varepsilon_{t+1} \;=\; \begin{cases} +2(1-\pi) & \text{with probability } \pi \\ -2\pi & \text{with probability } 1-\pi \end{cases}$$

This differs from equation (7) of the Vasicek model two respects: there is no mean reversion and (this is the key) the innovation takes on only two values. The parameter α governs the expected change or "drift" in the short rate and σ governs its conditional variance. The spread between the up and down states is 2σ. The variance is

$$Var_t(r_{t+1}) \;=\; 4\sigma^2\pi(1-\pi),$$

so the standard deviation is σ when $\pi = 0.5$, less than σ for other values. We refer to π and $1-\pi$ as the true probabilities to distinguish them from their risk-neutral counterparts.

Our approach to pricing has been to use a pricing kernel. Suppose the cash flows c_{t+1} are either c_u in the up state or c_d in the down state. Equation (5) tells us that they are worth

$$p_t \;=\; E_t(m_{t+1}c_{t+1}) \;=\; \pi m_u c_u + (1-\pi) m_d c_d, \tag{24}$$

where m_u and m_d represent the values of the kernel in the two states. Despite the discreteness, the principle is the same one we outlined in Section 3.

The pricing kernel highlights the interaction of probabilities and risk. Given a choice of m, a lower probability reduces the value of a payment in that state. Given probabilities, m summarizes the market's attitude toward risk. To see this, suppose m is constant. Since the probabilities sum to one (they're probabilities, after all), this constant value is just the price of a one-period bond: $m_u = m_d = b_t^1 = \exp(-r_t)$. The pricing relation (24) then becomes

$$p_t \;=\; e^{-r_t}\left[\pi c_u + (1-\pi)c_d\right] \;=\; e^{-r_t}E_t(c_{t+1}).$$

The price, in other words, is the discounted value of expected cash flows. Since only the expected cash flow affects the price, we might regard pricing as risk neutral, this being the way a person who didn't care about risk would value it. In general, of course, m is not the same in all states, and we can think of variations in m across states as reflecting attitudes toward risk of market participants.

A second theoretical language to describe pricing is based on *state prices*. Let q_u be the value now of one dollar next period in the up state and q_d the analogous value in the down state. We define these state prices by

$$\begin{aligned} q_u &= \pi m_u \\ q_d &= (1-\pi) m_d. \end{aligned}$$

Then equation (24) can be rewritten as

$$p_t = q_u c_u + q_d c_d. \tag{25}$$

State prices, then, are an equivalent approach to valuation.

The most common language for describing binomial models is based on *risk-neutral probabilities*. We denote them by $(\pi^*, 1-\pi^*)$ and define them by

$$\begin{aligned} \pi^* &= q_u/(q_u+q_d) = e^{r_t} \pi m_u \\ 1-\pi^* &= q_d/(q_u+q_d) = e^{r_t}(1-\pi) m_d, \end{aligned}$$

the second equality following from $q_u + q_d = b_t^1 = \exp(-r_t)$. The pricing relation becomes

$$p_t = \exp(-r_t)\left[\pi^* c_u + (1-\pi^*) c_d\right] = \exp(-r_t) E_t^*(c_{t+1}), \tag{26}$$

where $E_t^*(c_{t+1})$ means the conditional expectation of c_{t+1} computed from the risk-neutral probabilities.

The terminology deserves some explanation. People refer to $(\pi^*, 1-\pi^*)$ as *probabilities* for one obvious reason: they are positive and they sum to one. But unless m is constant, they are not the true probabilities. The modifier *risk neutral* is added to distinguish them from the true probabilities, and because the form of (26) is the same as our risk-neutral pricing with constant m. This is a little misleading: the effects of risk aversion are built in.

6.2 Pricing Risk

We now have three equivalent ways of describing models: a pricing kernel, state prices, and risk-neutral probabilities. The choice among them is a matter of convenience. Given the relations between them we can address a more substantive question: In a binomial model, what is the analog to the pricing kernel? The missing ingredient here is λ. Analogs to the other parameters are apparent from (23):

θ has been replaced by α_t, φ has been set equal to one, and σ plays similar roles in both binomial and continuous-state models. λ, however, remains a mystery.

Suppose we start, as commonly done in binomial models, with the risk neutral probability π^*. Given such a choice (one-half comes to mind), we apply (26) to value cash flows at each date. With a little effort, we can use the definitions of risk-neutral probabilities to compute the pricing kernel. The result is

$$-\log m_{t+1} \;=\; \delta^* + r_t + \lambda \varepsilon_{t+1},$$

where $\delta^* = \pi \log(\pi/\pi^*) + (1-\pi)\log[(1-\pi)/(1-\pi^*)]$ and

$$2\lambda \;=\; \log(\pi/\pi^*) - \log[(1-\pi)/(1-\pi^*)].$$

In words: the risk parameter λ is implicit in the difference between true and risk-neutral probabilities. This relation, moreover, is independent of the short rate process, whose only role here is to define the true probabilities used in (24).

7 Das's Jump Model

The original Vasicek and CIR models are based on continuous-time "diffusions" which means, essentially, that the innovations ε are normal. In fact, innovations in interest rates appear markedly non-normal, typically with fat tails indicative of kurtosis. Table 2B is suggestive: one-month changes in the short rate exhibit excess kurtosis of about 10. Since departures from normality can have a significant impact on prices of options and related derivatives, we discuss them at some length.

In continuous time, departures from normality over short time intervals are modeled with "point processes" or "jumps." In discrete time, we simply choose a non-normal distribution for the innovation ε. We illustrate this idea by modifying the Vasicek model described by equations (7) and (8). The approach is adapted from Das (1994) and Das and Foresi (1996). One of the simplest "abnormal" distributions is a mixture of normals:

$$\varepsilon_{t+1} \;=\; \begin{cases} \varepsilon_{1t+1} & \text{with probability } 1-\pi \\ \varepsilon_{2t+1} & \text{with probability } \pi, \end{cases}$$

with each ε_{it} an independent draw from a normal distribution with mean zero and variance τ_i. The mean of ε is therefore zero and the variance is $(1-\pi)\tau_1 + \pi\tau_2 = 1$ (the latter a continuation of our unit-variance convention).

Despite the modification, bond prices remain log-linear functions of the state variable z:

$$-\log b_t^n \;=\; A_n + B_n z_t.$$

We compute coefficients the usual way, starting with $A_0 = B_0 = 0$ and applying (6) to relate (A_n, B_n) to (A_{n+1}, B_{n+1}). The only difficulty involves terms of the form

$$E_t\left(e^{c\varepsilon_{t+1}}\right) \;=\; (1-\pi)e^{c^2\tau_1/2} + \pi e^{c^2\tau_2/2}$$

for an arbitrary constant c. The recursions are

$$\begin{aligned} A_{n+1} &= A_n + \delta + B_n(1-\varphi)\theta + \log\left[(1-\pi)e^{(\lambda+B_n\sigma)^2\tau_1/2} + \pi e^{(\lambda+B_n\sigma)^2\tau_2/2}\right] \\ B_{n+1} &= 1 + B_n\varphi. \end{aligned}$$

The choice

$$\delta = \log\left[(1-\pi)e^{\lambda^2\tau_1/2} + \pi e^{\lambda^2\tau_2/2}\right]$$

delivers $A_1 = 0$ and $B_1 = 1$, and thus sets the short rate r equal to z.

This model introduces some new parameters to the model: those governing the behavior of the mixture. Otherwise, our approach to choosing parameters is identical to our earlier treatment of the Vasicek model. We set the mixing probability $\pi = 0.05$ in the interest of simplicity. This is not an easy parameter to estimate precisely, although maximum likelihood or other methods can be applied. Our interest, however, is in reproducing the kurtosis of short rate innovations, which we label γ_2^ε. We estimate this to be 9.302, a slightly smaller value than we report in Table 2B for short rate changes. This value is computed from the residuals of a first-order autoregression for the short rate. The kurtosis of the model's innovation is

$$\gamma_2^\varepsilon \;=\; 3\frac{(1-\pi)+\pi\tau^2}{[(1-\pi)+\pi\tau]^2} - 3,$$

where $\tau = \tau_2/\tau_1$. With $\pi = 0.05$, our estimated value of γ_2^ε implies $\tau = 14.56 = 3.815^2$. This means that there is a five percent chance of drawing an interest rate innovation from a distribution whose standard deviation is almost 4 times its usual value. Given τ, the variances τ_1 and τ_2 are chosen to produce an overall variance of one:

$$1 \;=\; (1-\pi)\tau_1 + \pi\tau_2 \;=\; \tau_1(1-\pi+\pi\tau).$$

The result is $\tau_1 = 0.7720^2$ and $\tau_2 = 2.945^2$.

We proceed to identify the remaining parameters. As in the Vasicek model of Section 4.1, the autocorrelation, mean, and standard deviation of the short rate determine $\varphi = 0.976$, $\theta = 0.004428$, and $\sigma = 0.0005560$. We again choose λ to reproduce the average 10-year bond yield, setting $\lambda = -0.0817$. The resulting mean yield curve is indistinguishable from that of the Vasicek model (Figure 1).

We see, then, that Das's jump model provides a good approximation to the kurtosis in the short-term rate of interest and its innovations. What it cannot do is account for differences in kurtosis of yields and yield changes for different maturities: Since bond yields are linear functions of the same state variable z, their levels and changes have identical excess kurtosis for all maturities. Still, it provides a useful starting point for thinking about the role of jumps in pricing fixed income derivatives. In related work on currencies, we have found a Gram-Charlier expansion to be a more tractable non-normal distribution for pricing options; see Backus, Foresi, Li, and Wu (1998). Most of that work can be translated directly to fixed income.

8 Multifactor Models

We turn now to multifactor models, in which bond yields are governed by the movements in two or more state variables. The motivation for such models should be clear from Section 4.4: single-factor models cannot account for the average shape of the yield curve, the dynamics of interest rate spreads, or the pattern of interest-rate volatilities across maturities. For these reasons and others, practitioners often use models with multiple factors.

The Vasicek model is again the archetype. Its simple structure makes it relatively easy to understand and to assign parameter values. We develop the two-factor version at some length. We follow with a more cursory study of other multifactor "affine" or linear models, including two-factor CIR and Longstaff and Schwartz (1992) models. Our goal is the resolution of two discrepancies between the one-factor Vasicek and CIR models and observed bond yields. One is the average shape of the yield curve. As we saw in Figures 1 and 2, these one-factor models cannot simultaneously reproduce the observed curvature of the yield curve and persistence in the short rate. The second is the dynamics of yield spreads. In linear one-factor models, yield spreads have the same persistence as the short rate. In the data (Table 2), they are substantially less persistent. A two-factor model allows improvement along both dimensions.

8.1 Multifactor Vasicek

We base a multifactor generalization of Vasicek on independent state variables or factors z_i following

$$z_{it+1} \;=\; \varphi_i z_{it} + \sigma_i \varepsilon_{it+1}, \tag{27}$$

with innovations ε_{it} normally distributed with mean zero and variance one and independent across i and t. The pricing kernel is

$$-\log m_{t+1} \;=\; \delta + \sum_i \left(\lambda_i^2/2 + z_{it} + \lambda_i \varepsilon_{it+1} \right). \tag{28}$$

The kernel implies that the short rate is

$$r_t \;=\; \delta + \sum_i z_{it}. \tag{29}$$

Note that we have set the means of z_i equal to zero. In their place, we use δ to reproduce the mean of the short rate. This choice is dictated by the data: there is only one mean and it can determine only one parameter.

As in the one-factor model, each parameter has a clear role and interpretation. δ is the mean short rate. The variance and autocorrelation of the short rate — and other rates, as well — are controlled by $\{\sigma_i\}$ and $\{\varphi_i\}$. Finally, the λ's govern the correlation between innovations in the state variables and the pricing kernel: risk, in other words.

We construct bond prices from these components by the usual method. Bond prices remain log-linear functions of the state variables:

$$-\log b_t^n \;=\; A_n + \sum_i B_{in} z_{it}, \tag{30}$$

for some choice of coefficients $\{A_n, B_{in}\}$. The pricing relation (6) implies that the coefficients satisfy the recursions

$$\begin{aligned} A_{n+1} &= A_n + \delta + \frac{1}{2}\sum_i \left[\lambda_i^2 - (\lambda_i + B_{in}\sigma_i)^2\right] \\ B_{in+1} &= 1 + B_{in}\varphi_i, \end{aligned}$$

starting with $A_0 = B_{i0} = 0$. The solution implies forward rates of

$$f_t^n \;=\; \delta + \frac{1}{2}\sum_i \left[\lambda_i^2 - \left(\lambda_i + \frac{1-\varphi_i^n}{1-\varphi_i}\sigma_i\right)^2\right] + \sum_i \varphi_i^n z_{it}. \tag{31}$$

Note that more persistent factors (those with larger φ's) have relatively greater influence on long forward rates and yields. If a particular φ_i is close to one, then its effect is similar across maturities and state variable i has little effect on a yield or forward rate spread. In the two-factor case, we can estimate (roughly speaking) the parameters of the more persistent factor from long rates and those of the less persistent factor from yield spreads.

We consider parameter values in the two-factor version, a close relative of Brennan and Schwartz's (1979) two-factor model. The choice of parameters follows familiar logic, but the greater complexity of the model makes some of the steps a little more difficult. We now have seven parameters: δ and two choices of the triplet $(\varphi_i, \sigma_i, \lambda_i)$. We compute them from the moments in Tables 1 and 2 as follows:

- We estimate δ from the mean short rate: $\delta = 0.004428$.

- We compute $\{\sigma_i, \varphi_i\}$ to reproduce the variances and autocorrelations of the short rate and the spread between the short rate and the 5-year yield. The difficulty is that volatility (σ_i) and persistence (φ_i) parameters are now intertwined. Theory implies that an arbitrary yield or yield spread s is a linear function of the state variables:

  $$s_t \;=\; c_0 + c_1 z_{1t} + c_2 z_{2t},$$

  with coefficients $\{c_i\}$ related to $\{A_n, B_{in}\}$ as indicated by (1, 30). Each spread has variance

  $$Var(s) \;=\; c_1^2\, Var(z_1) + c_2^2\, Var(z_2) \tag{32}$$

  and autocorrelation

  $$Auto(s) = \frac{c_1^2\, Var(z_1)}{c_1^2\, Var(z_1) + c_2^2\, Var(z_2)}\, \varphi_1 + \frac{c_2^2\, Var(z_2)}{c_1^2\, Var(z_1) + c_2^2\, Var(z_2)}\, \varphi_2. \tag{33}$$

  Observations of variances and autocorrelations for two spreads allow us, in principle, to compute two σ's and two φ's. The difficulty is that the parameters must be computed simultaneously. Since B_{ni} in this model depends only on φ_i, each c_i is a function of φ_i alone. We can then compute the moments of spreads in this order. Given the φ's, we compute c_1 and c_2 and, from (32), the variances of the state variables. From them, we compute the volatility parameters σ_i. Given variances of the state variables, we can compute the autoregressive parameters $\{\varphi_i\}$ from (33). We are done when this circular path returns to the same values of $\{\varphi_i\}$ with which we started.

  The results of these computations, using "spreads" $s_t = r_t$ and $s_t = y^{60} - r_t$, are $\varphi_1 = 0.997$, $\sigma_1 = 0.000177$, $\varphi_2 = 0.858$, and $\sigma_2 = 0.000511$. Note that the first factor is the more persistent one.

- We estimate the λ's from the mean yields for maturities of 60 and 120 months. The intermediate 60-month rate captures the curvature that the one-factor model failed to reproduce. The implied values are $\lambda_1 = -0.0240$ and $\lambda_2 = -0.2884$.

These parameters go some ways toward resolving two of the problems with the one-factor model. We come much closer to the curvature of the average yield curve by using a small λ on the more persistent factor (the first one) and a larger one (in absolute value) on the second factor. This comes considerably closer to mean yields than the one-factor model (see Figure 3). We also reproduce the difference in autocorrelations of the short rate and the 5-year spread. The short rate is dominated by the more persistent factor and therefore inherits its persistence. The 5-year spread, on the other hand, emphasizes the less persistent factor and is therefore less highly autocorrelated.

This model isn't the last word in bond pricing, but it illustrates clearly how multiple factors can help to account for the unusual shifts and twists of the yield curve. Using the forward rate as a guide, equation (31) shows us that an increase in z_i is almost a parallel shift in the forward rates, since φ_1^n declines very slowly with n. The second factor is much less persistent, however, so an upward shift in z_2 has greater impact at short maturities (a "twist").

8.2 Affine Models

The multifactor Vasicek model is an example of a larger class of "affine" models, in which bond prices are log-linear functions of a vector of state variables. The underlying theory was developed by Duffie and Kan (1996). At the risk of increasing the level of abstraction, we summarize this class now to spare ourselves the effort of solving special cases separately. Casual readers should turn immediately to Section 8.3.

Expressed in discrete time, Duffie and Kan's affine models are based on a k-dimensional vector of state variables z that follow

$$z_{t+1} = (I - \Phi)\theta + \Phi z_t + V(z_t)^{1/2}\varepsilon_{t+1}, \tag{34}$$

where $\{\varepsilon_t\} \sim \text{NID}(0, I)$, $V(z)$ is a diagonal matrix with typical element

$$v_i(z) = \alpha_i + \beta_i^\top z,$$

β_i has nonnegative elements, and Φ is stable with positive diagonal elements. The process for z requires that the volatility functions v_i be positive, which places restrictions on the parameters. The pricing kernel takes the form

$$-\log m_{t+1} = \delta + \gamma^\top z_t + \lambda^\top V(z_t)^{1/2}\varepsilon_{t+1}. \tag{35}$$

Details are given in Duffie and Kan (1996) and translated into discrete time by Backus, Foresi, and Telmer (1996).

The multifactor Vasicek model is a special case. The parameters are related by

Affine Model	Vasicek Model
θ_i	0
Φ	$\text{diag}(\varphi_1, \ldots, \varphi_k)$
α_i	σ_i^2
β_i	0
δ	$\sum_i \lambda_i^2/2$
γ_i	1
$\lambda_i \alpha_i^{1/2}$	λ_i

With these choices, the affine structure reduces to the model of Section 8.1.

The primary benefit here of reviewing the general affine model is that we can characterize its solution once and be done with it. Bond prices are again log-linear functions of the state:

$$-\log b_t^n \;=\; A_n + B_n^\top z_t.$$

Applying the pricing relation (6) generates the recursions:

$$A(n+1) \;=\; A(n) + \delta + B(n)^\top (I-\Phi)\theta - \frac{1}{2}\sum_{j=1}^{k} (\lambda_j + B(n)_j)^2 \alpha_j \tag{36}$$

$$B(n+1)^\top \;=\; \left(\gamma^\top + B(n)^\top \Phi\right) - \frac{1}{2}\sum_{j=1}^{k} (\lambda_j + B(n)_j)^2 \beta_j^\top, \tag{37}$$

starting with $A(0) = 0$ and $B(0) = 0$. Moments of bond yields follow directly from those of the state variables. The state vector z has mean θ, so mean yields are

$$E\left(y^n\right) \;=\; n^{-1}\left(A_n + B_n^\top \theta\right).$$

The covariance matrix for z is Γ_0, which can be computed as

$$vec(\Gamma_0) \;=\; (I - \Phi \otimes \Phi)^{-1}\, vec[V(\theta)],$$

where $vec(A)$ is the vector formed from the columns of the matrix A. Autocovariance matrices obey

$$\Gamma_{j+1} \;=\; \Phi \Gamma_j,$$

for $j \geq 0$. Thus the first autocorrelation of an arbitrary linear combination $x = c^\top z$ can be calculated as

$$Corr(x_t, x_{t-1}) \;=\; \frac{c^\top \Gamma_1 c}{c^\top \Gamma_0 c}.$$

These relations allow us to compute means, variances, and autocorrelations of bond yields and yield spreads, as needed. The details will be familiar to readers with previous exposure to vector time series methods, mysterious to most others. The relevant material is described in Harvey (1989, ch 8).

None of this holds much interest for us in the abstract, but it can be applied directly to the special cases to which we now turn.

8.3 Other Affine Examples

Two-Factor CIR

The two-factor CIR model is a special case of the affine model in which z is two-dimensional, Φ is diagonal with elements φ_i, $\alpha = 0$, β_i has a single nonzero element

σ_i^2 in its ith position, $\delta = 0$, $\gamma_i = 1 + \lambda_i^2/2$, and λ_i corresponds to $\lambda_i \beta_{ii}^{1/2}$. The short rate is then $r_t = z_{1t} + z_{2t}$. We compute bond prices using (36,37).

This model has 8 parameters, one more than the two-factor Vasicek model, and therefore requires one more feature of the data to estimate. The extra parameter is one of θ_1 and θ_2, which was replaced by the single parameter δ in the Vasicek model. θ_1 and θ_2 are, however, notoriously hard to pin down. In the two-factor Vasicek model, θ_1 and θ_2 are not separately identified. In the CIR model, we have found no features of the data that lead to clear and precise estimates of their values. We could use, for example, estimates of the conditional variance or unconditional higher moments of the short rate. None of the alternatives are easy, and none work very well. Interested readers might consult Chen and Scott (1993), Duffie and Singleton (1997), or Backus, Foresi, Mozumdar, and Wu (1998).

The following models can be viewed as attempts to restrict the two-factor affine model further, thereby simplifying the estimation process.

Longstaff and Schwartz

The Longstaff and Schwartz (1992) model is a special case of the two-factor CIR model in which one of the risk parameters has been set equal to zero. The model consists of the equations

$$\begin{aligned}
z_{1t+1} &= (1-\varphi_1)\theta_1 + \varphi_1 z_{1t} + \sigma_1 z_{1t}^{1/2} \varepsilon_{1t+1} \\
z_{2t+1} &= (1-\varphi_2)\theta_2 + \varphi_2 z_{2t} + \sigma_2 z_{2t}^{1/2} \varepsilon_{2t+1} \\
-\log m_{t+1} &= (1+\lambda_1^2/2) z_{1t} + z_{2t} + \lambda_1 z_{1t}^{1/2} \varepsilon_{1t+1}.
\end{aligned}$$

The model implies $r_t = z_{1t} + z_{2t}$. Note the absence of λ_2 in the last equation.

Central Tendency

A model that addresses directly the difficulty of estimating the means of the two-factor CIR model is the Balduzzi-Das-Foresi (1998) model of central tendency. Our version consists of

$$\begin{aligned}
z_{1t+1} &= (1-\varphi_1) z_{2t} + \varphi_1 z_{1t} + \sigma_1 z_{1t}^{1/2} \varepsilon_{1t+1} \\
z_{2t+1} &= (1-\varphi_2)\theta_2 + \varphi_2 z_{2t} + \sigma_2 z_{2t}^{1/2} \varepsilon_{2t+1} \\
-\log m_{t+1} &= (1+\lambda_1^2/2) z_{1t} + (\lambda_2^2/2) z_{2t} + \lambda_1 z_{1t}^{1/2} \varepsilon_{1t+1} + \lambda_2 z_{2t}^{1/2} \varepsilon_{2t+1}.
\end{aligned}$$

The short rate is $r_t = z_{1t}$. Balduzzi, Das, and Foresi refer to z_2 as the "central tendency," since the short rate adjusts toward it. As a result, there is no θ_1 in the model. That leaves us with 7 parameters: θ_2, φ_1, φ_2, σ_1, σ_2, λ_1, and λ_2. We could estimate them using the same features of bond yields we used to estimate the 7 parameters of the two-factor Vasicek model.

Stochastic Volatility

Another interesting feature of the CIR model and its affine generalizations is that they exhibit stochastic volatility. In the one-factor CIR model, the conditional variance of the short rate is proportional to the short rate:

$$Var_t(r_{t+1}) = \sigma^2 r_t.$$

In the two-factor CIR and Longstaff and Schwartz models, the conditional variance depends on both state variables:

$$Var_t(r_{t+1}) = \sigma_1^2 z_{1t} + \sigma_2^2 z_{2t}.$$

Here we consider a third stochastic volatility model in which the conditional variance is a state variable in its own right:

$$\begin{aligned} z_{1t+1} &= (1-\varphi_1)\theta_1 + \varphi_1 z_{1t} + z_{2t}^{1/2}\varepsilon_{1t+1} \\ z_{2t+1} &= (1-\varphi_2)\theta_2 + \varphi_2 z_{2t} + \sigma_2 z_{2t}^{1/2}\varepsilon_{2t+1} \\ -\log m_{t+1} &= z_{1t} + (\lambda_2^2/2 + \lambda_2^2/2) z_{2t} + \lambda_1 z_{1t}^{1/2}\varepsilon_{1t+1} + \lambda_2 z_{2t}^{1/2}\varepsilon_{2t+1}. \end{aligned}$$

With this structure, z_1 is the short rate and z_2 its conditional variance. This model, too, has 7 parameters.

Three-Factor Models

Additional affine models can be constructed by combining elements of those listed above or heading off in new directions. Interested readers should consult Balduzzi, Das, Foresi, and Sundaram (1996), Chen and Scott (1993), and Dai and Singleton (1997).

9 Introduction to Options

Thus far, we have focused our efforts on bonds, ignoring entirely the use of these models to value fixed income derivatives. We now provide a brief overview of European options on zero-coupon bonds with two goals in mind. The first is to illustrate the principles involved: we apply the same pricing relation we used for bonds. As we noted in Section 3, options and other derivatives are simply more complex applications of equation (4) and its successors. The second is to describe, in a log-normal environment, the behavior of implied volatility across maturities of options and bonds. We note, in particular, the role of mean reversion in determining the volatility of long options and options on long bonds.

Suppose, for the sake of concreteness, that interest rate derivatives are governed by the Vasicek model of Section 4.1. The convenient implication in this context is that call prices obey the Black-Scholes formula. A call option on a zero must specify three terms: the strike price k, the maturity τ of the option, and the maturity n at expiration of the bond on which the option is written. The price of such a call option is

$$c_t^{\tau,n} \;=\; E_t\left[M_{t,t+\tau}\left(b_{t+\tau}^n - k\right)^+\right],$$

a direct application of (6) with $\log M_{t,t+n} = \sum_{i=1}^n \log m_{t+i}$. The expression x^+ means the positive part of x, $\max(0,x)$. The result is

$$c_t^{\tau,n} \;=\; b_t^{\tau+n}N(d) - kb_t^\tau N(d - v_{\tau,n}), \tag{38}$$

where N is the cumulative normal distribution function and

$$\begin{aligned} d &= \frac{\log[b_t^{\tau+n}/(b_t^\tau k)] + v_{\tau,n}^2/2}{v_{\tau,n}} \\ v_{\tau,n}^2 &= Var_t(\log b_{t+\tau}^n) \;=\; \left(\frac{1-\varphi^{2\tau}}{1-\varphi^2}\right)\left(\frac{1-\varphi^n}{1-\varphi}\right)^2\sigma^2. \end{aligned} \tag{39}$$

The tedious details of this calculation have been worked out in a number of places, including Backus, Foresi, and Zin (1998, Appendix A.4).

Equation (39) reminds us that volatility $v_{\tau,n}$ is a two-dimensional array. Practitioners treat it as such, most commonly in a matrix of swaption volatilities. To us, the most interesting feature of (39) is the role of the mean reversion parameter. If φ were one, volatility would be

$$v_{\tau,n}^2 \;=\; \tau\,(n\sigma)^2.$$

As in the original version of the Black-Scholes formula, volatility squared is proportional to the maturity τ of the option. The more complex form of (39) stems from two distinct roles played by the mean reversion parameter φ in determining prices of long bonds in the Vasicek model. Mean reversion appears, first, in the impact of short rate innovations on future short rates:

$$Var_t(z_{t+\tau}) \;=\; \sigma^2\left(1+\varphi^2+\varphi^4+\cdots\varphi^{2(\tau-1)}\right) \;=\; \sigma^2\left(\frac{1-\varphi^{2\tau}}{1-\varphi^2}\right),$$

a direct implication of (7). The second role of mean reversion concerns the impact of short rate movements on long bond prices. In the Vasicek model, a unit fall in the short rate is associated with a rise in the logarithm of the n-period bond price of $(1 + \varphi + \cdots + \varphi^{n-1}) = (1 - \varphi^n)/(1 - \varphi)$, which follows (with some effort) from (12). Thus mean reversion attenuates the impact of short rate innovations on long bond prices, an implication we see in the declining volatilities of yield changes with maturity (Table 2B). These patterns come from a relatively simple model, but they illustrate the challenges facing a practitioner who would like to value options that vary across both maturities (option and bond). The details depend on the model, but any internally consistent model will place restrictions on the two-dimensional array of option volatilities.

Other theoretical settings introduce additional issues, including departures from Black-Scholes associated with jumps (as in Das's jump model) and stochastic volatility (the CIR and multifactor affine models). We leave these issues for another time and place.

10 Final Remarks

We have applied a single theoretical approach to a number of bond pricing models, illustrating their solution and the choice of parameter values. The approach has two elements, discrete time and a pricing kernel, neither of which is original to us. The theory of Section 3 is reviewed in Duffie (1992). Applications to bond pricing include Campbell, Lo, and MacKinlay (1997), Sun (1992), and Turnbull and Milne (1991), as well as our work with various coauthors. The models themselves are largely examples from the affine class, whose structure was characterized in continuous-time by Duffie and Kan (1996). We think the affine class of models holds great promise for practitioners. Notably, the number of parameters is linear in both the maturity of the assets and the number of factors, which suggests substantial computational savings over binomial models.

Our catalog of models might leave a practitioner in despair at the range of choices. The choice of models must depend, we think, on the use to which the model is put. If one would like to value bonds, swaps, and short-dated options on them, a relatively simple model might suffice. Practical application will probably dictate that the use of time- or maturity-dependent drift and volatility parameters, perhaps as outlined in our summaries of Ho and Lee (1986) or Heath, Jarrow, and Morton (1992). If one would like to value options on fixed income instruments over a wide range of maturities, perhaps including options on spreads, the benefits of multifactor models may outweigh the increase in complexity. For out-of-the-money options, jumps and stochastic volatility may play a role. Such models are complicated, but in our view the complications are demanded by the complexity of modern financial markets. Good luck!

References

Backus, David, Silverio Foresi, Kai Li, and Liuren Wu, 1998, "Accounting for biases in Black-Scholes," manuscript, Stern School of Business, New York University.

Backus, David, Silverio Foresi, Abon Mozumdar, and Liuren Wu, 1998, "Predictable changes in yields and forward rates," NBER Working Paper No. 6379, January.

Backus, David, Silverio Foresi, and Chris Telmer, 1996, "Affine models of currency pricing," NBER Working Paper No. 5623, June.

Backus, David, Silverio Foresi, and Stanley Zin, 1998, "Arbitrage opportunities in arbitrage-free models of bond pricing," *Journal of Business and Economic Statistics* 16, 13-26.

Balduzzi, Pierluigi, Sanjiv Das, and Silverio Foresi, 1998, "The central tendency: A second factor in bond yields," *Review of Economics and Statistics* 80, 60-72.

Balduzzi, Pierluigi, Sanjiv Das, Silverio Foresi, and Rangarajan Sundaram, 1996, "A simple approach to three-factor affine term structure models," *Journal of Fixed Income* 6 (December), 43-53.

Black, Fischer, Emanuel Derman, and William Toy, 1990, "A one-factor model of interest rates and its application to treasury bond options," *Financial Analysts Journal* 46 (Jan-Feb), 33-39.

Brennan, Michael, and Eduardo Schwartz, 1979, "A continuous time approach to the pricing of bonds," *Journal of Banking and Finance* 3, 133-155.

Campbell, John, Andrew Lo, and Craig MacKinlay, 1996, *The Econometrics of Financial Markets*, Princeton NJ: Princeton University Press.

Chen, Ren-Raw and Louis Scott, 1993, "Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates," *Journal of Fixed Income* 3 (December), 14-31.

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

Dai, Qiang, and Kenneth Singleton, 1997, "Specification analysis of affine term structure models," manuscript, Stanford University, February.

Das, Sanjiv, 1994, "Bond and option pricing for jump-diffusion processes," manuscript, Harvard Business School.

Das, Sanjiv, and Silverio Foresi, 1996, "Exact solutions for bond and option prices with systematic jump risk," *Review of Derivatives Research* 1, 7-24.

Duffie, Darrell, 1992, *Dynamic Asset Pricing Theory*, Princeton, NJ: Princeton University Press.

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

Duffie, Darrell, and Kenneth Singleton, 1997, "An econometric model of the term structure of interest rate swap yields," *Journal of Finance* 52, 1287-1321.

Gibbons, Michael, and Krishna Ramaswamy, 1993, "A test of the Cox, Ingersoll, Ross model of the term structure," *Review of Financial Studies* 6, 619-658.

Harvey, Andrew, 1989, *Forecasting, Structural Time Series Models, and the Kalman Filter*, New York: Cambridge University Press.

Harvey, Andrew, 1993, *Time Series Models (Second Edition)*, Cambridge, MA: MIT Press.

Heath, David, Robert Jarrow, and Andrew Morton, 1992, "Bond pricing and the term structure of interest rates," *Econometrica* 60, 225-262.

Ho, Thomas S.Y., and Sang-Bin Lee, 1986, "Term structure movements and pricing interest rate contingent claims," *Journal of Finance* 41, 1011-1029.

Hull, John, and Alan White, 1990, "Pricing interest-rate-derivative securities," *Review of Financial Studies* 3, 573-592.

Hull, John, and Alan White, 1993, "One-factor interest-rate models and the valuation of interest-rate derivative securities," *Journal of Financial and Quantitative Analysis* 28, 235-254.

Longstaff, Francis, and Eduardo Schwartz, 1992, "Interest rate volatility and the term structure: A two-factor general equilibrium model," *Journal of Finance* 47, 1259-1282.

McCulloch, J. Huston, and Heon-Chul Kwon, 1993, "US term structure data, 1947-91," manuscript and computer diskettes, Ohio State University.

Pearson, Neil, and Tong-Sheng Sun, 1994, "Exploiting the conditional density in estimating the term structure: An application to the Cox, Ingersoll, and Ross model," *Journal of Finance* 54, 1279-1304.

Sun, Tong-sheng, 1992, "Real and nominal interest rates: A discrete-time model and its continuous-time limit," *Review of Financial Studies* 5, 581-611.

Turnbull, Stuart, and Frank Milne, 1991, "A simple approach to interest-rate option pricing," *Review of Financial Studies* 5, 87-120.

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

Table 1
Properties of US Government Bond Yields

Maturity	Mean	St Dev	Skewness	Kurtosis	Auto
1 month	5.314	3.064	0.886	0.789	0.976
3 months	5.640	3.143	0.858	0.691	0.981
6 months	5.884	3.178	0.809	0.574	0.982
9 months	6.003	3.182	0.776	0.480	0.982
12 months	6.079	3.168	0.730	0.315	0.983
24 months	6.272	3.124	0.660	0.086	0.986
36 months	6.386	3.087	0.621	−0.066	0.988
48 months	6.467	3.069	0.612	−0.125	0.989
60 months	6.531	3.056	0.599	−0.200	0.990
84 months	6.624	3.043	0.570	−0.349	0.991
120 months	6.683	3.013	0.532	−0.477	0.992

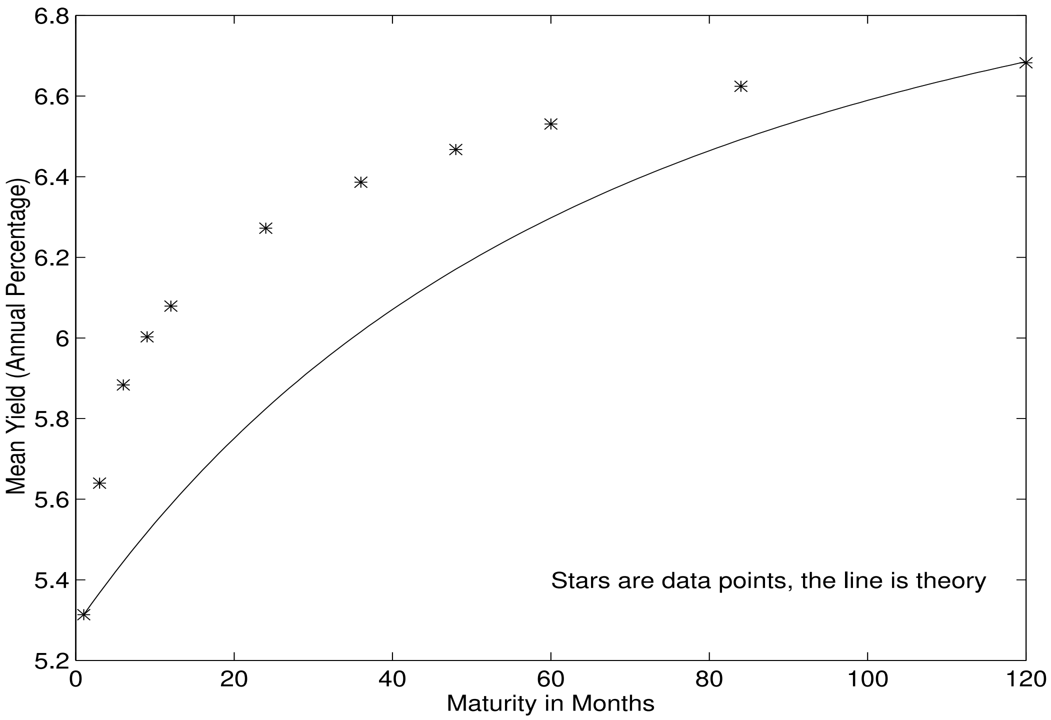
The data are monthly estimates of annualized continuously-compounded zero-coupon US government bond yields computed by McCulloch and Kwon (1993), January 1952 to February 1991 (470 observations). Mean is the sample mean, St Dev the sample standard deviation, Skewness an estimate of the skewness measure γ_1, Kurtosis an estimate of the (excess) kurtosis measure γ_2, and Auto an estimate of the first autocorrelation. The skewness and kurtosis measures are defined, specifically, in terms of central moments μ_j: $\gamma_1 = \mu_3/\mu_2^{3/2}$ and $\gamma_2 = \mu_4/\mu_2^2 - 3$. Both are zero for normal random variables. Our estimates replace population moments with sample moments.

Table 2
Properties of Yield Spreads and Monthly Changes in Yields

Maturity	Mean	St Dev	Skewness	Kurtosis	Auto
		A. Spreads Over Short Rate			
3 months	0.326	0.303	2.036	7.079	0.353
6 months	0.570	0.437	1.457	5.350	0.556
9 months	0.689	0.521	1.362	5.032	0.630
12 months	0.765	0.593	1.271	4.964	0.686
24 months	0.959	0.796	0.531	2.606	0.793
36 months	1.073	0.927	0.275	1.988	0.831
48 months	1.154	1.011	0.098	1.554	0.851
60 months	1.217	1.078	0.032	1.333	0.864
84 months	1.305	1.178	−0.001	1.092	0.879
120 months	1.369	1.237	−0.087	0.815	0.885
		B. Monthly Changes in Yields			
1 month	0.008	0.644	−1.172	10.224	0.023
3 months	0.009	0.575	−1.751	14.008	0.110
6 months	0.009	0.570	−1.619	15.618	0.150
9 months	0.009	0.571	−1.240	14.680	0.148
12 months	0.010	0.547	−0.783	12.824	0.152
24 months	0.011	0.487	−0.398	11.474	0.132
36 months	0.011	0.441	−0.032	8.128	0.100
48 months	0.011	0.409	0.052	6.359	0.087
60 months	0.011	0.382	0.077	5.142	0.077
84 months	0.012	0.340	0.040	3.548	0.069
120 months	0.012	0.309	−0.205	3.288	0.068

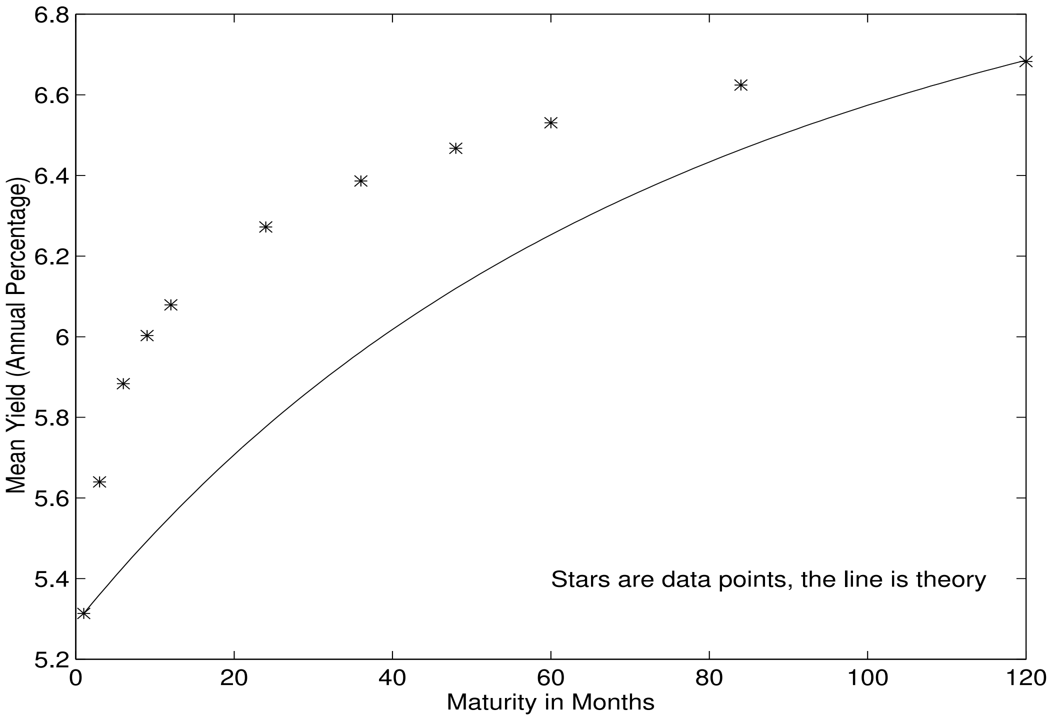
See Table 1 for notes.

Figure 1
Mean Yields in the Vasicek Model



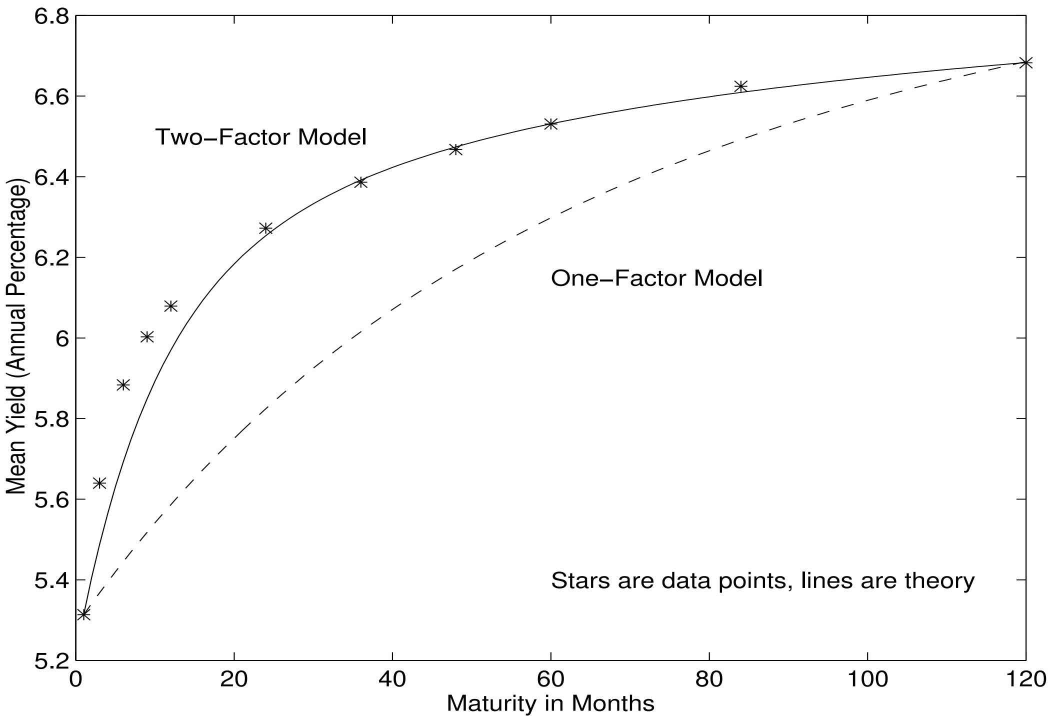
Asterisks are mean yields on US treasury securities, as reported in Table 1. The line represents mean yields in the Vasicek model using parameter values reported in the text.

Figure 2
Mean Yields in the Cox-Ingersoll-Ross Model



Asterisks are mean yields on US treasury securities, as reported in Table 1. The line represents mean yields in the Cox-Ingersoll-Ross model using parameter values reported in the text.

Figure 3
Mean Yields in One- And Two-Factor Vasicek Models



Asterisks are mean yields on US treasury securities, as reported in Table 1. The lines represents mean yields in the one-factor (dashed line) and two-factor (solid line) Vasicek models using parameter values reported in the text.