NBER WORKING PAPER SERIES

REVERSE ENGINEERING THE YIELD CURVE

David K. Backus
Stanley E. Zin

Working Paper No. 4676

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 1994

The authors thank Ned Elton, Stephen Figlewski, Silverio Foresi, James MacKinnon, Fallaw Sowell, and Chris Telmer for helpful comments and suggestions. Backus thanks the National Science Foundation for financial support. This paper is part of NBER's research program in Asset Pricing. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #4676
March 1994

REVERSE ENGINEERING THE YIELD CURVE

ABSTRACT

Prices of riskfree bonds in any arbitrage-free environment are governed by a pricing kernel: given a kernel, we can compute prices of bonds of any maturity we like. We use observed prices of multi-period bonds to estimate, in a log-linear theoretical setting, the pricing kernel that gave rise to them. The high-order dynamics of our estimated kernel help to explain why first-order, one-factor models of the term structure have had difficulty reconciling the shape of the yield curve with the persistence of the short rate. We use the estimated kernel to provide a new perspective on Hansen-Jagannathan bounds, the price of risk, and the pricing of bond options and futures.

David K. Backus
Stern School of Business
Management Education Center
New York University
New York, NY 10012-1126
and NBER

Stanley E. Zin
Graduate School of Industrial Administration
Carnegie-Mellon University
Pittsburgh, PA 15213-3890
and NBER

1 Introduction

A newcomer to the theory of bond pricing would be struck by the enormous variety of models used by academics and practitioners alike. Prominent examples include Black, Derman, and Toy (1990), Brennan and Schwartz (1979), Cox, Ingersoll, and Ross (1985), Heath, Jarrow, and Morton (1992), Ho and Lee (1986), and Vasicek (1977), and one could easily add many others to the list. The common element linking these diverse theories is what Sargent (1987) terms the *pricing kernel*: the stochastic process governing prices of state-contingent claims. A theory of bond prices that does not admit pure arbitrage opportunities implies that a pricing kernel exists. Hence a theory of bond prices is essentially a choice of pricing kernel.

Given a pricing kernel, we can compute prices of bonds and related assets. We describe and implement a procedure for doing the reverse: of using prices of riskfree government bonds to deduce the pricing kernel. We do this in a discrete-time log-linear theoretical framework that has been used by, among others, Campbell (1986), den Haan (1993), and Turnbull and Milne (1991), and is closely related to the one-factor Gaussian interest rate model of Jamshidian (1989) and Vasicek (1977). This framework does not encompass all of the models in the literature, or even all of those listed above, but its simple structure makes the relation between bond yields and the pricing kernel relatively transparent. Given this structure, we can trace observed properties of bond prices to properties of the kernel. We show, for example, that the upward slope of the mean yield curve and the positive autocorrelation of interest rates provide information about the pricing kernel that bear on the pricing of related assets.

Our estimated pricing kernel provides a useful perspective on both the general equilibrium foundations of asset pricing and the properties of interest-rate derivative assets. With respect to the foundations of asset pricing, the widely-documented discrepancies between representative agent theories and observed asset prices have been linked, most notably by Hansen and Jagannathan (1991), to variability of the pricing kernel. Like them, we find that the pricing kernel has substantially greater variability than theory based on a representative agent with power utility. We find, in addition, that most of this variability is short term. With respect to derivative assets, our estimated kernel provides some insight into their pricing. One example concerns bond options. In our theoretical framework, an increasing mean yield curve implies mean reversion in the kernel and the short rate. As a result, approaches to option pricing that do not incorporate mean reversion must compensate by positing a declining term structure of volatility. A second example concerns the relation between forward and futures prices. Although these prices need not bear any particular relation to each other in our framework, our estimated kernel implies that futures prices are less than forward prices, as they seem to be in the data. This property is a general consequence, in our framework, of kernels that imply a positively auto-correlated short rate, and holds even when the mean yield curve is decreasing. These examples illustrate, we think, the benefits of reducing the theory of bond pricing to its least common denominator, the pricing kernel.

We develop these issues in the following pages, starting with a quick look at the salient properties of bond yields and forward rates.

2 Properties of Bond Yields

We review two features of yields and forward rates on US government bonds: the tendency for long rates to be higher than short rates, on average, and the high degree of persistence in yields and forward rates of all maturities. The data are monthly, and were constructed by McCulloch and Kwon (1993).

To fix the notation, let b_t^n denote the dollar *price* at date t of an n-period discount bond: the claim to one dollar in all states at date $t+n$. The dollar *yield* on a bond of maturity n, for $n > 0$, is

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

The yield on a one-period bond is simply the short rate: $r_t \equiv y_t^1 = -\log b_t^1$. *Forward rates* are implicit in the prices of n- and $(n+1)$-period bonds:

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

From definitions (1) and (2) we see that yields are averages of forward rates:

$$y_t^n = n^{-1} \sum_{j=1}^{n} f_t^{j-1}. \tag{3}$$

Thus we can express the maturity structure of riskfree bonds in three equivalent ways: with prices, yields, or forward rates.

The most common method of representing the term structure is with yields, to which we turn now. The first feature of interest is the slope of the yield curve.

We see in Table 1 and Figure 1 that the yield curve has been, on average, upward sloping. This characterization holds for the postwar period as a whole (1952-91) and for a shorter recent period (1982-91). Certainly yields have been higher over the last decade than they were earlier, but there has been no qualitative change in their tendency to increase with maturity.

The second feature we want to stress is the high degree of persistence in bond yields. The first-order autocorrelation, reported in Table 1, is at least 0.9 for all maturities. For the period as a whole, the autocorrelation of the one-month yield is 0.976, and for the last decade 0.906. This pattern of somewhat smaller autocorrelations over the recent past applies to all maturities. The autocorrelations are similar across maturities, but increase slightly with maturity in both sample periods. Some authors have interpreted the high degree of persistence as suggesting the possibility of a unit root in the short rate. Chan et al. (1992, p 1217), for example, note that there is "only weak evidence of mean reversion." Perhaps for this reason, the theoretical models of Ho and Lee (1986) and Black, Derman, and Toy (1990) start with unit roots in, respectively, the short rate and its logarithm.

Since the theory is expressed most simply with forward rates, we report analogous properties for them in the second panel of Table 1. Both features of bond yields show up in forward rates, too. One minor exception is the slope of the mean forward rate curve: there is a modest nonmonotonicity between five and ten years. The departures from monotonocity are small, however, relative to sampling variability and the accuracy of the forward rate data.

These properties of bond yields and forward rates will come as no surprise, but they serve as useful guides in thinking about the theory that follows.

3 A Theoretical Framework

Our theoretical framework starts with the pricing kernel: the stochastic process for state-contingent claims prices. Among the many notable applications of this approach to bond pricing, real and nominal, are Cox, Ingersoll, and Ross (1985), Constantinides (1992), Duffie (1993), and Longstaff and Schwartz (1992). Our discrete-time, log-linear structure builds on earlier work by Campbell (1986), den Haan (1993), and Turnbull and Milne (1991).

The pricing kernel is a stochastic process for a positive random variable m satisfying

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

for all maturities n, where $R_{t+1}^{n+1} = b_{t+1}^n / b_t^{n+1}$ is the one-period (gross) return on an $(n+1)$-period bond and E_t denotes the expectation conditional on the date-t information set, which includes the history of the pricing kernel. In representative agent economies, m is the nominal intertemporal marginal rate of substitution and (4) is a first-order condition. More generally, there exists a positive random variable m satisfying (4) if the economy admits no pure arbitrage opportunities. See, for example, the discussion in Duffie (1993, Section 1A). More important for our purposes, equation (4) allows us to price bonds recursively:

$$b_t^{n+1} = E_t(m_{t+1} b_{t+1}^n). \tag{5}$$

By convention $b_t^0 = 1$ (a dollar today costs one dollar).

The pricing relation (5) becomes a theory of bond prices once we characterize the pricing kernel m. A convenient and tractable choice is the infinite moving average,

$$-\log m_t = \delta + \sum_{j=0}^{\infty} \alpha_j \epsilon_{t-j}, \tag{6}$$

for $\{\epsilon_t\}$ normally and independently distributed with mean 0 and variance σ^2. The logarithm guarantees that m is positive; the negative sign is chosen to produce simple expressions for interest rates. We normalize by setting $\alpha_0 = 1$, so that σ^2 is the variance of the innovation ϵ. Stationarity requires that the coefficients be square summable: $\sum_{j=0}^{\infty} \alpha_j^2 < \infty$. The dynamics of m are governed by the moving average coefficients $\{\alpha_j\}$. In most respects, this specification of the pricing kernel is simply a translation into discrete time of the one-factor Gaussian interest rate model. The difference is that our formulation allows more complex interest rate dynamics than the diffusions used by Jamshidian (1989) and Vasicek (1977). We will see later that this difference is essential.

We are now in a position to compute bond prices. We show, by induction, that bond prices can be expressed in log-linear form as

$$-\log b_t^n = \mu^n + \sum_{j=0}^{\infty} \beta_j^n \epsilon_{t-j}, \tag{7}$$

for some parameters $\{\mu^n, \beta_j^n\}$. For $n = 0$ we have $\log b_t^0 = 0$, so

$$\mu^0 = \beta_j^0 = 0, \tag{8}$$

for all $j \geq 0$. Given the price function (7) of an n-period bond, we use the pricing relation (5) to compute the parameters of the price function of an $(n+1)$-period bond. The first step is to evaluate the conditional expectation in (5). Note that

$$\log b_{t+1}^n + \log m_{t+1} = -(\delta + \mu^n) - \sum_{j=0}^{\infty}(\alpha_j + \beta_j^n)\epsilon_{t+1-j}.$$

The conditional mean and variance of this expression are

$$E_t(\log b_{t+1}^n + \log m_{t+1}) = -(\delta + \mu^n) - \sum_{j=0}^{\infty}(\alpha_{j+1} + \beta_{j+1}^n)\epsilon_{t-j}$$

and

$$Var_t(\log b_{t+1}^n + \log m_{t+1}) = (\alpha_0 + \beta_0^n)^2\sigma^2.$$

We now apply a property of expectations of log-normal random variables: if $\log x$ is normal with mean μ and variance σ^2, then $\log E(x) = \mu + \sigma^2/2$. Applying this formula to the pricing relation (5) and collecting terms, we find that the $(n+1)$-period bond price function has the infinite moving average form (7) with parameters

$$\mu^{n+1} = \mu^n + \delta - (\alpha_0 + \beta_0^n)^2\sigma^2/2 \tag{9}$$

$$\beta_j^{n+1} = \beta_{j+1}^n + \alpha_{j+1}, \tag{10}$$

for $n, j \geq 0$. Evaluating the bond price functions, then, is simply a matter of running through the recursions (9,10), starting with the initial conditions (8).

As in Heath, Jarrow and Morton (1992), the pricing formulas are simpler for forward rates than for bond prices. We start by defining partial sums of moving average coefficients,

$$A_n = \sum_{j=0}^{n} \alpha_j.$$

From the recursion (10) and the initial value (8), we see that the moving average coefficients for the bond price function are

$$\beta_j^n = \sum_{i=1}^{n} \alpha_{j+i} = A_{n+j} - A_j. \tag{11}$$

Since $\beta_0^n = A_n - A_0$, the intercept in the bond price function is

$$\mu^n = n\delta - (\sigma^2/2)\sum_{j=1}^{n} A_{j-1}^2.$$

Thus bond price functions are governed by the partial sums of moving average coefficients, A_n.

Turning to forward rates, we see from equations (2,7) that they can be expressed

$$f_t^n = (\mu^{n+1} - \mu^n) + \sum_{j=0}^{\infty}(\beta_j^{n+1} - \beta_j^n)\epsilon_{t-j}.$$

From (9,11) we find that the intercept in the forward rate function is

$$\mu^{n+1} - \mu^n = \delta - A_n^2\sigma^2/2$$

and the moving average coefficients are

$$\beta_j^{n+1} - \beta_j^n = A_{n+j+1} - A_{n+j} = \alpha_{n+j+1}.$$

Forward rates are summarized, then, by

Proposition 1 *Forward rates in this economy have the infinite moving average representation*

$$f_t^n = \delta - A_n^2\sigma^2/2 + \sum_{j=0}^{\infty}\alpha_{n+1+j}\epsilon_{t-j}, \tag{12}$$

with parameters $\{\delta, \sigma, \alpha_j\}$ *of the pricing kernel (6) and* $A_n \equiv \sum_{j=0}^{n}\alpha_j$.

Proposition 1 characterizes forward rates in this economy. From it, we can construct prices and yields on bonds of any maturity. Yields, for example, are averages of forward rates [see (3)],

$$y_t^n = \delta - (\sigma^2/2n)\sum_{j=1}^{n} A_{j-1}^2 + n^{-1}\sum_{j=0}^{\infty}(A_{n+j} - A_j)\epsilon_{t-j}, \tag{13}$$

and bond prices are simple functions of yields,

$$-\log b_t^n = n\delta - (\sigma^2/2)\sum_{j=1}^{n} A_{j-1}^2 + \sum_{j=0}^{\infty}(A_{n+j} - A_j)\epsilon_{t-j}. \tag{14}$$

Thus, if we know the parameters of the pricing kernel we can compute bond prices, yields, and forward rates. Our plan is to do the reverse, and use observations of bond prices to uncover the parameters $\{\delta, \sigma, \alpha_j\}$ of the kernel.

Remarks. 1. Although the state is infinite dimensional, in the sense that bond prices depend on the infinite history of ϵ's, in another sense this is really a one-factor model. One-period returns on bonds of all maturities are (except for a small nonlinearity) perfectly correlated, since the only source of uncertainty is next period's innovation. Our framework, however, allows more general dynamics than most other one-factor models. Multifactor extensions are considered in Appendix B. 2. Proposition 1 can be proved more directly by attacking forward rates instead of bond prices (see Appendix A). The cost is that we must start with a somewhat less intuitive pricing relation for forward contracts.

4 Reverse Engineering 1: An Example

One of the intriguing features of bond yields is the close relation between their "time series" and "cross section" properties: the pricing kernel dictates both the dynamic behavior of interest rates and the slope of the yield curve at a point in time. In this section we describe how time series and cross section information can be used to characterize the underlying pricing kernel. Both are necessary: the parameters of the kernel cannot be inferred from the time series or cross section of bond prices alone. An example illustrates in a more concrete setting how both kinds of information can be used to estimate the parameters of the kernel.

One source of information about the pricing kernel is dynamic properties of interest rates, like autocorrelations and autocovariances. The short rate, for example, is

$$r_t = f_t^0 = \delta - \sigma^2/2 + \sum_{j=0}^{\infty} \alpha_{j+1}\epsilon_{t-j},$$

so its autocorrelations are determined by $\{\alpha_1, \alpha_2, \ldots\}$. These moving average coefficients are those of the pricing kernel shifted over one position, so we lose one piece of information about the kernel when we look at the short rate. As we move up the forward rate curve, we lose additional moving average coefficients [see (12)], so the dynamics of the short rate are more informative than the dynamics of longer-maturity forward rates. The time series evidence, however, is not sufficient: the variance σ^2 of innovations to the pricing kernel cannot be identified from time series alone.

A second source of information about the pricing kernel is the cross section of bond yields, exemplified by the yield curve at a point in time or the mean yield or forward rate curve. The mean forward rate curve, for example, is

$$E(f^n) = \delta - A_n^2\sigma^2/2,$$

and the mean yield curve is

$$E(y^n) = \delta - (\sigma^2/2n)\sum_{j=1}^{n} A_{j-1}^2. \tag{15}$$

Information on mean yields or forward rates, however, is not generally sufficient to identify the parameters of the kernel. One reason is that mean forward rates depend on squared partial sums, A_n^2, and a given sequence of squared partial sums can be generated by more than one sequence of moving average coefficients. The two sequences of moving average coefficients,

$$\{1, -.5, -.3, -.1, \ldots\} \text{ and } \{1, -1.5, .7, -.1, \ldots\},$$

both produce the same values of A_n^2, so they generate the same mean forward rate curve. Another reason why cross section evidence is not enough is that n mean forward rates depend on $n+2$ parameters: δ, σ, and the first n α's. We gain one degree of freedom from the normalization $\alpha_0 = 1$, but that leaves us short one piece of information. The cross section of forward rates, then, like the dynamics of the short rate, is insufficient to identify the pricing kernel. The kernel can generally be identified, however, by a combination of time series and cross section evidence, which we do shortly.

Although the cross section evidence cannot generally tell us all of the parameters of the kernel, the slope of the mean forward rate curve places limits on the dynamics of the short rate. To produce an increasing mean forward rate curve, the parameters must satisfy

$$A_0^2 \geq A_1^2 \geq A_2^2 \geq \cdots \geq A_n^2 \geq \cdots. \tag{16}$$

This implies, by itself, mean reversion in the short rate. This follows since an increasing mean forward rate curve requires the positive sequence $\{A_n^2\}$ to be decreasing, so it must converge. This rules out pricing kernels whose partial sums diverge, including the random walk model of Ho and Lee (1986) and our own fractional difference model (Backus and Zin, 1993). In the random walk model the moving average coefficients are $\alpha_j = \alpha_1$ for $j \geq 1$, so that $A_n = 1 + n\alpha_1$. As long as α_1 is nonzero, the squared partial sum A_n^2 eventually grows without bound as n increases, and the mean yield and forward rate curves eventually decline with maturity. Unless this effect appears at maturities greater than ten years, the random walk model is inconsistent with our observation (in Table 1 and Figure 1) of increasing mean yield and forward rate curves. Dybvig (1989) makes a related point about the Ho and Lee model and suggests an alternative that exhibits mean reversion.

An example illustrates how the time series and cross section properties of bond yields might be combined to estimate the parameters of the pricing kernel. We approach this estimation problem formally in the next section, but its logic is apparent from an informal moment matching exercise. Let us say that the pricing kernel is ARMA(1,1):

$$-\log m_t = (1-\varphi)\delta - \varphi \log m_{t-1} + \epsilon_t + \theta\epsilon_{t-1}.$$

The moving average coefficients are then $\alpha_0 = 1$, $\alpha_1 = \varphi + \theta$, and $\alpha_j = (\varphi + \theta)\varphi^{j-1}$ for $j > 1$, so the coefficients are square summable if $|\varphi| < 1$. The nice feature of this example is that it delivers the first-order autoregressive short rate studied by Turnbull and Milne (1991), Vasicek (1977), and others. The short rate for this example is

$$r_t = (1-\varphi)(\delta - \sigma^2/2) + \varphi r_{t-1} + (\varphi + \theta)\epsilon_t. \tag{17}$$

An estimate of this equation over the same sample period as Panel A includes the autoregressive parameter, $\varphi = 0.976$, and an estimate of the innovation variance,

$$Var[(\varphi + \theta)\epsilon] = (\varphi + \theta)^2\sigma^2 = 0.000534^2.$$

The time series evidence thus gives us two pieces of information that we can use to estimate the four parameters of the pricing kernel.

To identify the parameters of the kernel we need cross section evidence as well. In this example we use mean bond yields, which we've seen depend on the squared partial sums of moving average coefficients. In the ARMA(1,1) example the partial sums take the form

$$A_n = 1 + (\varphi + \theta)(1 - \varphi^n)/(1 - \varphi), \tag{18}$$

for $n \geq 0$. The mean short rate is (see Table 1)

$$E(r) = \delta - \sigma^2/2 = 5.314/1200,$$

which is one of the required pieces of information. Each long yield gives us an additional piece of information, so in this example the mean yield on any long bond enables us to estimate the four parameters of the kernel.

We combine the time series and cross section evidence informally using a graphical technique we refer to as the Quattro Method of Moments. Given a value for θ, the time series information allows us to compute φ and σ from

$$\begin{aligned} \varphi &= 0.976 \\ \sigma &= 0.000534/|\varphi + \theta|, \end{aligned}$$

and the mean short rate gives us

$$\delta = 5.314/1200 + \sigma^2/2.$$

The final step is to choose θ to bring the theory "close" to observed mean yields on long bonds, with the other parameters adjusting to satisfy the first three conditions. We experiment with different values of θ until the theoretical mean yield curve looks similar to the data when we graph the two together. The result is pictured in Figure 2A, where the black squares represent cross section evidence (the mean yields reported in Table 1) and the line represents theory (equation [15]). The parameter values are $\varphi = 0.976$, $\theta = -0.982$, $\sigma = 0.0890$, and $\delta = 0.00839$.

We see in Figure 2A that this procedure provides only a rough approximation to the data. We can replicate the steep slope of the short end of the yield curve with smaller values of θ or the flatness of the long end with larger values, but we cannot do both at the same time. The chosen value of θ leans toward the latter of these two objectives. This difficulty is reduced if we use moments from the 1982-91 period in our exercise. For example, with $\varphi = 0.906$, $\theta = -0.9081$, $\sigma = 0.30704$, and $\delta = 0.05337$ the mean yield curve has greater curvature, and we see in Figure 2B that the implied mean yield curve is quite close to the data. The critical difference here is the choice of φ. For the complete sample period, there is tension between φ's role in determining the autocorrelation of the short rate (indicating a high value) and its role in determining the curvature of the yield curve (indicating a low value). If we choose $\varphi = 1$, so that the short rate is a random walk, the match between theory and data is worse: With some values of θ the mean yield curve is even less concave than with $\varphi = 0.976$, and with others it slopes down over the relevant range of maturities.

Our method of estimating the parameters of the pricing kernel differs in two respects from the procedure pioneered by Ho and Lee (1986) and widely adopted by practitioners. In the Ho and Lee procedure, what we have called the time series information consists of an autoregressive parameter for the short rate and a volatility parameter for short rate innovations. Ho and Lee also introduce time-dependent drift in the short rate. They set the autoregressive parameter equal to one (so that short rate changes are independent) and the volatility and drift parameters are inferred from cross section information: prices of long bonds and derivative assets, like bond options, at a point in time. [This is implicit in their work, explicit in later treatments (Duffie 1993, Exercise 3.12, for example)]. Thus the Ho and Lee procedure differs, first, in disregarding time series evidence on the short rate and, second, in adding time-dependent drift parameters that can be used to match observed asset prices more closely.

This procedure's widespread use suggests that it works well in practice – indeed, the additional time-dependent parameters allow the model to replicate an entire yield curve exactly, which our model does not (the rough fit of Figure 2A being an example). We worry, though, that the additional parameters may mask weaknesses in the theory's foundations, particularly regarding the implied dynamics of the short rate. We've seen, for example, that with a random walk short rate the theory produces insufficiently concave, and eventually decreasing, mean yield curves, for which the time-dependent drift parameters compensate. Our procedure, on the other hand, should give us a clearer picture of the impact of interest rate dynamics on prices of bonds and related assets. With luck this will lead to future improvements in the pricing of interest-rate derivative securities.

5 Reverse Engineering 2: Estimation

In the previous section we used an informal procedure to estimate the parameters of an ARMA(1,1) pricing kernel. Here we extend the theory to finite ARMA pricing kernels and estimate their parameters by the generalized method of moments, or GMM (Hansen, 1982).

Finite ARMA pricing kernels are those that can be expressed in the form

$$-\Phi(L)\log m_t = \Phi(1)\delta + \Theta(L)\epsilon_t, \tag{19}$$

with autoregressive and moving average polynomials

$$\Phi(L) = 1 - \varphi_1 L - \cdots - \varphi_p L^p$$

and

$$\Theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q,$$

respectively, for finite nonnegative integers p and q. As in statistical time series analysis, the hope is that a model of this kind will be able to approximate a wide range of dynamic behavior with a small number of parameters. In Appendix C we describe the relation between short rate dynamics and the form of the pricing kernel.

Our estimation procedure chooses the parameters of an ARMA pricing kernel to make the theory "close" to the data in a well-defined statistical sense. As in the ARMA(1,1) example, we use both time series and cross section information to estimate the parameters of the kernel. We summarize the time series behavior of interest rates with a subset of the autocovariances of the short rate,

$$Cov(r_t, r_{t+k}) = \sigma^2 \sum_{j=1}^{\infty} \alpha_j \alpha_{j+k},$$

for $k = 0, 1, 3, 12, 24$ months. Differences between the autocovariances of this formula and sample estimates give us five moment conditions. There is enough smoothness in the autocovariance function that these five moments capture a large part of the dynamics of the short rate. And, as we have seen, autocovariances of long rates contain no additional information regarding the dynamics of the pricing kernel. We summarize the cross section of bond yields with mean yield spreads,

$$E(y^n - y^1) = (A_0 - n^{-1} \sum_{j=1}^{n} A_{j-1}^2)\sigma^2/2,$$

for maturities $n = 3, 12, 36, 60, 120$ months. Differences between the expressions on the right side of the equation and sample means of yield spreads give us five more moment conditions. These 10 moment conditions are used to estimate the standard deviation of the kernel (σ) and the autoregressive and moving average parameters (φ_j and θ_j). The remaining parameter (δ) simply shifts the yield curve up and down by a constant, and is chosen to equate the mean short rate,

$E(r) = \delta - \sigma^2/2$, to its sample mean.

Estimated pricing kernels are reported in Table 2 for several ARMA models. The estimates are computed using a common two-step implementation of GMM. In step one, we weight each of the 10 moment conditions equally and use the estimated parameter values to construct a weighting matrix. The weighting matrix is computed by Newey and West's method with a window width of 96. We use the weighting matrix from our most general model, the ARMA(2,3), for each of the models in the table. In step two, we use the estimated weighting matrix to compute the parameter values reported in Table 2. This procedure places greater weight on those moments that are estimated precisely than on those that are not. Since we have more moment conditions than parameters, the parameters will generally not permit the pricing kernel to match all of the moments exactly, just as our informal procedure (summarized by Figure 2) did not match the mean yield curve exactly. The magnitude of these differences between theory and data is summarized by the J-statistic. In the language of sampling theory, the J-statistic for an ARMA(p,q) model is distributed asymptotically as a chi-square with $s = 10 - (p + q + 1)$ degrees of freedom if this model generated the data. A large J-statistic thus indicates that the differences between theory and data are large relative to their sampling variability.

Estimates of low-order models give us an idea of the demands the data place on the theory. The ARMA(1,1) model, for example, illustrates how time series and cross section evidence is combined by our estimation procedure. We saw in the previous section that the autocorrelation of the short rate for the full sample implied a value of φ_1 of 0.976, but that the curvature of the mean yield curve required a smaller value. Our estimated value (φ_1 is 0.7073) is a compromise between the two, which the J-statistic indicates is not liked by either. A casual look at the autoregressive and moving average parameters for this model suggests that they are equal, and that we could eliminate both from the model by cancellation in (19). In fact the small difference between the two is both economically essential (an increasing mean yield curve requires $\alpha_1 = \varphi_1 + \theta_1 < 0$) and statistically significant (the difference of -0.0051 has a standard error of 0.0006).

The J-statistics for the ARMA(2,2) and ARMA(2,3) models suggest that both are considerable improvements over the ARMA(1,1). The ARMA(2,3), in particular, appears to provide an adequate approximation to the features of bond yields captured by our ten moment conditions. Moreover, the t-statistic of 2.1 for the third moving average parameter (θ_3) indicates that this model is a useful step beyond the ARMA(2,2).

The striking improvement of moving from the ARMA(1,1) to the ARMA(2,3) kernel may help us understand why Stambaugh (1988) and Gibbons and Ramaswamy (1993), among others, found that first-order, one-factor models do not provide an inadequate description of bond yields, returns, and forward rates. Both of these studies use the Cox-Ingersoll-Ross model. Like our ARMA(1,1), the Cox-Ingersoll-Ross model implies first-order dynamics for the short rate (see Appendix D). Gibbons and Ramaswamy (1993, Section 5.3) find, as we do with the ARMA(1,1), that the degree of short rate persistence implied by cross section evidence is substantially smaller than that implied by the dynamics of the short rate (their parameter ρ corresponds to our φ_1). In our case, the tension between time series and cross section evidence is greatly reduced when we increase the order of the model. The ARMA(2,3) delivers both the concave shape of the yield curve, as we see in Figure 3, and the highly autocorrelated short rate that we see in the data (the first autocorrelation for this model is 0.938). In this respect, the high-order dynamics allowed by our theoretical structure are a useful extension of Vasicek's (1977) one-factor Gaussian interest rate model.

Our estimated ARMA(2,3) model gives us a detailed description of the pricing kernel, whose properties are easily related to those of bond yields. One feature of our estimated kernel is that the moving average coefficients α_j are negative after the initial coefficient $\alpha_0 = 1$. These coefficients, which are plotted in Figure 4, illustrate the response of the logarithm of the kernel to a unit innovation. Since the coefficients all have the same sign, the autocovariances of forward rates and yields are all positive. Since they are much smaller than one, the partial sums $A_n = \sum_{j=0}^{n} \alpha_j$ are positive and decline with maturity, thus generating increasing mean yield and forward rate curves. We see, in short, that while the moving average coefficients in the theory can take a variety of forms, the data imply a pricing kernel with much more structure to it.

6 Implications

The pricing kernel has been used, in different contexts, as a metric for assessing general equilibrium asset pricing theories and a building block for pricing fixed income securities. Our work sheds some light on each. With regard to the general equilibrium foundations of asset pricing theory, we relate the variability of our estimated pricing kernel to estimated lower bounds constructed by Hansen and Jagannathan (1991). With regard to fixed income security pricing, we examine the pricing of bond options, forwards, and futures.

Variability of the pricing kernel. One of the open issues in asset pricing theory concerns the relation between asset prices and economic fundamentals. Although general equilibrium treatments of asset prices date back at least to Merton (1973), they have yet to be taken seriously as descriptions of actual economies, and with good reason. Attempts to apply representative agent theory, for example, have produced asset prices and returns much different from those we observe. Perhaps the most elegant characterization of the discrepancy between theory and data is Hansen and Jagannathan's (1991) lower bound: observed asset returns imply substantially larger standard deviations of the pricing kernel than we get from representative agent theory with power utility and moderate risk aversion. This lower bound is a yardstick against which other theories, like Telmer's (1993) incomplete markets economy, can be measured.

Our estimated pricing kernel implies a standard deviation that can be compared to lower bounds derived by Hansen and Jagannathan (1991) and estimated by Bekaert and Hodrick (1992). Hansen and Jagannathan show that the return from a balanced portfolio strategy (the return, that is, from a strategy that requires no initial payment, like an excess return) places a lower bound on the standard deviation of the pricing kernel m. If the return from such a strategy is labeled x, then the Hansen-Jagannathan bound is

$$[Var(m)]^{1/2}/E(m) \quad \geq \quad |E(x)|/[Var(x)]^{1/2}.$$

The right side is the Sharpe ratio of the investment strategy (the ratio of the return's mean to its standard deviation), so the inequality tells us that large Sharpe ratios imply high variability of the pricing kernel. Bekaert and Hodrick (1992, esp Section IV and Table XI) report Sharpe ratios ranging from 0.004 to 0.78 for portfolios of US and foreign equity and currencies. Their data are monthly for the 1980s, so they correspond most closely to our estimates for the recent period. Using these estimates, and the log-normal structure of our theory, we find

$$[Var(m)]^{1/2}/E(m) \quad \cong \quad [Var(\log m)]^{1/2} = 1.047.$$

This number is larger than Bekaert and Hodrick's highest lower bound, but not by an enormous margin.

Our estimates of the variability of the pricing kernel are based on a more restrictive theoretical structure than those of Bekaert and Hodrick (1992), but in return for this structure we get more detailed information concerning the kernel's dynamic structure. One way of expressing the dynamic structure is the autocorrelation function. Our pricing kernel is negatively autocorrelated at all lags, but the magnitudes are small (the largest in absolute value is the first autocorrelation, -0.05). Another way of expressing the dynamics of the kernel is with its conditional variance at different time horizons:

$$Var_t(\log m_{t+n}) = \sigma^2 \sum_{j=0}^{n} \alpha_j^2.$$

As the forecast horizon n increases, this approaches the square of our estimated standard deviation. We find that our estimated parameter values imply that almost all the variation is in the first period. In this sense, the variability of the pricing kernel indicated by earlier work seems to consist primarily of short term variability. Cochrane and Hansen (1992, Section 2.7) come to a similar conclusion by a different route.

Both the variability and the dynamics of the kernel indicate large discrepancies between observed asset prices and representative agent theory with power utility. In this framework, the pricing kernel might be expressed

$$-\log m_t = \rho + \gamma \log(c_t/c_{t-1}) + \log(p_t/p_{t-1}),$$

where ρ is the agent's discount rate, c is consumption, γ is the risk aversion parameter, and p is the dollar price of the consumption good. Thus the properties of the kernel are inherited from the growth rate of consumption and the rate of inflation. Neither component accounts for the properties of our estimated kernel. Consumption growth, for example, exhibits too little variability to account for the standard deviation of our kernel, at least with reasonable values of the risk aversion parameter. Since the standard deviation of monthly consumption growth is about 0.005 (0.5 percent per month), the theory requires a risk aversion parameter of about 200 to match the standard deviation of the kernel. The standard deviation of inflation is even smaller (about 0.003 monthly), and inflation dynamics are much different from those of our estimated kernel. Since inflation is highly persistent, its conditional variance increases slowly, with most of the unconditional variation showing up far into the future. Our estimated kernel thus highlights the weaknesses of representative agent theory and provides a target at which alternative theories might aim.

The price of risk. Many treatments of dynamic asset pricing theory start with with a parameter termed the price of risk, which summarizes the relation between risk and expected return. Two notable examples are Hull (1993, ch

12), who describes this approach in a general setting, and Vasicek (1977), who applies it to a one-factor interest rate model much like ours. Our theory makes no mention of a price of risk, but we can easily derive one. We find that this helps us understand the origins of the price of risk and gives us a somewhat different perspective on it.

The standard definition of the price of risk is the Sharpe ratio: the ratio of the expected excess return on an asset to its standard deviation. For a bond with maturity $n+1$, the (net) return in our theoretical framework is $\log(b^n_{t+1}/b^{n+1}_t)$. This definition is approximate in our discrete time model, but exact in continuous time. The excess return over the short rate can be expressed (see equation [14])

$$x^{n+1}_{t+1} = (A_0^2 - A_n^2)\sigma^2/2 + (A_0 - A_n)\epsilon_{t+1},$$

which has conditional mean

$$E_t x^{n+1}_{t+1} = (A_0^2 - A_n^2)\sigma^2/2 = (A_0 - A_n)(A_0 + A_n)\sigma^2/2$$

and conditional variance

$$Var_t x^{n+1}_{t+1} = (A_0 - A_n)^2\sigma^2.$$

The analog in our model of Vasicek's price of risk (see his equation [14]) is the ratio of the mean to the standard deviation, which we label q:

$$q^{n+1} = (\sigma/2)(A_0 + A_n)\ \mathrm{sign}(A_0 - A_n)$$

for each maturity $n+1$.

Our formula for the price of risk differs in two ways from applications that treat it as a parametric constant. One difference is that it depends on maturity: bonds with different maturities have different prices of risk. We see in Figure 5 that with our estimated parameter values the range of variation in q is extremely small. A second, more fundamental difference is that the price of risk need not be positive. The required parameter values may be unrealistic, but they are nonetheless consistent with arbitrage-free bond pricing. If, however, we require the theory to produce an increasing mean forward rate curve, then $A_0 = 1 > A_n$ and the price of risk is positive. Thus the property of the kernel that delivers an increasing mean forward rate curve also delivers a positive price of risk.

The behavior of the price of risk is transparent in our ARMA(1,1) example, which like Vasicek's (1977, Section 5) popular continuous-time example has a first-order autoregressive short rate. Partial sums A_n are given by (18) and the price of risk is

$$q^{n+1} = -(\sigma/2)[2 + (\varphi + \theta)(1 - \varphi^n)/(1 - \varphi)] \operatorname{sign}(\varphi + \theta).$$

From the formula we see that q is positive when $\varphi + \theta$ is negative, which is necessary to produce an increasing mean forward rate curve.

Bond options: volatility and maturity. Practitioners commonly apply the Black-Scholes formula to bond options, at least as a first approximation, but find that with constant volatility the Black-Scholes formula either undervalues short options or overvalues long options. This is typically corrected by choosing smaller values of the volatility parameter for long options, resulting in a so-called term structure of volatility. Our theoretical framework produces exactly this effect as the result of the mean-reversion of interest rates implied by an increasing mean forward rate curve.

Our benchmark is an interpretation of the Black-Scholes formula for bond options adapted from Hull (1993, Section 15.6). Consider a European call at date t on an n-period discount bond with strike price k and expiration date $t + \tau$. Its price can be expressed

$$c_t^{\tau,n} = b_t^{n+\tau} N(d_1) - b_t^{\tau} k N(d_2),$$

where $N(\cdot)$ is the standard normal distribution function and

$$d_1 = [\log(b_t^{n+\tau}/k) - \log b_t^{\tau} + v^2\tau/2]/(v\tau^{1/2}), \qquad d_2 = d_1 - v\tau^{1/2},$$

for some volatility parameter v. For example, we might choose v to match the one-period volatility of an an n-period bond:

$$v^2 = Var_t(\log b_{t+1}^n) = (A_n - A_0)^2\sigma^2.$$

The formulas then tell us how to extend call prices to longer maturities τ. Practitioners generally find that they must use smaller values of v, so-called implied volatilities, for longer maturities τ to explain observed prices of options.

The option formula for our theoretical framework has the same form, but a different volatility parameter. The call price is the solution to

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

where $M_{t,\tau} \equiv \prod_{j=1}^{\tau} m_{t+j}$ is the τ-period pricing kernel and $x^+ \equiv \max\{0, x\}$ is the nonnegative part of x. This gives us a call price of

$$c_t^{\tau,n} = b_t^{\tau+n} N(d_1) - b_t^{\tau} k N(d_2) \tag{20}$$

with

$$d_1 = [\log(b_t^{\tau+n}/k) - \log b_t^{\tau} + \sigma_{\tau,n}^2/2]/\sigma_{\tau,n} \qquad d_2 = d_1 - \sigma_{\tau,n}$$

and

$$\sigma_{\tau,n}^2 = Var_t(\log b_{t+\tau}^n) = \sigma^2 \sum_{j=0}^{\tau-1} (A_{n+j} - A_j)^2. \tag{21}$$

Similar formulas are reported by Jamshidian (1989, equation 9) and Turnbull and Milne (1991, Theorem 1) for closely related environments.

The difference between our formula and Black-Scholes lies in the choice of volatility parameters. In our application of Black-Scholes, volatility $v\tau^{1/2}$ increases with the square root of the time to expiration. In our exact formula, volatility also increases with τ, but the rate of increase depends on the moving average coefficients $\{\alpha_j\}$ of the pricing kernel through the partial sums $A_n = \sum_{j=0}^{n} \alpha_j$. In principle, the theory can generate a wide range of volatility patterns. With a random walk short rate, as in the popular Ho and Lee (1987) model, volatility increases with the square root of τ. But if we restrict ourselves to pricing kernels that give rise to increasing mean forward rates, including those estimated in the previous section, then volatility generally increases less rapidly than $\tau^{1/2}$. We have seen that an increasing mean forward rate curve requires convergence of A_n^2. As a consequence, the volatility parameter $\sigma_{\tau,n}$ in equation (21) also converges, and the implied volatility $v = \sigma_{\tau,n}/(\sigma_{1,n}\tau^{1/2})$ approaches zero as we increase the maturity τ of the option. Thus mean reversion in the pricing kernel and interest rates implies, as a general feature of our theoretical framework, a term structure of volatilities for the Black-Scholes formula that declines for long options. Similar properties are implicit in Jamshidian (1989) and Turnbull and Milne (1991). What is new is the connection between this condition on implied volatilities and the slope of the mean forward rate curve.

Our ARMA(1,1) example gives us some idea of the magnitudes involved. From equation (18) we know that

$$A_{n+j} - A_j = \varphi^j (1 - \varphi^n)(\varphi + \theta)/(1 - \varphi),$$

so the volatility parameter [equation (21)] for the call option is

$$\sigma_{\tau,n} = \sigma[(1-\varphi^n)|\varphi+\theta|/(1-\varphi)][(1-\varphi^{2\tau})/(1-\varphi^2)]^{1/2}.$$

If we choose $v = \sigma_{1,n}$ to match the volatility formula in Black-Scholes to the true volatility for $\tau = 1$, then for more distant exercise dates we must choose implied volatilities

$$v_\tau = [(1-\varphi^{2\tau})/(1-\varphi^2)]^{1/2}/\tau^{1/2}.$$

The sequence $\{v_\tau\}$ is an example of a term structure of volatilities. Using the same parameter as Section 4, $\varphi = 0.976$, this results in a declining term structure of volatility (Figure 6).

A related feature of our example is that the volatility parameter, defined by equation (21), generally declines with the maturity n of the underlying bond. Thus our model, like Jamshidian's (1989), automatically produces the positive relation between "duration" (n in this case) and volatility addressed by Schaefer and Schwartz (1987).

Forward and futures prices. Cox, Ingersoll, and Ross (1981) have traced differences in prices of forward and futures contracts to correlation between bond and futures prices. The same is true in our theoretical framework, but given our estimated parameter values we can trace this feature back to properties of the pricing kernel.

Forward contracts are implicit in bond prices and forward rates. Consider a contract at date t specifying payment at date $t+\tau$ of $F_t^{\tau,n}$ dollars in return for an n-period bond at the same date, the claim to one dollar at date $t+\tau+n$. We can replicate this cashflow with two bond transactions at date t: we buy a bond with maturity $\tau+n$ and we sell (short) the equivalent value of τ-period bonds. If bond prices are labeled b_t^n, then the second transaction involves the quantity $b_t^{\tau+n}/b_t^\tau$ of τ-period bonds and a cash flow at date $t+n$ of the same amount. Arbitrage requires

$$F_t^{\tau,n} = b_t^{\tau+n}/b_t^\tau.$$

From equation (14) we see that the equilibrium forward price is

$$-\log F_t^{\tau,n} = n\delta - (\sigma^2/2)\sum_{j=\tau+1}^{\tau+n} A_{j-1}^2 + \sum_{j=0}^{\infty}(A_{\tau+n+j} - A_{\tau+j})\epsilon_{t-j}, \quad (22)$$

with the usual definitions of parameter values.

Now consider a futures contract specifying receipt of the same n-period bond at date $t+\tau$. Unlike the forward contract, the futures requires the owner to post collateral each period equal to the current value of the contract. Cox, Ingersoll, and Ross (1981, Proposition 2) show that the value of the futures contract is the value of the payment

$$b^n_{t+\tau} / \prod_{j=0}^{\tau-1} b^1_{t+j}$$

at date $t+\tau$, the date the contract matures. The date-t value of the contract is

$$G^{\tau,n}_t = E_t \left(\prod_{j=0}^{\tau-1} (m_{t+j+1}/b^1_{t+j}) b^n_{t+\tau} \right).$$

This has the recursive representation,

$$G^{\tau+1,n}_t = E_t(m^*_{t+1} G^{\tau,n}_{t+1}),$$

starting with $G^{0,n}_t = b^n_t$, where $m^*_{t+1} \equiv m_{t+1}/b^1_t$ defines the equivalent martingale measure (Duffie 1993, Section 2G). The price of a futures contract with maturity $\tau > 0$ is thus

$$\begin{aligned} -\log G^{\tau,n}_t &= n\delta - \sigma^2/2 \left(\sum_{j=1}^{n} A^2_{j-1} + \sum_{j=0}^{\tau-1} (A_{n+j} - A_j + A_0)^2 - \tau A^2_0 \right) \\ &\quad + \sum_{j=0}^{\infty} (A_{\tau+n+j} - A_{\tau+j}) \epsilon_{t-j}. \end{aligned} \quad (23)$$

Note that the coefficients of ϵ_{t-j} are the same as those for the forward contract.

We are now in a position to compare prices of forward and futures contracts, given by equations (22) and (23). In our theoretical framework, the ratio of the two prices is constant, with

$$\log F^{\tau,n}_t - \log G^{\tau,n}_t = \sigma^2 \sum_{j=0}^{\tau-1} (A_j - A_{n+j})(A_0 - A_j).$$

This price differential can have either sign, but with our estimated kernel it is positive, as we seem to see in the data (Meulbroek, 1992, Table IV).

More interesting to us is that the conditions that deliver a positive price differential are different from those that give us an increasing mean forward rate curve. Instead, the critical ingredient seems to be the persistence of the short rate: pricing kernels that produce highly autocorrelated short rates also tend to produce positive price differentials. We can see this most easily by looking at two examples. One example is a kernel that generates a random walk short rate, which is obviously highly persistent. In this case the partial sums are $A_n = 1 + n\alpha_1$, so the price differential is

$$\log F_t^{\tau,n} - \log G_t^{\tau,n} = \sigma^2 n\tau(\tau - 1)\alpha_1^2/2,$$

which is positive for maturities $\tau > 1$. A second example is our ARMA(1,1), with A_n given by equation (18). The first-order autocorrelation of the short rate is φ. The price differential in this example is

$$\log F_t^{\tau,n} - \log G_t^{\tau,n} = \sigma^2[(\varphi + \theta)/(1 - \varphi)]^2(1 - \varphi^n)\sum_{j=0}^{\tau-1}\varphi^j(1 - \varphi^j),$$

which is positive if $\varphi > 0$, negative otherwise. Note that this condition can hold even in models that do not produce increasing mean yield curves. With $\varphi > 0$, for example, the slope of the yield curve is determined by the sign of $\varphi + \theta$, but this second condition has no bearing on the sign of the price differential.

Positive autocorrelation of the short rate is not sufficient to guarantee a positive price differential, but it's close. A sufficient condition is that the moving average coefficients of the short rate, α_j for $j \geq 1$, all have the same sign, which guarantees positive autocorrelations at all lags. This condition says that an innovation to the short rate increases the conditional mean of the short rate at all future dates. It holds for the random walk model, in which these coefficients are the same, and for the ARMA(1,1) example with positive φ, in which $\alpha_{j+1} = \varphi\alpha_j$. Although this result is specific to our log-linear framework, it is suggestive of two related results. The first is the positive price differentials produced by the Cox-Ingersoll-Ross (1981, Section 4) model. As we show in Appendix D, their model is a generalization of our ARMA(1,1) example in which the analog of φ is positive. Perhaps the same mechanism is at work in their model and our ARMA(1,1) example. Our second conjecture is that the dependence of the price differential on persistence that characterizes our model may extend to other environments. Since other popular models of bond pricing imply highly persistent interest rates, perhaps they also imply a positive premium of forward prices over futures prices.

7 Final Remarks

We have described, in the context of a log-linear model of bond pricing, how information on bond yields can be used to estimate the pricing kernel that gave rise to them. The pricing kernel is the building block of fixed income security pricing, so information about its properties has direct application in the pricing of interest-rate derivative securities. For this reason, we think that our procedure of reverse engineering may help us understand the features of a theory that lead to observed properties of asset prices, and lead to improvements in pricing derivative assets. Our log-linear framework, however, is only one step in this direction. It allows us to relate properties of asset prices to those of the kernel in a relatively simple and transparent way, but it rules out, among other things, the changing volatility apparent in bond prices and returns. One extension, outlined in Appendix E, introduces a stochastic process for the conditional variance of the pricing kernel, which was constant in the framework of this paper. Another is to model the kernel along the lines of the semi-nonparametric models used by Gallant and Tauchen (1989). This approach, in principle, would allow us to approximate an arbitrary nonlinear pricing kernel. Perhaps future work will tell us which extensions are the most useful for understanding the prices of bonds and related assets.

A Alternate Proof of Proposition 1

The proof starts with Jamshidian's (1988) suggestion to translate the pricing relation for bond prices into an analogous relation for forward rates. From definition (2) we note that the one-period return on an $(n+1)$-period bond satisfies

$$\begin{aligned} \log R^{n+1}_{t+1} &= \log b^n_{t+1} - \log b^{n+1}_t \\ &= f^n_t + \log b^n_{t+1} - \log b^n_t. \end{aligned}$$

The pricing relation (4) then implies

$$\exp(-f^n_t) = E_t(m_{t+1} b^n_{t+1}/b^n_t), \tag{24}$$

and we compute forward rates and bond prices recursively. For $n=0$ we have $b^0_t = 1$, so

$$\exp(-f^0_t) = E_t m_{t+1}.$$

We now apply a property of expectations of log-normal random variables: if $\log x$ is normal with mean μ and variance σ^2, then $\log E(x) = \mu + \sigma^2/2$. That gives us

$$f^0_t = \delta - \alpha_0^2\sigma^2/2 + \sum_{j=0}^{\infty} \alpha_{j+1}\epsilon_{t-j}.$$

This in turn defines the one-period bond price as

$$-\log b^1_t = f^0_t = \delta - \alpha_0^2\sigma^2/2 + \sum_{j=0}^{\infty} \alpha_{j+1}\epsilon_{t-j},$$

which is clearly linear in the innovations $\{\epsilon_t\}$.

Bond prices and forward rates for higher maturities follow by induction. Suppose the n-period bond price can be expressed

$$-\log b^n_t = \mu^n + \sum_{j=0}^{\infty} \beta^n_j \epsilon_{t-j},$$

for some parameters $\{\mu^n, \beta^n_j\}$. We've just derived the parameters for $n=1$. Given the parameters for an arbitrary maturity $n > 0$, we derive the price of an $(n+1)$-period bond from the forward rate pricing relation, equation (24). Note that

$$-\log(m_{t+1}b^n_{t+1}/b^n_t) = \delta + (\alpha_0 + \beta^n_0)\epsilon_{t+1} + \sum_{j=0}^{\infty}(\alpha_{j+1} + \beta^n_{j+1} - \beta^n_j)\epsilon_{t-j}.$$

This random variable has conditional mean

$$-E_t[\log(m_{t+1}b^n_{t+1}/b^n_t)] = \delta + \sum_{j=0}^{\infty}(\alpha_{j+1} + \beta^n_{j+1} - \beta^n_j)\epsilon_{t-j}.$$

and conditional variance

$$Var_t[\log(m_{t+1}b^n_{t+1}/b^n_t)] = (\alpha_0 + \beta^n_0)^2\sigma^2.$$

Hence

$$f^n_t = \delta - (\alpha_0 + \beta^n_0)^2\sigma^2/2 + \sum_{j=0}^{\infty}(\alpha_{j+1} + \beta^n_{j+1} - \beta^n_j)\epsilon_{t-j}.$$

Proposition 1 follows from running through the recursions.

B Multifactor Models

Several papers have looked at similar bond pricing models with more than one factor, including Heath, Jarrow, and Morton (1992, Section 6) and Turnbull and Milne (1991, Section 5). We do the same here for our theoretical framework.

We develop the theory in a two-factor setting; extensions to higher dimensions should be transparent. Let us say that the pricing kernel follows a two-dimensional process,

$$-\log m_t = \delta + \sum_{j=0}^{\infty}\alpha_{1j}\epsilon_{1,t-j} + \sum_{j=0}^{\infty}\alpha_{2j}\epsilon_{2,t-j},$$

where the ϵ's are independent normal random variables with zero means and variances σ_1^2 and σ_2^2. By convention we set $\alpha_{10} = \alpha_{20} = 1$. Using the same methods as Proposition 1, we find that forward rates are

$$f^n_t = \delta - (A^2_{1n}\sigma^2_1 + A^2_{2n}\sigma^2_2)/2 + \sum_{j=0}^{\infty}\alpha_{1,n+1+j}\epsilon_{1,t-j} + \sum_{j=0}^{\infty}\alpha_{2,n+1+j}\epsilon_{2,t-j},$$

where $A_{in} \equiv \sum_{j=0}^{n} \alpha_{ij}$. This is similar to equation (12), but has two components for both the mean and moving average terms.

There are two differences between the one and two factor models worth noting. One difference is that the dynamics can be more complex with two factors than with one. The combination of two AR(1) factors, for example, produces ARMA(2,1) dynamics in the pricing kernel. This aspect of the two-factor theory is easily replicated by choosing appropriate moving average coefficients in the one-factor model. The second difference is more fundamental: the two factor model gives agents a finer information partition. They know the individual innovations, ϵ_1 and ϵ_2, not just their sum (which is the innovation to the pricing kernel). For this reason, a two-factor model is not observationally equivalent to a more complex one-factor model. This is easily verified by comparing mean forward rates in a two-factor model with those from the one-factor Wold representation of its pricing kernel.

Example. Let the pricing kernel be

$$-\log m_t = \delta + x_{1t} + x_{2t},$$

with

$$x_{it} = \rho_i x_{i,t-1} + \epsilon_{it}.$$

This gives us $\alpha_{ij} = \rho_i^j$ for $i = 1,2$ and $j \geq 0$. Forward rates are

$$f_t^n = \delta - (A_{1n}^2\sigma_1^2 + A_{2n}^2\sigma_2^2)/2 + \rho_1^{n+1}x_{1t} + \rho_2^{n+1}x_{2t},$$

with $A_{in} = (1-\rho_i^{n+1})/(1-\rho_i)$.

For comparison, consider the infinite moving average (Wold) representation of the kernel, obtained by projecting $\log m_t$ onto its past innovations. In the example (and more generally with the normalization $\alpha_{i0} = 1$), innovations are $\epsilon_t = \epsilon_{1t} + \epsilon_{2t}$. Their variance is $\sigma^2 = \sigma_1^2 + \sigma_2^2$. The Wold representation is equation (6) with

$$\alpha_j = Cov(\alpha_{1j}\epsilon_{1,t-j} + \alpha_{2j}\epsilon_{2,t-j}, \epsilon_{t-j})/Var(\epsilon_{t-j}) = (\rho_1^j\sigma_1^2 + \rho_2^j\sigma_2^2)/\sigma^2.$$

Unless $\rho_1 = \rho_2$, forecasts with this one factor representation are inferior to those of the two factor model, since they are based on a coarser information partition. The one-factor representation also generates different forward rates, as you can see by comparing $A_n^2\sigma^2$ with $A_{1n}^2\sigma_1^2 + A_{2n}^2\sigma_2^2$ for $n \geq 1$.

This two-factor example leads to a two-dimensional vector autoregression in (say) f^0 and f^1, and thus bears some resemblance to the interest rate dynamics in Brennan and Schwartz (1979).

C ARMA Pricing Kernels

We show how the dynamics of the pricing kernel translate into dynamics of the short rate. The only subtle feature of this translation is the impact of autoregressive coefficients of the pricing kernel on moving average coefficients of interest rates. To see how this works, consider an ARMA(2,0) pricing kernel:

$$-\log m_t = -(1-\varphi_1-\varphi_2)\delta - \varphi_1 \log m_{t-1} - \varphi_2 \log m_{t-2} + \epsilon_t.$$

The moving average coefficients of the Wold representation of the pricing kernel, equation (6), are $\alpha_0 = 1$, $\alpha_1 = \varphi_1$, and $\alpha_j = \varphi_1\alpha_{j-1} + \varphi_2\alpha_{j-2}$ for $j > 1$. The short rate is

$$\begin{aligned}
r_t &= (\delta - \sigma^2/2) + \sum_{j=0}^{\infty} \alpha_{j+1}\epsilon_{t-j} \\
&= (\delta - \sigma^2/2) + \alpha_1\epsilon_t + \sum_{j=1}^{\infty} (\varphi_1\alpha_j + \varphi_2\alpha_{j-1})\epsilon_{t-j} \\
&= (\delta - \sigma^2/2) + \alpha_1\epsilon_t + \varphi_1 \sum_{j=0}^{\infty} \alpha_{j+1}\epsilon_{t-j-1} + \varphi_2 \sum_{j=0}^{\infty} \alpha_j\epsilon_{t-j-1} \\
&= (1-\varphi_1-\varphi_2)(\delta - \sigma^2/2) + \varphi_1 r_{t-1} + \varphi_2 r_{t-2} + \varphi_1\epsilon_t + \varphi_2\epsilon_{t-1}.
\end{aligned}$$

Thus, the autoregressive parameters φ_j of the pricing kernel show up in the short rate as both autoregressive and moving average parameters, and an ARMA(2,0) pricing kernel generates an ARMA(2,1) short rate.

More generally, consider the ARMA(p,q) pricing kernel of Section 5. The kernel can be expressed in moving average form as

$$-\log m_t = [\Phi(1)/\Phi(L)]\delta + [\Theta(L)/\Phi(L)]\epsilon_t.$$

The short rate is

$$r_t = [\Phi(1)/\Phi(L)](\delta - \sigma^2/2) + [\Theta(L)/\Phi(L)]\epsilon_{t+1} - \epsilon_{t+1},$$

or

$$\Phi(L)r_t = \Phi(1)(\delta - \sigma^2/2) + [\Theta(L) - \Phi(L)]\epsilon_{t+1}.$$

Since $\varphi_0 = \theta_0 = 1$, ϵ_{t+1} drops out of the equation, leaving us with the remaining $p-1$ autoregressive and $q-1$ moving average parameters in the moving average component of the short rate. Therefore, an ARMA(p,q) pricing kernel generates an ARMA(p,max(p,q)-1) short rate.

The same logic can be applied to show that all forward rates have the same autoregressive polynomial, $\Phi(L)$, but the calculation of the MA parameters is more difficult for higher maturity forward rates. Yields also have the autoregressive polynomial $\Phi(L)$, since they are averages of forward rates, but again the MA parameters are more complicated.

D Cox-Ingersoll-Ross as an ARMA(1,1)

We express Sun's (1992) discrete time version of the Cox-Ingersoll-Ross (1985) model of bond pricing in a form similar to our ARMA(1,1) example. The starting point is a state variable z that obeys the "square root" process

$$z_t = (1-\varphi)\delta + \varphi z_{t-1} + \lambda z_{t-1}^{1/2}\epsilon_t,$$

with $\{\epsilon_t\} \sim \text{NID}(0, \sigma^2)$. This relation is AR(1), despite the unusual form of the innovation. With the square-root process the conditional variance of the innovation is proportional to z, which reduces the chance of getting a negative value. As the time interval approaches zero, so does the probability that z turns negative. The pricing kernel is

$$-\log m_t = z_{t-1} + z_{t-1}^{1/2}\epsilon_t,$$

which can be rewritten as

$$-\log m_t \;=\; (1-\varphi)\delta - \varphi \log m_{t-1} + z_{t-1}^{1/2}\epsilon_t + (\lambda - \varphi) z_{t-2}^{1/2}\epsilon_{t-1}.$$

With the exception of the square-root terms, this is just our ARMA(1,1) example with $\theta = \lambda - \varphi$. Since the kernel is conditionally log-normal, we can approach bond pricing in much the same way we did in Section 3. Forward rates and yields are linear functions of the state variable z. The short rate, for example, is $r_t = (1 - \sigma^2/2)z_t$, which allows us to express yields as functions of r rather than z.

E Stochastic Volatility

One obvious limitation of our pricing kernel is its constant conditional variance, so stochastic volatility is high on our list of extensions. The cost is considerably greater complexity in expressions for bond prices and forward rates. We illustrate how this might work in an environment that is simple enough to retain linearity of forward rate functions.

Let us say that the pricing kernel is equation (6) with volatility $Var_t(\epsilon_{t+1}) = h_t$ that varies through time. With $h_t = \sigma^2$ we are back in the constant volatility world of the paper, and with $h_t = z_t$ we get the Cox-Ingersoll-Ross model. We focus on a tractable alternative, in which h is an independent linear process:

$$h_{t+1} = \sigma^2 + 2\sum_{j=0}^{\infty} \gamma_j \eta_{t-j},$$

with $\{\eta_t\} \sim \text{NID}(0, \sigma_\eta^2)$ and $Cov(\eta_t, \epsilon_t) = 0$. If $\gamma_j = 0$ this reduces to the model of the text. With this kernel, forward rates are

$$f_t^n = \delta - A_n^2 h_t/2 + \sum_{j=0}^{\infty} \alpha_{n+1+j}\epsilon_{t-j} - \sigma_\eta^2 \left(\sum_{j=1}^{n} A_{j-1}^2 \gamma_{n-j} \right)^2,$$

for $n \geq 1$. This is similar to equation (12), with h_t replacing the constant σ^2, but the last term illustrates how the dynamics of the kernel are intertwined with those of volatility. The dynamics of the short rate reflect both the moving average coefficients of the kernel, the α_j's, and those of volatility, the γ_j's.

Despite the increase in complexity, some of the salient features of the constant volaility model extend to this environment. One example is the relation between mean reversion and the slope of the forward rate curve. Mean forward rates are

$$E(f^n) \quad = \quad \delta - A_n^2\sigma^2/2 - \sigma_\eta^2 \left(\sum_{j=1}^{n} A_{j-1}^2 \gamma_{n-j} \right)^2.$$

For this to be increasing we again need the sequence $\{A_n^2\}$ to be decreasing. If A_n^2 converges to a nonzero limit, a similar requirement applies to the dynamics of volatility – namely, the coefficients $\{\gamma_j\}$ must be summable. Together they imply eventual decreasing conditional volatility of bond prices, which we used in our discussion of bond options.

References

Backus, David, and Stanley Zin (1993), "Long-memory inflation uncertainty: Evidence from the term structure of interest rates," *Journal of Money, Credit, and Banking* 25, 681-700.

Bekaert, Geert, and Robert Hodrick (1992), "Characterizing predictable components in excess returns on equity and foreign exchange markets," *Journal of Finance* 47, 467-509.

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*, 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 (1986), "Bond and stock returns in a simple exchange economy," *Quarterly Journal of Economics* 101, 785-803.

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

Cochrane, John, and Lars Hansen (1992), "Asset pricing explorations for macroeconomics," in O. Blanchard and S. Fischer, (eds.), *NBER Macroeconomics Annual*. Cambridge, MA: MIT Press, 115-165.

Constantinides, George (1992), "A theory of the nominal term structure of interest rates," *Review of Financial Studies* 5, 531-552.

Cox, John, Jonathan Ingersoll, and Stephen Ross (1981), "The relation between forward prices and futures prices," *Journal of Financial Economics* 9, 321-346.

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

den Haan, Wouter (1993), "The term structure of interest rates in real and monetary economies," unpublished manuscript, University of California, San Diego.

Duffie, Darrell (1993), *Dynamic Asset Pricing Theory*. Princeton, NJ: Princeton University Press.

Dybvig, Philip (1989), "Bond and bond option pricing based on the current term structure," unpublished manuscript, Washington University, February.

Gallant, A. Ronald, and George Tauchen (1989), "Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications," *Econometrica* 57, 1091-1120.

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.

Hansen, Lars (1982), "Large sample properties of generalized method of moments estimators," *Econometrica* 50, 1029-1054.

Hansen, Lars, and Ravi Jagannathan (1991), "Implications of security market data for models of dynamic economies," *Journal of Political Economy* 99, 225-262.

Heath, David, Robert Jarrow, and Andrew Morton (1992), "Bond pricing and the term structure of interest rates," *Econometrica* 60, 77-105.

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 (1993), *Options, Futures, and Other Derivative Securities, Second Edition*. Englewood Cliffs, NJ: Prentice-Hall.

Jamshidian, Farshid (1988), "Notes on Heath, Jarrow, and Morton," unpublished manuscript, Merrill Lynch, April.

Jamshidian, Farshid (1989), "An exact bond option formula," *Journal of Finance* 44, 205-209.

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," unpublished manscript and computer diskettes, Ohio State University.

Merton, Robert (1973), "An intertemporal capital asset pricing model," *Econometrica* 41, 867-888.

Meulbroek, Lisa (1992), "A comparison of forward and futures prices of an interest sensitive financial asset," *Journal of Finance* 47, 381-396.

Sargent, Thomas (1987), *Dynamic Macroeconomic Theory*. Cambridge: Harvard University Press.

Schaefer, Stephen, and Eduardo Schwartz (1987), "Time-dependent variance and the pricing of bond options," *Journal of Finance* 42, 1113-1128.

Stambaugh, Robert (1988), "The information in forward rates: Implications for models of the term structure," *Journal of Financial Economics* 21, 41-70.

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.

Telmer, Chris (1993), "Asset pricing and incomplete markets," *Journal of Finance* 49, 1803-1832.

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.

</ref_section>

Table 1
Yields and Forward Rates: Summary Statistics

The data are monthly estimates of annualized continuously-compounded zero-coupon US government bond yields and instantaneous forward rates computed by McCulloch and Kwon (1993). Mean is the sample mean, St Dev the standard deviation, and Auto the first autocorrelation.

Maturity	1952:1 - 1991:2			1982:1 - 1991:2		
	Mean	St Dev	Auto	Mean	St Dev	Auto
A. Yields						
1 month	5.314	3.064	0.976	7.483	1.828	0.906
3 months	5.640	3.143	0.981	7.915	1.797	0.920
6 months	5.884	3.178	0.982	8.190	1.894	0.926
9 months	6.003	3.182	0.982	8.372	1.918	0.928
12 months	6.079	3.168	0.983	8.563	1.958	0.932
24 months	6.272	3.124	0.986	9.012	1.986	0.940
36 months	6.386	3.087	0.988	9.253	1.990	0.943
48 months	6.467	3.069	0.989	9.405	1.983	0.946
60 months	6.531	3.056	0.990	9.524	1.979	0.948
84 months	6.624	3.043	0.991	9.716	1.956	0.952
120 months	6.683	3.013	0.992	9.802	1.864	0.950
B. Forward Rates						
1 month	5.552	3.140	0.979	7.781	1.753	0.915
3 months	5.963	3.200	0.981	8.334	1.961	0.921
6 months	6.225	3.256	0.976	8.579	1.990	0.923
9 months	6.263	3.169	0.981	8.925	2.050	0.933
12 months	6.358	3.169	0.984	9.320	2.149	0.942
24 months	6.516	3.037	0.986	9.472	2.093	0.943
36 months	6.696	3.071	0.989	9.923	1.966	0.943
48 months	6.729	3.026	0.990	9.833	2.050	0.949
60 months	6.839	3.062	0.991	10.182	1.972	0.953
84 months	6.838	2.997	0.992	10.068	1.900	0.952
120 months	6.822	2.984	0.991	10.058	1.522	0.908

Table 2
Estimates of ARMA Pricing Kernels

The table lists GMM estimates of ARMA pricing kernels. Date are monthly from 1952:1 to 1990:2 with the first 24 observations reserved for calculating autocovariances of the short rate. Numbers in parentheses are standard errors. The number in brackets below the J-statistic is its significance probability or p-value. The mean is denoted δ, the innovation standard deviation σ, the autoregressive parameters φ_j, and the moving average parameters θ_j. The mean δ is fixed to match the sample mean of the short rate. The moment conditions used to estimate the remaining parameters are yield spreads for maturities 3, 12, 36, 60, and 120 months and short-rate autocovariances of orders 0, 1, 3, 12, and 24 months, for a total of 10 conditions. The weighting matrix is computed from first-stage ARMA(2,3) estimates by the method of Newey and West, using 48 autocovariances.

Parameter/Statistic	ARMA(2,3)	ARMA(2,2)	ARMA(1,1)
δ	0.528022	0.206633	0.030679
σ	1.023141	0.635673	0.228415
	(0.000733)	(0.001033)	(0.002688)
φ_1	1.031253	1.234310	0.707288
	(0.176372)	(0.117500)	(0.014194)
φ_2	-0.073191	-0.278337	
	(0.166909)	(0.107638)	
θ_1	-1.031448	-1.235127	-0.712387
	(0.176429)	(0.117473)	(0.013901)
θ_2	0.073011	0.279004	
	(0.167110)	(0.107642)	
θ_3	0.000322		
	(0.000153)		
J-Statistic	0.3683	9.0746	50.9837
	[0.9850]	[0.1061]	[0.0000]

Figure 1
Mean Yield Curves for US Govt Bonds

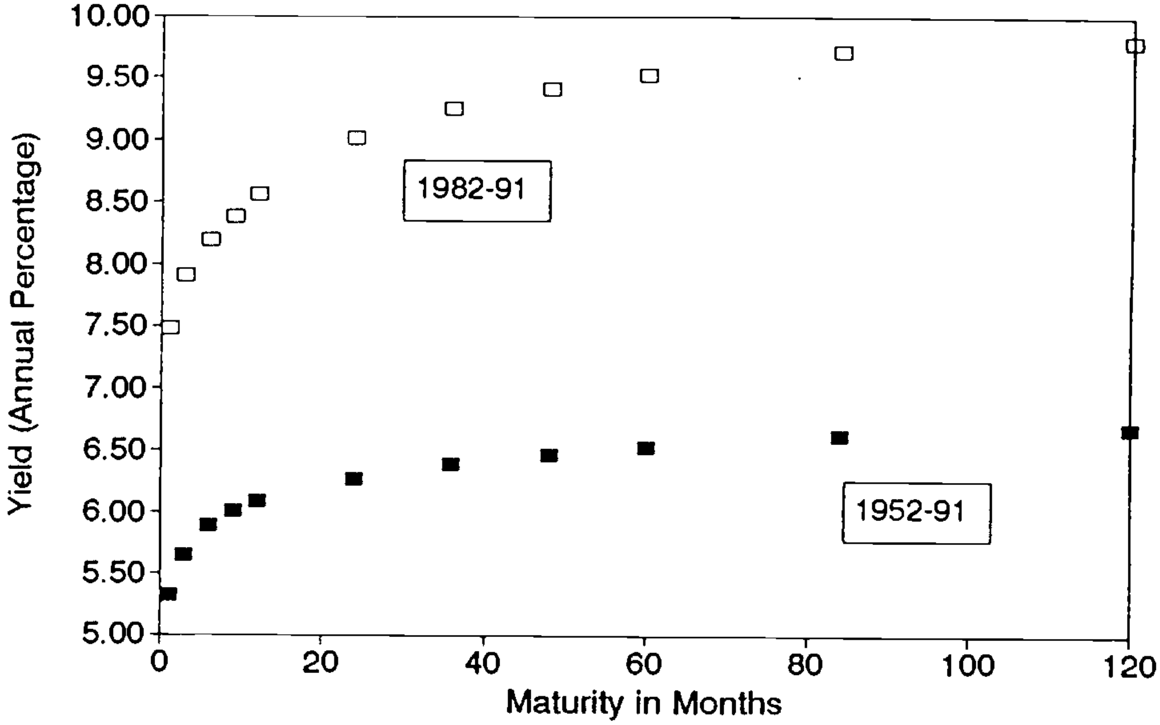
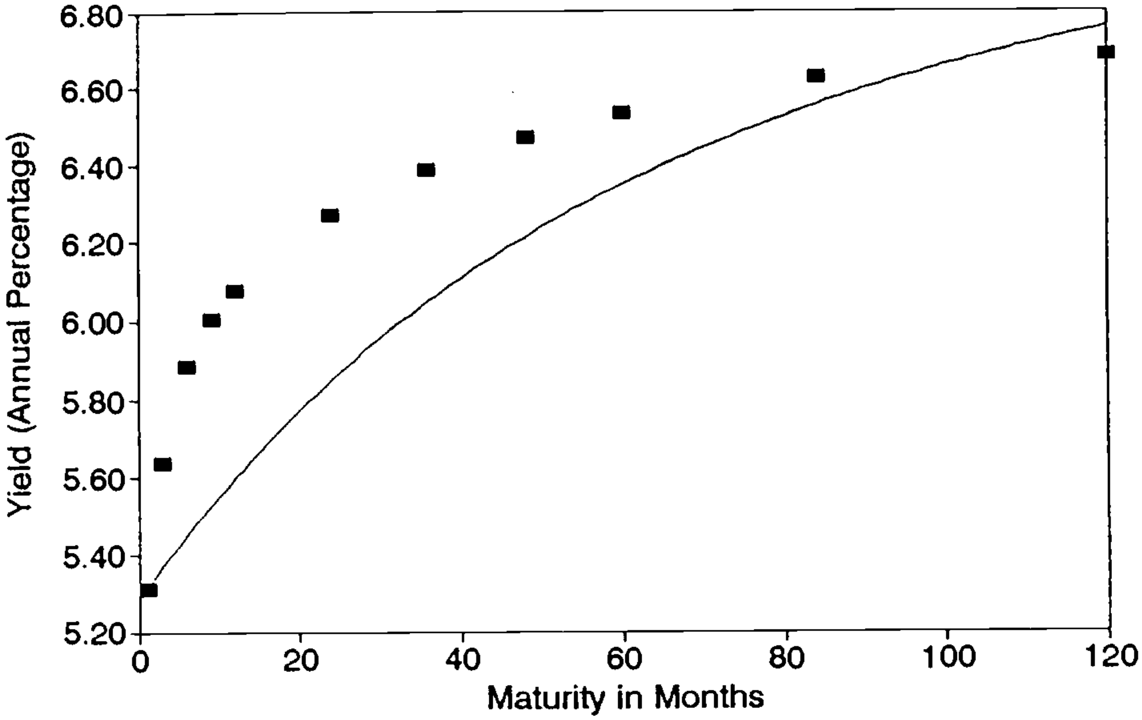
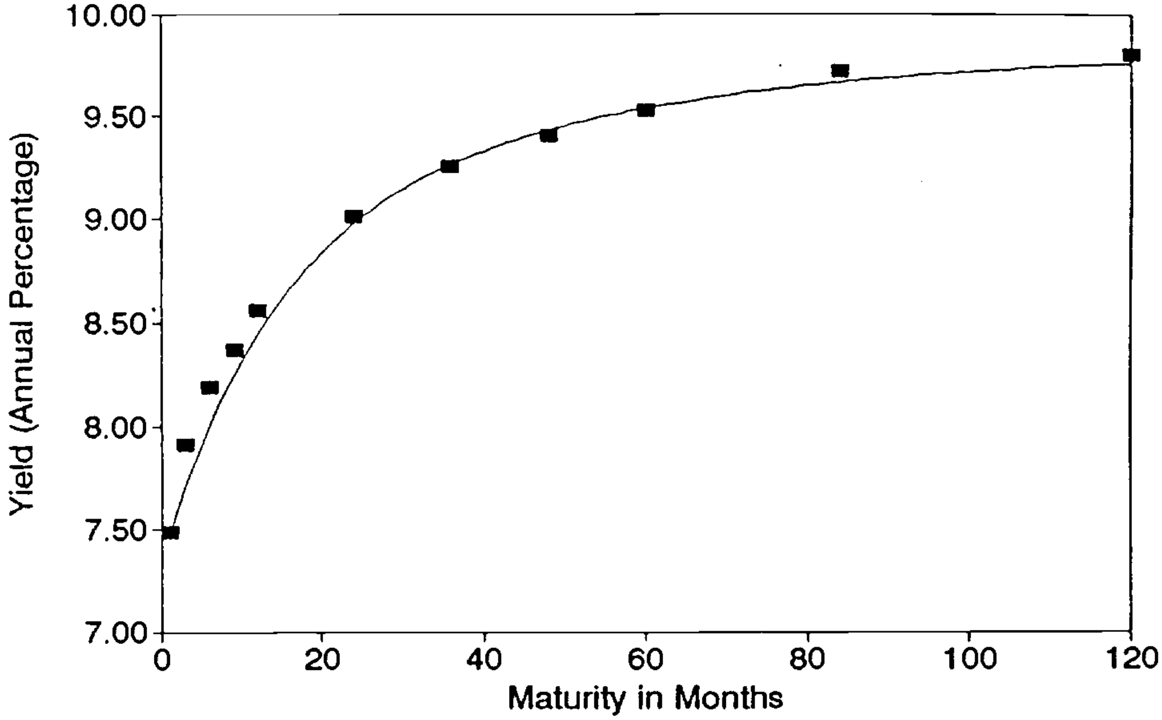


Figure 2A
Mean Yield Curve for ARMA(1,1) Example



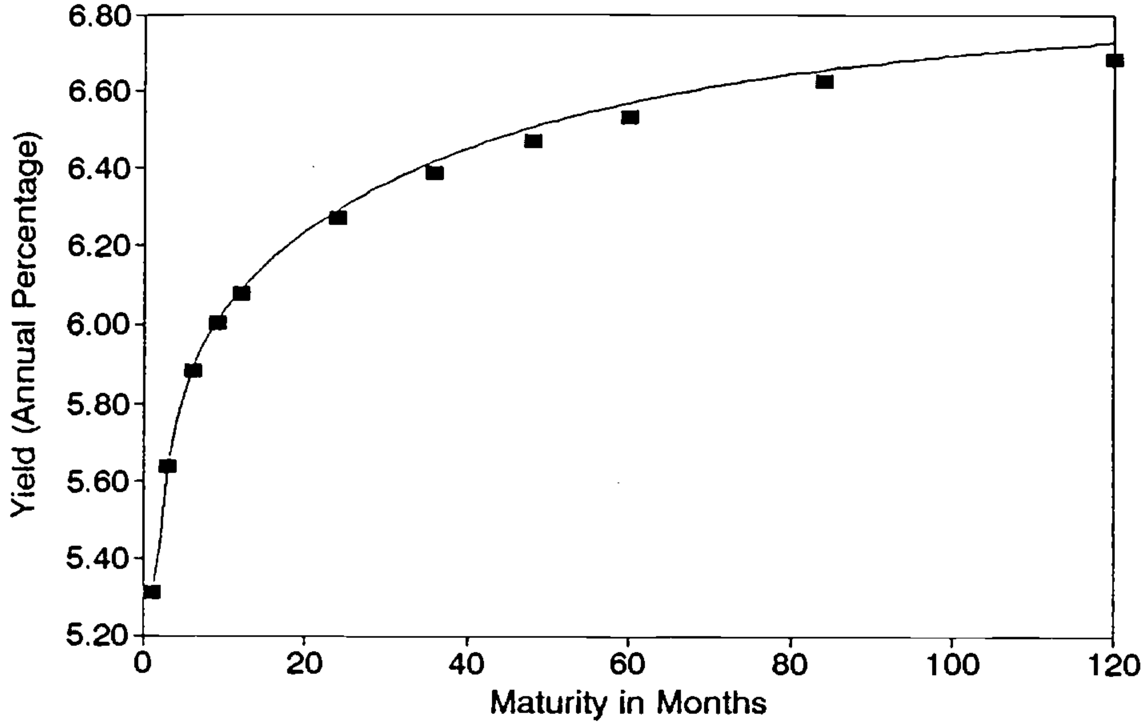
Black squares are data points, the line is theory.

Figure 2B
ARMA(1,1) Example: 1982-91



Black squares are data points, the line is theory.

Figure 3
Mean Yield Curve for ARMA(2,3)



Black squares are data points, the line is theory.

Figure 4
Dynamics of Estimated Pricing Kernel

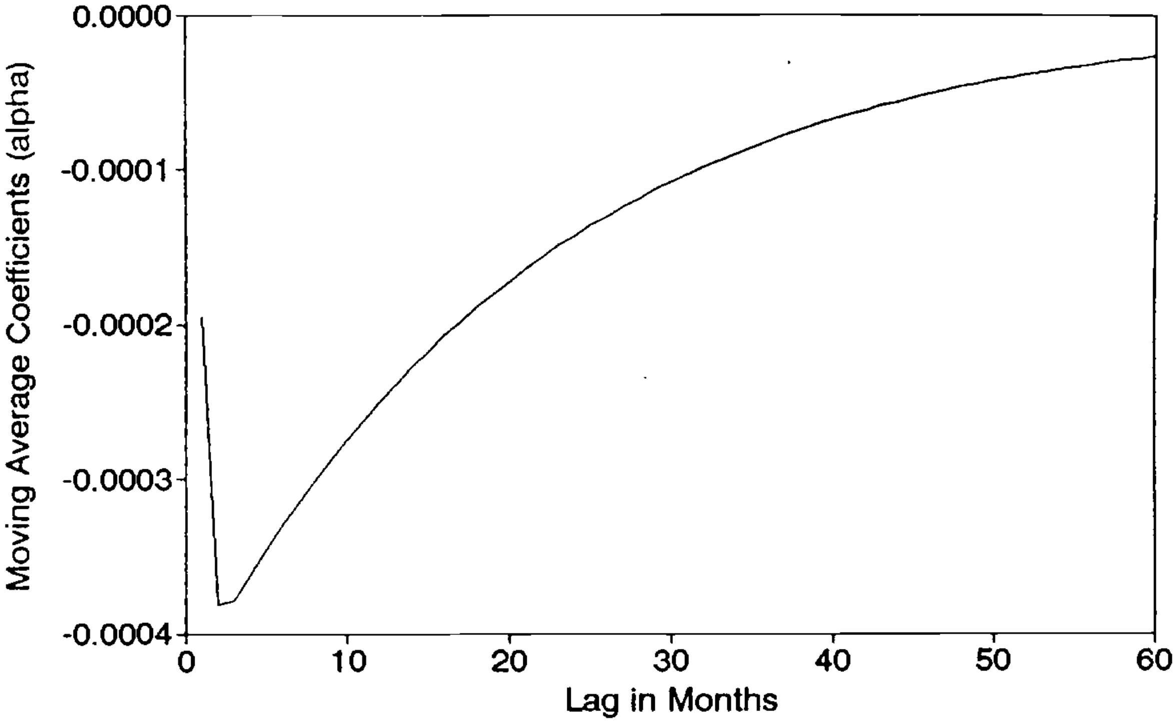


Figure 5
Price of Risk in Estimated Kernel

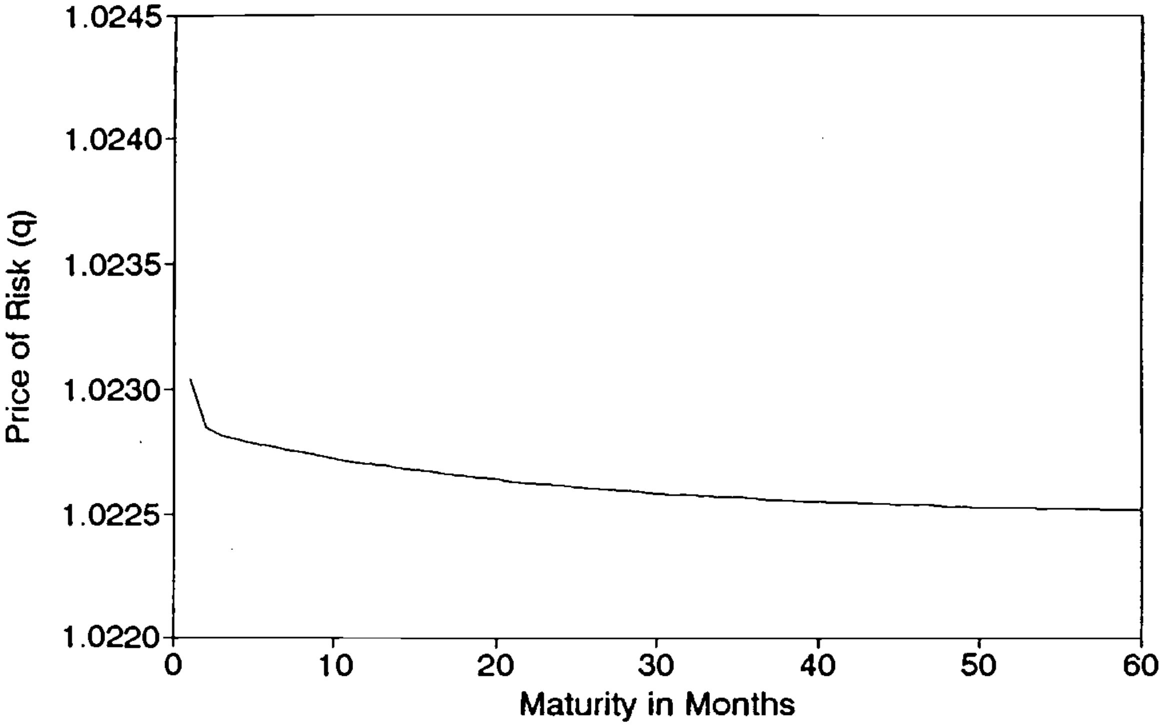


Figure 6
Term Structure of Volatility