Forces That Shape the Yield Curve

MARK FISHER

Fisher is a senior economist and policy adviser in the financial section of the Atlanta Fed's research department. *He thanks Lucy Ackert, Christian Gilles, Frank King, Stephen LeRoy, Saikat Nandi, Steve Smith, and Larry Wall for helpful comments on an earlier draft.*[1]

MONETARY POLICYMAKERS AND OBSERVERS PAY SPECIAL ATTENTION TO THE SHAPE OF THE YIELD CURVE AS AN INDICATOR OF THE IMPACT OF CURRENT AND FUTURE MONETARY POLICY ON THE ECONOMY. THE YIELD CURVE SHOWS HOW THE YIELD ON A GOVERNMENT BOND DEPENDS ON THE MATURITY OF THE BOND. HOWEVER, DRAWING INFERENCES FROM the yield curve is much like reading tea leaves if one does not have the proper tools for yield-curve analysis. The purpose of this article is to provide a rigorous yet accessible introduction to those tools using high-school algebra.

The yield curve is shaped by expectations of the future path of short-term interest rates and by uncertainty regarding the path. Uncertainty affects the yield curve through two channels: investor attitudes toward risk (risk aversion) as reflected in risk premia and the nonlinear relation between yields and bond prices (known as convexity). In order to present the theory behind the yield curve correctly, uncertainty must be taken seriously. Nevertheless, the source of uncertainty can be modeled quite simply: all uncertainty is resolved by a single flip of a coin. In this setting, all three forces (expectations, risk aversion, and convexity) that shape the yield curve can be rigorously presented. The analysis is organized around the conditions that guarantee the absence of arbitrage opportunities. An arbitrage is a trading strategy that produces something for nothing.

The basic ideas are developed first in an introductory section by the use of an analogy. Next, bond pricing is introduced in a world of perfect certainty, in which no-arbitrage conditions are first worked out algebraically. (In this setting, the absence-of-arbitrage conditions are equivalent to the expectations hypothesis of the term structure of interest rates.[2]) Next, uncertainty is introduced via the coin flip, and the no-arbitrage conditions for bond prices are worked out again. These no-arbitrage conditions are shown to imply the existence of a risk premium that depends on the price of risk (which reflects risk aversion and is the same for all bonds) and the amount of risk (which is measured by the volatility of a bond's price).[3] The last section discusses how to translate the no-arbitrage condition for bond prices into a no-arbitrage condition for yields. The nonlinearity of the price-yield relation brings the convexity term into play.

What Is the Yield Curve?

The simplest kind of bond is called a zero-coupon bond. A zero-coupon bond (also known as a discount bond) makes a single payment on its maturity date. By contrast, a coupon bond makes periodic interest payments, called coupon payments, prior to its maturity when it also makes a final payment that represents repayment of principal. A coupon bond may be thought of as a portfolio of zero-coupon bonds.

A default-free bond is a bond for which all payments are certain to be made in full and on time. U.S. Treasury securities are generally considered to be default-free. The Treasury issues both coupon bonds and zero-coupon bonds. Treasury bills are zero-coupon bonds with original maturities of one year of less. Treasury notes and bonds are coupon bonds with original maturities of two years or more (bonds have original maturities of twenty years or more) that pay interest twice a year. Since the mid-1980s, investors have been able to trade the coupon payments of certain Treasury notes and bonds separately as zero-coupon bonds in what is known as the STRIPS market.[4]

Bonds with different maturities typically have different yields. For example, the yield on a five-year bond is often higher than the yield on a two-year bond. But sometimes the yield on the two-year bond is higher. At any given point in time, the yield curve can be plotted to show the relation between yields and maturity.

Any explanation of the shape of a particular yield curve should be consistent with a combination of expectations, risk premia, and convexity.

In order to focus on the relation between yields and maturity, it is helpful to abstract from a number of other factors that can also affect a bond's yield. For example, bonds issued by private corporations or municipalities (including states and cities) are subject to credit risk, which means simply that the bonds are not default-free. In addition, corporate and municipal bonds are not as actively traded as Treasury securities, and this illiquidity can affect their yields. Some bonds (municipal bonds in particular but also some Treasury securities known as flower bonds) receive special tax treatment.[5] Many bonds, including some Treasury coupon bonds, are callable, which means the issuer has the right to buy them back at a predetermined price at some point in the future. The analysis of bond prices in this article abstracts from all of these factors other than maturity itself. As such, the analysis is most directly applicable to the default-free zero-coupon bonds traded in the STRIPS market.[6]

The Expectations Hypothesis

Historically, the expectations hypothesis has been the most widely used analytical tool to understand the shape of the yield curve.[7] In a nutshell, the expectations hypothesis says that the yield on long-term bonds equals the average of the expected one-period interest rates. If the expectations hypothesis were correct, the slope of the term structure could be used to forecast the future path of the interest rate. For example, if the yield curve were to slope upward at the short end, it would be because the interest rate is expected to rise. One problem with this version of the expectations hypothesis is that in fact the yield curve slopes upward at the short end on average even though interest rates do not rise on average. One way to explain this divergence is to assume that investors are simply wrong on average.[8] But a good theory should not imply that investors are wrong on average.

The expectations hypothesis can be easily modified to account for this persistent upward slope in a way that does not require systematic errors on the part of investors. Since bond prices do fluctuate over time, there is uncertainty (even for default-free bonds) regarding the return from holding a long-term bond over the next period. Moreover, the amount of uncertainty increases with the maturity of the bond. If there were a risk premium associated with that uncertainty, then the yield curve could slope upward on average without implying that interest rates increase on average. If the risk premium were constant, then changes in the slope of the yield curve would forecast changes in the future path of the interest rate. For example, if the slope of the yield curve were to increase, then it would have to be because the path of future interest rates is expected to be higher. This increase in the slope would also imply that future bond yields would be higher. But there is a problem with this version of the hypothesis as well.

Empirical tests of this extended version of the expectations hypothesis (using U.S. data) have shown that changes in the slope of the term structure do a poor job of forecasting changes in the bond yields. In fact, one widely used test shows that an increase in the slope of the yield curve may actually signal a decrease in the future yields. What went wrong in the theory? What went wrong was assuming that the risk premium was constant while in fact it varies over time. Movements in risk premia over time are responsible for a sizable fraction of the movements of the slope of the term structure. When risk premia increase, so does the slope even though expectations are unchanged. As a result, changes in the slope of the yield curve are often negatively correlated with changes in realized yields. It should be noted that the changes in risk premia that bring about this effect can (and do) occur without any change in the risk of the bonds. Risk premia are essentially covariances that change when either the amount of risk or the price of risk changes. In the discussion below, the effects of changing the amount of risk without changing the price of risk will be seen.

There is another feature of the yield curve that the expectations hypothesis has difficulty explaining. The zero-coupon yield curve slopes downward on average at the long end, typically over the range of twenty to thirty years. In other words, the yield on a thirty-year zero-coupon bond is typically below the yield on a twenty-year bond. The expectations hypothesis would suggest that this slope is due to either (1) a persistently incorrect belief that the interest rate will begin to fall about twenty years from now or (2) a decrease in the risk premium for bonds with maturities beyond twenty years, even though the uncertainty of the holding-period return for thirty-year bonds is greater than that for twenty-year bonds. Neither of these reasons is sensible.[9]

There is a sensible explanation, although it may seem counterintuitive at first, for the persistent downward slope of the term structure at the long end. The explanation has to do with the uncertainty regarding the future path of short-term rates. This uncertainty underlies the risk of holding bonds. (If there were no uncertainty regarding the future path, there would be no risk to holding default-free bonds.) Increases in this uncertainty lead (1) to increases in risk premia that increase the slope of the yield curve at the short end and (2) to decreases in the slope of the yield curve at the long end via the effect of convexity. Convexity (technically known as Jensen's inequality) arises from the nonlinear relation between bond yields and bond prices. As a consequence, a symmetric increase in uncertainty about yields raises the average price of bonds, thereby lowering their current yields. This effect is trivial at the short end of the yield curve where it plays no significant role, but it becomes noticeable and even dominant at the long end. The overall shape of the yield curve involves the trade-off between the competing effects of risk premia (which cause longer-term yields to be higher) and convexity (which cause longer term yields to be lower). Typically, the maximum yield occurs in the fifteen- to twenty-five-year maturity range of the zero-coupon yield curve.[10]

This article emphasizes that expectations do in fact play an important role in determining changes in the shape of the yield curve. The reason the expectations hypothesis fails is not that expectations do not matter; rather, the hypothesis fails because it says that nothing else matters. But as has been discussed, the expected future path of interest rates is only one of a number of important forces that shape the yield curve. Any explanation of the shape of a particular yield curve should be consistent with a combination of expectations, risk premia, and convexity.

1. This article is based in part on a memo written at the Federal Reserve Board coauthored with Christian Gilles.
2. "Term structure of interest rates" is another way of referring to the yield curve.
3. This implication—that the absence of arbitrage implies the existence of a risk premium that depends on the price of risk and the amount of risk—is the central message of the article. It is quite general and applies to other asset prices, not just bond prices.
4. The Treasury STRIPS program was introduced in February 1985. STRIPS is the acronym for separate trading of registered interest and principal of securities. The STRIPS program lets investors hold and trade the individual interest and principal components of eligible Treasury notes and bonds as separate securities.
5. Taxability is treated in the companion working paper (Fisher 2001).
6. Even in the STRIPS market, there are other factors at play. Although STRIPS are subject to taxation, once taxes are treated explicitly, the analysis that ignores taxes is essentially correct. Only in the comparison of taxable bonds with tax-exempt bonds is there a need to explicitly account for the effects of taxes. Other factors are more relevant for the internal structure of the STRIPS market. Principal STRIPS often trade at a premium relative to coupon STRIPS because principal STRIPS implicitly contain certain options. Consequently, the analysis presented here is most applicable to coupon STRIPS. (An explanation of the technical reasons for this relationship is beyond the scope of this article.)
7. Actually there are a number of different but related hypotheses, each of which is called the expectations hypothesis. See Cox, Ingersoll, and Ross (1981) for a discussion of a number of these competing hypotheses. The version described here is the one most often used.
8. Another way to explain the divergence is to assume that investors give some weight to very large increases in the interest rate that have not yet been observed. This is sometimes called the "peso problem."
9. There is another explanation—not related to the expectations hypothesis—that is sensible. The downward slope at the long end of the yield curve could, in principle, reflect a substantial demand for the longest-maturity (default-free) zero-coupon bond (for example, to insulate the value of insurance companies' long-term liabilities from interest-rate risk). Although the explanation is not unreasonable, it is unnecessary given the convexity effect discussed below.
10. It should be stressed that the yield curve typically reported in the newspaper is not the zero-coupon yield curve and may display a somewhat different shape owing to a variety of factors.

No-Arbitrage Conditions: An Introduction

This article has shown that the expectations hypothesis is not a good tool for studying the shape of the yield curve after all, but what will replace it? The fundamental problem with the expectations hypothesis is that it is taken from a world of perfect certainty, in which the expectations hypothesis is a condition for the absence of arbitrage opportunities, and transplanted into a world where there is uncertainty, in which the expectations hypothesis is not a condition for the absence of arbitrage opportunities. Fortunately, in recent years the theory of finance has produced better tools that allow one to directly apply the conditions guaranteeing the absence of arbitrage opportunities in a world where there is uncertainty. The tools were developed as an outgrowth of the famous Black-Scholes model of option prices. The revolution in asset pricing that was initiated by the Black-Scholes model ultimately carried over to bond pricing and the term structure.[11]

It is necessary to have a firm grasp of the no-arbitrage conditions in order to make sense of the shape of the yield curve.

An arbitrage involves trading securities in such a way as to generate something for nothing. Therefore, the conditions that guarantee the absence of arbitrage opportunities have to do with bond prices rather than bond yields. Thus, there is a bit of a paradox: in order to understand the term structure (bond yields), one must move away from the expectations hypothesis (which focuses on yields) and focus instead on bond prices.

The most powerful tool for understanding the term structure of interest rates is called the absence of arbitrage. (This phrase is shorthand for "the conditions that guarantee the absence of arbitrage opportunities.") An opportunity for arbitrage exists when there is an inconsistency in the prices of securities that allows a valuable payoff to be obtained at no cost. For example, if there are two ways to obtain a given payoff and if one way is cheaper than the other, then one can take advantage of this situation by buying the payoff the inexpensive way ("buy low") and selling it the expensive way ("sell high"). The difference is the profit from an arbitrage.[12]

Anyone who prefers more to less would like to take advantage of an arbitrage opportunity. Smart and greedy investors are constantly on the lookout for arbitrage opportunities. In an active and liquid market such as the market for U.S. Treasury securities, any arbitrage opportunities that appear are taken advantage of almost immediately. What happens to an arbitrage opportunity when someone tries to take advantage of it? Buying the payoff in the inexpensive way puts upward pressure on the cost of doing so, and selling the payoff in the expensive way puts downward pressure on the cost of doing so. The result is that an opportunity for arbitrage tends to go away when someone tries to take advantage of it.

In order to understand the conditions that guarantee the absence of arbitrage opportunities, it is useful to think of financial securities as claims to state-dependent payoffs. Different securities contain differing amounts of each possible payoff. Insurance policies are particularly simple in this regard because an insurance policy pays only when a specific state of the world occurs (for example, flood insurance pays only if there is a flood). Other securities may contain a wide variety of payoffs. Derivative securities, such as options, allow for the "disbundling" of the payoffs. For example, one can write a put option on a stock to insure against a fall in its price.

In principle, each of the payoffs in a security's bundle has a separate price. From this perspective, the price of the security is the sum of the (implicit) prices of the payoffs. Here is the key: As long as all of the individual payoffs have positive prices, there will be no opportunities for arbitrage. In other words, arbitrage opportunities arise only if one or more of the payoffs has a zero or negative price. The simplest example of an arbitrage is free insurance. (Free insurance generates something for nothing, but only in some states of the world.) More generally, a trading strategy that generates something for nothing involves buying and selling securities in such a way as to isolate and extract the mispriced payoffs.

These ideas can be illustrated concretely in a mundane setting. Consider a smart shopper at the grocery store. To keeps things simple, suppose the store sells only apples and oranges. Ordinarily when one goes to a store, one sees the posted prices for the produce. If one were to buy a bag containing, for example, two apples and three oranges, the price for the bag of produce would be computed from the prices posted for apples and oranges.

But consider a different kind of store. First of all, apples and oranges are sold mixed together in color-coded grocery bags. There are two combinations available: red bags each contain two apples and three oranges, and blue bags each contain three apples and two oranges. The store posts prices for the bags but not for apples or oranges separately. Even so, a smart shopper can figure out the implicit prices of apples and oranges from the prices of the bags. As long as the implicit prices of apples and oranges are both positive, there will be no arbitrage opportunities. But if the implicit price of either fruit is zero or negative, then one can get something for nothing.

There is another important difference between this store and an ordinary grocery store. Here one can sell bags of produce as well as buy them. For example, if one has two apples and three oranges, one can put them in a red bag (which the store conveniently supplies for free), sell it to the store, and receive the posted price. This repackaging allows a smart shopper who wants only apples to buy only apples. For example, the shopper can buy three red bags (containing a total of nine apples and six oranges), sell two blue bags (containing a total of four apples and six oranges), and end up with five apples left over. The net cost of the apples is the difference between the revenue from selling the two blue bags and the expense of buying the three red bags. Suppose the price of red bags is two dollars and the price of blue bags is three dollars. Then the net cost of apples is zero, and our smart shopper's "trading strategy" involving red and blue bags is an arbitrage: the smart shopper gets something for nothing.[13]

Faced with this arbitrage opportunity, why would the smart shopper limit the size of the trading strategy? Why not buy 3,000 red bags and sell 2,000 blue bags, netting 5,000 apples? Or why not buy three million red bags and sell two million blue bags, netting five million apples? Or why not buy three billion…? The reason, of course, is that at some point the purchases and sales will affect the prices of the bags, driving up the price of a red bag and driving down the price of a blue bag. The changing bag prices will indirectly affect the prices of the apples and oranges, raising the cost of apples. This dynamic reflects the general proposition stated earlier—attempting to take advantage of arbitrage opportunities tends to make them disappear.

How Useful Are No-Arbitrage Conditions?

For some securities, the absence of arbitrage may not be very useful. Consider the prices of Microsoft Corporation stock and Bank of America stock. The absence of arbitrage does not tell us much about the relation between these two stock prices because the state-contingent payoffs that the stocks "contain" do not overlap very much. For a different example, consider the price of Microsoft stock and an option to buy Microsoft stock. In this case, the payoffs are so closely related that the price of the option is completely determined by the no-arbitrage condition (that is, the Black-Scholes model).

The term structure of interest rates is more like the second example than the first. In the stock/option example, there are two risky securities but there is only one source of risk. Similarly for the term structure, there are more bonds than there are sources of risk. Because the payoffs to bonds of different maturities are highly correlated, the absence of arbitrage opportunities is quite useful. On the other hand, as noted above, there is an important difference between the term structure and the stock/option example. In that example, the state of the world is determined by the value of the stock. Because the stock is an asset, the formula for the value of an option is especially simple. In particular, investors' attitudes toward risk play no role. However, for the term structure, the state of the world is determined by the interest rate, and the interest rate is not the value of an asset. Consequently, investors' attitudes toward risk do play a role in the term structure.

Bond Prices and One-Period Returns

The Discount Function. A zero-coupon bond makes a single payment of one unit of payment at some fixed time in the future. For the purpose of exposition, let the unit of payment be the dollar, but the analysis would apply even if the payment were one peso or one "widget." Let $p(t, n)$ be the value at time t of a zero-coupon bond that matures at time $t + n$, where n is the term to maturity of the bond. Holding t fixed and varying n in $p(t, n)$ traces out the discount function at time t. The value of a zero-coupon bond tells how much a risk-free payment paid in the future is worth today. Two properties of bond prices

11. See Black and Scholes (1973). In the Black-Scholes model, the stock price summarizes the "state of the world" for option prices. In modeling the term structure, it is the interest rate, rather than the price of an asset, that summarizes the state of the world for bond prices. This difference accounts for the time lag in adapting the Black-Scholes paradigm to bond prices.
12. This example highlights the fact that when the "law of one price" is violated, an arbitrage opportunity exists.
13. In order to avoid arbitrage opportunities, the ratio of the cost of blue bags to the cost of red bags must be greater than two-thirds and less than three-halves. In this example the ratio was exactly three-halves, which allows arbitrage opportunities.

are immediately apparent. First, the value of one dollar to be delivered immediately is one dollar; that is, $p(t, 0) = 1$ (see the table). Second, the value of a dollar to be delivered in the infinite future is zero; that is, $\lim_{n\to\infty} p(t, n) = 0$.[14] Chart 1 shows a discount function.

One-Period Returns. Suppose one buys an n-period bond today and sells it next period when it becomes an $(n - 1)$-period bond. The bond that costs $p(t, n)$ today can be sold for $p(t + 1, n - 1)$ next period, as shown in the table. The holding-period return for this investment is

$$\frac{p(t+1, n-1)}{p(t, n)} - 1 = \frac{p(t+1, n-1) - p(t, n)}{p(t, n)},$$

which is the amount one has at the end of the period divided by the amount one invested at the beginning of the period minus one.

In general, it is not known in advance what the price of an $(n - 1)$-period bond will be in the next period, and consequently the holding period return is uncertain. The central point of this article is to uncover the relation between the average holding period return and this uncertainty.

For now, focus on the holding-period return on a one-period bond, which *is* known in advance since the one-period bond delivers one dollar without fail next period (see the table). This return can be defined as the one-period risk-free interest rate. A one-period bond can be purchased today for $p(t, 1)$. The amount repaid next period equals the amount loaned plus interest:

$$1 = [1 + r(t)]\, p(t,1). \tag{1}$$

Equation (1) can be solved for the one-period risk-free interest rate:

$$r(t) = \frac{1}{p(t,1)} - 1.$$

Today's Price: The Present Value of Next Period's Price

The relation between bond prices today and bond prices in the next period is examined below. This examination involves forming a portfolio today that costs nothing and finding out what it will be worth in the next period. An n-period bond will be purchased and financed by borrowing its cost at the one-period risk-free rate. (In other words, one-period bonds of equal value will be sold.) The net cash flow at time t is zero. Next period, the long-term bond will be sold and the debt repaid (principal plus interest). The table summarizes the net cash flows for this trading strategy.

Net Cash Flows

Buying an *n*-Period Bond and Holding until Maturity

Today (t)	At maturity ($t + n$)
$-p(t, n)$	1

Buying an *n*-Period Bond and Holding One Period

Today (t)	Next period ($t + 1$)
$-p(t, n)$	$p(t + 1, n - 1)$

Buying a One-Period Bond

Today (t)	Next period ($t + 1$)
$-p(t,1)$	1

Financing the Purchase of an *n*-Period Bond with One-Period Borrowing

Today (t)	Next period ($t + 1$)
0	$p(t + 1, n - 1) - [1 + r(t)]\, p(t, n)$

If it is known today that $p(t + 1, n - 1)$ will be greater than $[1 + r(t)]p(t, n)$, then the trading strategy is an arbitrage: something (next period) is obtained for nothing (today). On the other hand, if it is known today that $p(t + 1, n - 1)$ will be less than $[1 + r(t)]p(t, n)$, the trading strategy can be modified to make it an arbitrage. Instead of buying the long-term bond and selling some one-period bonds, sell the long-term bond and buy the one-period bonds. The net cash flows for this trading strategy are the same as for the original trading strategy except that the signs are reversed. The upshot is that in a world of no uncertainty, the absence-of-arbitrage condition for bond prices is

$$p(t + 1, n - 1) - [1 + r(t)]\, p(t, n) = 0. \tag{2}$$

Equation (2) can be solved for today's price of the long-term bond:

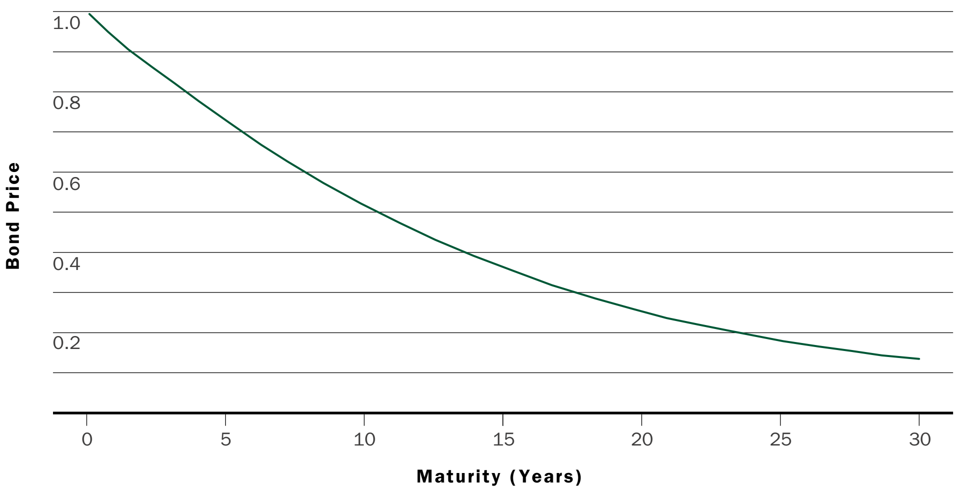
$$p(t, n) = \frac{p(t+1, n-1)}{1 + r(t)}. \tag{3}$$

In other words, the price of the bond today is the present value of its price in the next period. Another way to express this is

$$\frac{p(t+1, n-1) - p(t, n)}{p(t, n)} = r(t),$$

which says that the (net) return on a bond equals the risk-free interest rate.

CHART 1 The Discount Function: The Price of Zero-Coupon Bonds



Uncertainty

All of the bonds dealt with in this article are default-free—that is, all promised payments are made in full and on time. Nevertheless, these bonds have risk prior to maturity: they can gain or lose value. This uncertainty regarding bond prices can (and will) be linked to the uncertainty regarding interest rates, and this latter uncertainty can be viewed as more fundamental. Nevertheless, the effect of that uncertainty on bond prices and on the conditions that guarantee the absence of arbitrage opportunities can be studied without reference to the underlying interest rate uncertainty.

In the previous section, an absence-of-arbitrage condition based on knowing next period's bond value with certainty was established. (See equation [2].) What if the bond's value in the next period is not known with certainty? What if its possible values can make the net cash flow for a trading strategy sometimes positive and sometimes negative? In this case, the trading strategy is not an arbitrage. The conditions for the absence of arbitrage opportunities are not sufficiently restrictive to completely establish the relation between today's price and next period's price when there is uncertainty. Nevertheless, they do put enough structure on bond prices to provide useful results.

Heads or Tails? All bond prices tend to go up and down together. When the short-term interest rate rises, all bond prices tend to fall, and, conversely, when the short-term interest rate falls, all bond prices tend to rise. To keep things simple, suppose there are only two possible discount functions in the next period. The flip of an unbiased coin will determine which discount function is realized.[15] In other words, if one were to buy an n-period bond today, there would be two possibilities for the price of an $(n - 1)$-period bond in the next period, with the actual outcome determined by the flip of a coin. The notation can be simplified a bit by limiting consideration to just today (time t) and tomorrow (time $t + 1$). Let the price today of an n-period bond be p_n. (See the appendix for a list of the notations used in this article.) If the coin comes up heads, the price of the bond tomorrow will be p^H_{n-1} and if it comes up tails the price will be p^T_{n-1}. Let $\bar{p}_{n-1}$ denote the average price of the bond in the next period:

$$\bar{p}_{n-1} = \frac{p^H_{n-1} + p^T_{n-1}}{2}.$$

Let δ^p_{n-1} denote the volatility of the bond price in the next period:

$$\delta^p_{n-1} = \frac{p^H_{n-1} - p^T_{n-1}}{2}.$$

Volatility is a measure of the riskiness of the investment. It is related to the variance and the standard

14. This property holds if the interest rate is always positive. If the interest rate can be negative, then the discount function does not have to go to zero. So-called nominal interest rates cannot take on negative values because one can always hold currency instead (which has a nominal return of zero).
15. An unbiased coin has a fifty-fifty chance of coming up either heads or tails.

deviation. The variance is the average squared deviation from the mean,

$$\frac{1}{2}(p^{H}_{n-1} - \bar{p})^2 + \frac{1}{2}(p^{T}_{n-1} - \bar{p})^2 = (\delta^{p}_{n-1})^2,$$

and the standard deviation is the square root of the variance, which is the absolute value of the volatility, δ^{p}_{n-1}. Volatility is more useful than standard deviation because volatility's sign plays a role in characterizing whether the risk is bad or good. (An insurance policy is an example of an investment that has good risk because it pays off in bad times). The two possible values of the $(n-1)$-period bond in the next period as determined by the coin flip are $p^{H}_{n-1} = \bar{p}_{n-1} + \delta^{p}_{n-1}$ and $p^{T}_{n-1} = \bar{p}_{n-1} - \delta^{p}_{n-1}$. Chart 2 plots two postflip discount functions and their average.

Although there is no need to specify which of the two postflip prices is greater, for the sake of concreteness assume $p^{H}_{n-1} > p^{T}_{n-1}$ and therefore $\delta^{p}_{n-1} > 0$.

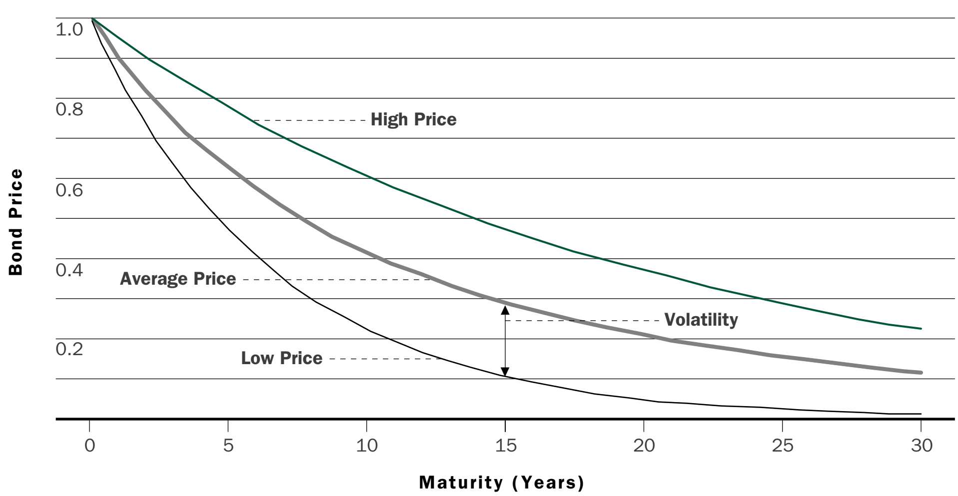
The Absence of Arbitrage Opportunities under Uncertainty: Part I. Recall that an arbitrage is a trading strategy that generates "something for nothing." Now that uncertainty has been introduced, what the absence of arbitrage means needs to be reexamined. Suppose there were a trading strategy that had zero net cash flow today. In other words, the trading strategy costs nothing. The conditions for absence-of-arbitrage opportunities can be stated in terms of the net cash flows next period as follows: Either (1) they are both zero (as they must be in the case of no uncertainty) or (2) one is positive and the other is negative.

To see why these conditions must be so, suppose the contrary were true. If, for example, the net cash flows next period were both positive, then the trading strategy would clearly generate an arbitrage: one would get something in the next period—in all states of the world—for nothing today. On the other hand, suppose only one net cash flow were positive next period and the other were zero. This situation too would generate an arbitrage: just like free insurance, it would cost nothing today and make positive payoffs in some states of the world next period, without the possibility of negative payoffs. Alternatively, if both payoffs were negative (or one negative and the other zero), one could reverse the signs of the payoffs by reversing the positions in the trading strategy (for example, selling instead of buying, lending instead of borrowing).

This analysis can be applied to the following simple trading strategy: Buy an n-period bond today and finance its purchase price with one-period risk-free borrowing. The net cash flow today is zero, and the two possible net cash flows in the next period are $p^{H}_{n-1} - (1 + r)p_n$ and $p^{T}_{n-1} - (1 + r)p_n$. If there were no uncertainty ($p^{H}_{n-1} = p^{T}_{n-1}$), the no-arbitrage condition would be that both net cash flows in the next period must be zero. But when there is uncertainty ($p^{H}_{n-1} \neq p^{T}_{n-1}$), the two net cash flows cannot both be zero. In this case, the no-arbitrage condition is that $(1 + r)p_n$ must lie between p^{H}_{n-1} and p^{T}_{n-1}, thereby guaranteeing that one net cash flow is positive and the other negative.[16]

Today's Price: The Present Value of Next Period's Adjusted Average Price. By aping the

CHART 2 Two Postflip Discount Functions and Their Average



Note: The volatility is a measure of the uncertainty.

relation between today's price and next period's price that was established when there was no uncertainty, some guidance in how to proceed can be obtained. The simplest and most natural way to modify equation (3) so that it makes sense when the value of a bond next period is not certain is to replace the uncertain price next period with its average:

$$p_n = \frac{\bar{p}_{n-1}}{1+r}, \tag{4}$$

where $r = r(t)$. Equation (4) says that today's bond price is the present value of the "expected value" of tomorrow's bond price.[17] Equation (4) can be written as

$$\frac{\bar{p}_{n-1} - p_n}{p_n} = r,$$

which says that the expected return on a long-term bond equals the risk-free rate (that is, the risk-free return on a one-period bond).

But why should investors be willing to earn exactly the risk-free rate on average? If the uncertainty associated with owning bonds contributes to the overall uncertainty of investors' lives, investors may require a higher average return to take on this additional risk. On the other hand, if the uncertainty associated with owning bonds reduces the overall uncertainty of their lives, they may accept an average return that is less than the risk-free rate.

In order to account for how investors feel about the kind of risk they face, an adjustment term (a_{n-1}) can be incorporated into the formula for today's bond price:

$$p_n = \frac{\bar{p}_{n-1} - a_{n-1}}{1+r}. \tag{5}$$

The numerator on the right-hand side of equation (5), $\bar{p}_{n-1} - a_{n-1}$, is referred to as the adjusted average price. Equation (5) says that today's price is the present value of next period's adjusted average price. Equation (5) can be rearranged to express the expected return for the bond:

$$\frac{\bar{p}_{n-1} - p_n}{p_n} = r + \frac{a_{n-1}}{p_n}. \tag{6}$$

Equation (6) says that the average holding-period return for a bond is the risk-free rate plus an additional term that somehow accounts for the amount and type of risk involved.

The adjustment term, which can be positive, negative, or zero, provides great flexibility within certain bounds. It has already been shown that $(1 + r)p_n$ must be between p^H_{n-1} and p^T_{n-1} in order to avoid arbitrage opportunities. Given equation (5), these boundaries imply that the adjusted average price, $p_{n-1} - a_{n-1}$ must also be between p^H_{n-1} and p^T_{n-1} (since $[1 + r]p_n$ equals the adjusted average price). Within these bounds, any bond price (or expected return) can be obtained with a suitable choice for the adjustment term. Stated differently, no bond prices in this range can be ruled out. In other words, thus far the theory of bond pricing under uncertainty provides very little structure. To obtain more structure, the way in which two different long-term bonds interact must be examined.

The Absence of Arbitrage Opportunities under Uncertainty: Part II. This section examines arbitrage opportunities that involve simultaneously buying and selling bonds with different maturities in order to form a risk-free portfolio.[18] In the course of this examination, the condition that guarantees the absence of arbitrage opportunities will be uncovered. As will be seen, that condition has something important to say about how the adjustment terms on different bonds relate to each other.

Consider the following portfolio of two bonds: Buy one n-period bond and buy (or sell) some m-period bonds (where m is different from n). Let b denote the number of m-period bonds purchased (where b is negative if they are sold). The cost of this portfolio today is $p_n + bp_m$, which may be positive, negative, or zero. Let π^H and π^T represent the possible values of this portfolio next period. The two values are $\pi^H = p^H_{n-1} + bp^H_{m-1}$ and $\pi^T = p^T_{n-1} + bp^T_{m-1}$.

Each of these two bonds is risky in isolation. But since the uncertainty for each of these bonds is driven by the same underlying source of risk, it is possible to combine the bonds in such a way as to reduce the overall risk. In fact, there is a value for b (call it b^*) that makes the portfolio completely risk-free. In other words, the value of the portfolio next period is the same in both states of the world

16. This condition guarantees that the realized return on the n-period bond is greater than r if the coin comes up heads and less than r if it comes up tails.

17. Equation (4) is an expectations hypothesis, albeit one based on bond prices rather than on interest rates. In the companion working paper (Fisher 2001), the typical statement of the expectations hypothesis is discussed, namely, that forward rates are expectations of future one-period returns.

18. See Vasicek (1977) for an early application of the absence of arbitrage to the term structure of interest rates.

so that $\pi^H = \pi^T$. For this statement of value to be true, b^* must satisfy

$$p^H_{n-1} + b^* p^H_{m-1} = p^T_{n-1} + b^* p^T_{m-1}. \quad (7)$$

Equation (7) can be solved for

$$b^* = -\left(\frac{p^H_{n-1} - p^T_{n-1}}{p^H_{m-1} - p^T_{m-1}}\right) = -\left(\frac{\delta^p_{n-1}}{\delta^p_{m-1}}\right). \quad (8)$$

Since b^* is negative, this portfolio involves selling some m-period bonds. In other words, b^* is a hedge ratio—it tells us how to use one bond to hedge the risk of another so that on balance there is no risk at all.[19] Let π^* denote the known payoff to this risk-free portfolio. Since π^* can be computed from either side of Equation (7), it must equal the average of the two sides:

$$\pi^* = \bar{p}_{n-1} + b^* \bar{p}_{m-1}.$$

The cost of the risk-free portfolio is $p_n + b^* p_m$. Consider a trading strategy in which the risk-free portfolio is financed with one-period borrowing. Since the net cash flow today is zero and the net cash flow next period is certain, there will be an arbitrage opportunity unless the cash flow next period is zero. Therefore, the condition for the absence of arbitrage opportunities is

$$\pi^* - (1 + r)(p_n + b^* p_m) = 0. \quad (9)$$

In order to see what this condition implies for the adjustment terms of the two bonds, equation (5) can be used to re-express the cost of this portfolio using the adjusted average prices:

$$\begin{aligned} p_n + b^* p_m &= \overbrace{\left(\frac{\bar{p}_{n-1} - a_{n-1}}{1+r}\right)}^{p_n} + b^* \overbrace{\left(\frac{\bar{p}_{m-1} - a_{m-1}}{1+r}\right)}^{p_m} \\ &= \frac{\overbrace{\left(\bar{p}_{n-1} + b^* \bar{p}_{m-1}\right)}^{\pi^*}}{1+r} - \frac{\left(a_{n-1} + b^* a_{m-1}\right)}{1+r} \\ &= \frac{\pi^*}{1+r} - \frac{\left(a_{n-1} + b^* a_{m-1}\right)}{1+r}. \end{aligned} \quad (10)$$

At this point $p_n + b^* p_m$ can be replaced in the no-arbitrage condition (9) by the last line on the right-hand side of equation (10), so that the no-arbitrage condition becomes

$$a_{n-1} + b^* a_{m-1} = 0. \quad (11)$$

Equation (11) shows that the adjustment terms play a central role in the condition that guarantees the absence of arbitrage opportunities.

The final expression for the absence-of-arbitrage condition can now be found. Substituting the solution for b^* given in equation (8) into equation (11) and rearranging produces

$$\frac{a_{n-1}}{\delta^p_{n-1}} = \frac{a_{m-1}}{\delta^p_{m-1}}. \quad (12)$$

Equation (12) says that the ratio of the adjustment term to the bond-price volatility must be the same for both bonds. This common ratio is called the price of risk. Let λ denote the price of risk, so that

$$\lambda = \frac{a_{n-1}}{\delta^p_{n-1}} = \frac{a_{m-1}}{\delta^p_{m-1}}.$$

The absence-of-arbitrage condition does not say whether the price of risk is big or small or even whether it is positive, negative, or zero; it says only that it must be the same for all bonds.

The Adjustment Term Is the Risk Premium. Given the absence-of-arbitrage condition just established, the adjustment term can be written as

$$a_{n-1} = \lambda \delta^p_{n-1}, \quad (13)$$

where λ is the price of risk and δ^p_{n-1} is the volatility of the bond's price. Equation (13) can be expressed as *risk premium = price of risk × amount of risk*. In other words, the adjustment term is the risk premium and the volatility of the bond price is the amount of risk that earns a premium.

The condition for the absence of arbitrage opportunities can be stated in terms of the expected return on a bond by substituting equation (13) into equation (6):

$$\frac{\bar{p}_{n-1} - p_n}{p_n} = r + \lambda\left(\frac{\delta^p_{n-1}}{p_n}\right), \quad (14)$$

where δ^p_{n-1}/p_n is the relative volatility of the bond price; it measures the volatility of the holding-period return. Equation (14) can be expressed as *expected return = risk-free rate +* (*relative*) *risk premium*, where the relative risk premium equals the price of risk times the amount of risk as measured by the relative volatility of the bond price. In other words, the extra return one gets (from the risk premium) depends on the amount of risk (δ^p_{n-1}/p_n) and the price of risk (λ). If either is zero, there is no risk premium.[20]

Bond Yields and Convexity

In this section, the yield to maturity is defined and the absence-of-arbitrage conditions are re-expressed in terms of yields.

Yield to Maturity. Suppose an n-period bond were purchased. If it were held until it matured, what would the return on the investment be? If the amount invested were $p(t, n)$ and the amount returned were one, the total gross return would be simply

$$\frac{1}{p(t,n)}.$$

From the total gross return, the gross return per period could be computed:

$$p(t,n)^{-1/n} = \frac{1}{\sqrt[n]{p(t,n)}}$$

since

$$\overbrace{\frac{1}{\sqrt[n]{p(t,n)}} \times \frac{1}{\sqrt[n]{p(t,n)}} \times \cdots \times \frac{1}{\sqrt[n]{p(t,n)}}}^{n \text{ times}} = \left(\frac{1}{\sqrt[n]{p(t,n)}}\right)^n = \frac{1}{p(t,n)}.$$

Typically, however, it is not the gross return period that is used to characterize the return but rather the net return per period. The net per-period return is called the yield to maturity (or simply the yield). The yield is like an "interest rate." There is a degree of freedom in computing interest rates: how many times per period is interest assumed to be compounded? The fact that there are only two points in time under consideration (the beginning of the period and the end of the period) does not resolve the issue since one is free to quote the interest rate as if there were subperiods over which compounding takes place. Let $y^i(t, n)$ denote the value of y that solves the following equation for a given i: $(1 + y/i)^i = p(t, n)^{-1/n}$ $i = 1, 2, 3, \cdots$. The solution is $y^i(t, n) = i[p(t, n)^{-1/(ni)} - 1]$. Given the price of the bond, each and every $y^i(t, n)$ has a right to be called the net return per period. How one chooses to quote the return (that is, the value one chooses for i) is merely a matter of convenience.

There are two rates of compounding that are particularly convenient to use, and they happen to lie at opposite ends of the compounding spectrum. The first case is called simple compounding, where interest is compounded only once per period ($i = 1$):

$$y^1(t,n) = \frac{1}{\sqrt[n]{p(t,n)}} - 1.$$

The one-period risk-free rate used above is computed using simple compounding: $r(t) = y^1(t, 1)$.

The second case is called continuous compounding, where interest is compounded infinitely many times per period ($i = \infty$). Let $y(t, n)$, without the symbol for infinity, denote continuously compounded yields. Fortunately there is a simple formula for continuously compounded yields:[21]

$$y(t,n) = \frac{-\log[p(t,n)]}{n}.$$

In discussing the yield curve, continuously compounded yields will be used. Chart 3 plots the yield curve computed from the discount function that is plotted in Chart 1.

A First Look at the Expectations Hypothesis. The expectations hypothesis can be expressed in a number of equivalent ways. Here is one way to express it: The long-term yield equals the average of the (expected) one-period yields. Of course, when there is no uncertainty, expected one-period yields equal the actual one-period yields. In this case the expectations hypothesis can be expressed as

$$y(t,n) = \frac{y(t,1) + y(t+1,1) + \cdots + y(t+n-1,1)}{n}. \quad (15)$$

But equation (15) is not just a statement of the expectations hypothesis; when there is no uncertainty it is also a statement of the absence of arbitrage opportunities.

That equation (15) is an absence of arbitrage condition when there is no uncertainty can be demonstrated as follows. According to equation (3), the value of an n-period bond today is the

19. The use of a hedge ratio is analogous to delta hedging in option pricing.
20. In an appendix to the companion working paper (Fisher 2001), it is shown that the risk premium can be interpreted as a covariance with a marketwide factor. As a consequence, equation (14) has the same form as the capital asset pricing model (CAPM), in which the expected return on an equity equals the risk-free rate plus a risk-premium that depends on the covariance with the market portfolio.
21. Formally, the continuously compounded yield is the limit of $y^i(t, n)$ as i goes to infinity.

present value of next period's value of an $(n-1)$-period bond:

$$p(t,n) = \frac{p(t+1,n-1)}{1+r(t)} = p(t,1)p(t+1,n-1). \quad (16)$$

The second equality follows from $p(t,1) = 1/[1+r(t)]$. Equation (3) can now be applied to the price of an $(n-1)$-period bond at time $t+1$:

$$\begin{aligned} p(t+1,n-1) &= \frac{p(t+2,n-2)}{1+r(t+1)} \\ &= p(t+1,1)p(t+2,n-2). \end{aligned} \quad (17)$$

Combining equations (16) and (17) produces $p(t,n) = p(t,1)\,p(t+1,1)\,p(t+2,n-2)$. This process can be continued until the price of a long-term bond ends up expressed as the product of one-period bond prices:

$$p(t,n) = p(t,1)\,p(t+1,1)\cdots p(t+n-1,1). \quad (18)$$

By taking logs of both sides of equation (18) (recall that $\log[ab] = \log[a] + \log[b]$) and dividing by $-n$, equation (15) is obtained.

Since the expectations hypothesis is equivalent to the absence-of-arbitrage conditions when there is no uncertainty, it is understandable that some people may have thought that the same equivalence is true where there is uncertainty—understandable, but wrong.

Uncertainty and Convexity. At this point, the effect of uncertainty on bond yields can be examined. The way in which uncertainty per se drives a wedge between the expected future yields and current yields will be seen. The relation between bond prices and bond yields is not linear; consequently, the yield computed from the average bond price is less than the average yield.[22] In this section, this point is demonstrated and its consequences explored.

The relation between bond yields and bond prices, $y_n = -\log(p_n)/n$, is plotted in Chart 4 for ten- and twenty-year bonds. The two primary features that are evident in the chart are (1) the negative slope and (2) the fact that the graph of the function is "bowed in" toward the origin—in other words, convex to the origin.[23] This second feature is called convexity. Chart 4 shows that a twenty-year bond has more convexity than a ten-year bond.

Convexity drives a wedge between the average yield and the yield of the average price (see Chart 5). There are two outcomes that depend on the flip of the coin: (1) high price and low yield or (2) low price and high yield. The average price and the average yield are at the midpoint of the straight line that connects the two outcomes. But the yield computed from the average price lies on the curved line, below the average yield.

The next step is to derive an algebraic expression for the effect of convexity. Using continuous compounding, the postflip yields can be computed from the two postflip bond prices:

$$y_{n-1}^H = \frac{-\log(p_{n-1}^H)}{n-1} \quad \text{and} \quad y_{n-1}^T = \frac{-\log(p_{n-1}^T)}{n-1}.$$

The postflip yields can be expressed as

$$y_{n-1}^H = \bar{y}_{n-1} + \delta_{n-1}^y \quad \text{and} \quad y_{n-1}^T = \bar{y}_{n-1} - \delta_{n-1}^y,$$

CHART 3 The Zero-Coupon Yield Curve Computed from the Discount Function in Chart 1

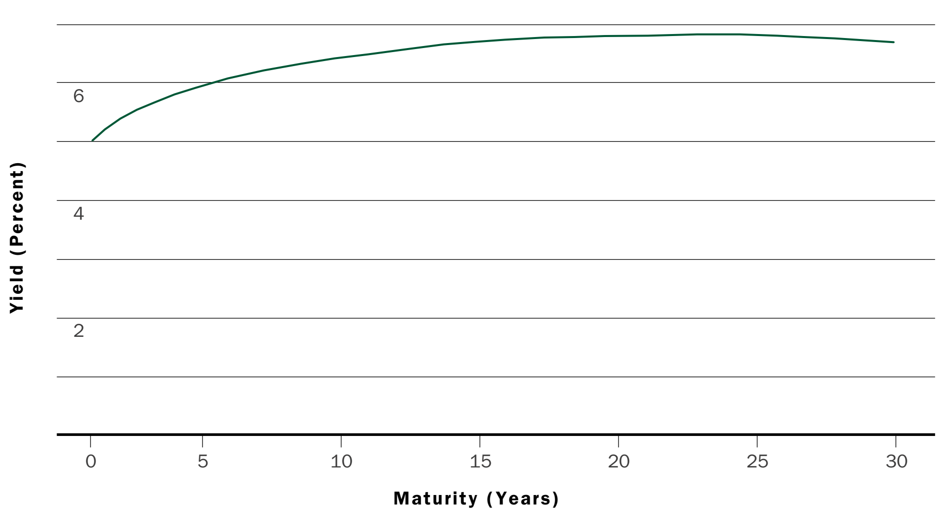


CHART 4 The Yield of Zero-Coupon Bonds as a Function of the Price

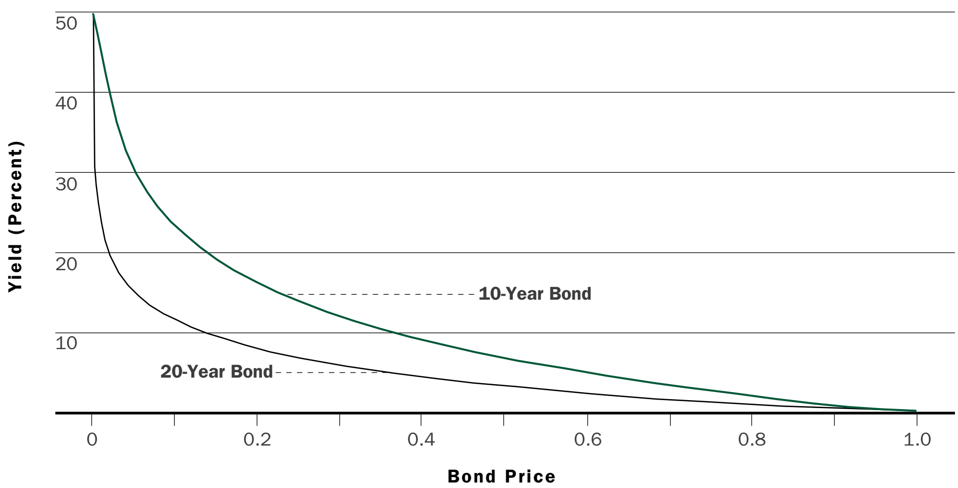
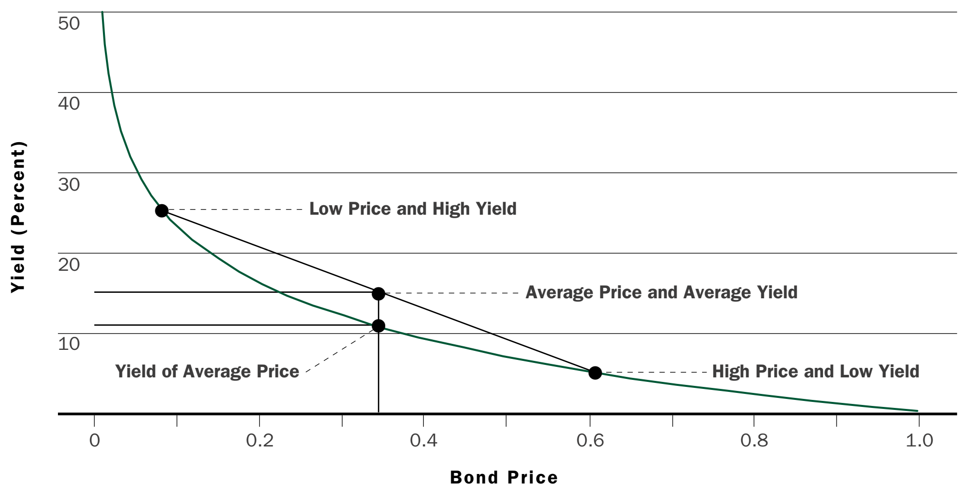


CHART 5 The Convexity Effect on the Average Yield and the Yield of the Average Price



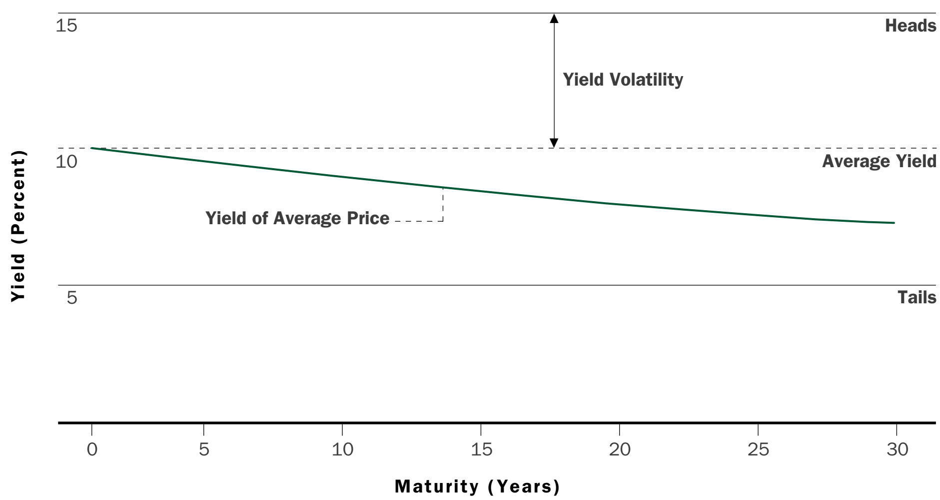
where the average yield and the volatility of the yield are given by

$$\bar{y}_{n-1} = \frac{y^H_{n-1} + y^T_{n-1}}{2} \quad \text{and} \quad \delta^y_{n-1} = \frac{y^H_{n-1} - y^T_{n-1}}{2}.$$

As an example, suppose that the current one-period yield equals the average long-term yield, $y_1 = \bar{y}_{n-1} = \bar{y}$, for all $n \geq 2$, and also suppose that the yield volatility is constant, $\delta^y_{n-1} = \delta^y$. Then the yield on an n-period bond (that is, the yield curve) can be approximated by $y_n \approx \bar{y} - n(1/2)(\delta^y)^2$ as long as n is not too big. This approximation illustrates the three main features of convexity: (1) Convexity has the effect of reducing yields. (2) The convexity effect is larger for longer-term bonds. (3) The convexity effect depends on the variance of the uncertainty about yields. See Chart 6 for an example in which $\bar{y} = 0.10$ and $\delta^y = 0.05$.[24]

22. This result is an example of Jensen's inequality.
23. These two features summarize the first two derivatives of the bond yield with respect to the price. The first derivative is negative, and the second derivative is positive.
24. The graph is drawn using the exact formula upon which the approximation is based. See Part 2 of the companion working paper (Fisher 2001) for the details.

CHART 6 An Example of the Convexity Effect on Zero-Coupon Yield Curves



Note: Where $\bar{y} = 0.10$ and $\delta^y = 0.05$

This example illustrates the depressing effect of uncertainty on bond yields via the convexity effect. As noted in the introductory section, risk premia will also have an effect on the shape of the term structure. Unfortunately, it is beyond the scope of this article to treat the effect of risk premia in a fully rigorous way. (The interested reader will find a full account in Part 2 of the companion working paper [Fisher 2001]).

Conclusion

The conditions that ensure the absence of arbitrage opportunities provide structure for the analysis of the yield curve by appealing to rationality at its most basic level. The central implication of the no-arbitrage conditions is that the risk premium for an asset can be decomposed into the amount of risk (measured by volatility) and the price of risk (which reflects investors' attitudes toward risk), where the price of risk is common to all assets. The forces that shape the yield curve are channeled through this feature. The nonlinear relation between bond yields and bond prices leads to surprising and even counterintuitive results. It is necessary to have a firm grasp of the no-arbitrage conditions in order to make sense of the shape of the yield curve. Analysis that ignores the implications of the no-arbitrage conditions will inevitably lead one astray.

REFERENCES

Black, Fischer, and Myron Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81 (May/June): 637–54.

Cox, John C., Jonathan E. Ingersoll Jr., and Stephen A. Ross. 1981. A reexamination of traditional hypotheses about the term structure of interest rates. *Journal of Finance* 36 (September): 769–99.

Fisher, Mark. 2001. Forces that shape the yield curve: Parts 1 and 2. Federal Reserve Bank of Atlanta Working Paper 2001–3, March.

Vasicek, Oldrich. 1977. An equilibrium characterization of the term structure. *Journal of Financial Economics* 5, no. 2:177–88.

Notations

Bond-Price Uncertainty

p_n Value of an n-period bond (same as $p[t, n]$)

r One-period interest rate (same as $r[t]$)

p^H_{n-1} Value next period of an $(n - 1)$-period bond if the coin comes up heads

p^T_{n-1} Value next period of an $(n - 1)$-period bond if the coin comes up tails

$\bar{p}_{n-1}$ Average value of an $(n - 1)$-period bond (preflip)

δ^p_{n-1} Volatility of the bond price (amount of risk)

a_{n-1} Adjustment term (risk premium)

Bond Portfolios

b Number of m-period bonds held in portfolio

b^* Number of m-period bonds held to make the portfolio risk-free

π^H Value of the portfolio next period if the coin comes up heads

π^T Value of the portfolio next period if the coin comes up tails

π^* Value of a risk-free portfolio (holding b^* m-period bonds)

λ Price of risk

Bond Yields and Compounding

$y^i(t, n)$ Yield at time t on an n-period bond, compounded i times per period

$y^1(t, n)$ Yield computed with simple compounding

$y^1(t, 1)$ Same as $r(t)$

$y(t, n)$ Continuously compounded yield (same as $y^\infty[t, n]$)

Bond Yield Uncertainty (Continuously Compounded)

y_n Yield at time t on an n-period bond (same as $y[t, n]$)

y^H_{n-1} Yield next period on an $(n - 1)$-period bond if the coin comes up heads

y^T_{n-1} Yield next period on an $(n - 1)$-period bond if the coin comes up tails

$\bar{y}_{n-1}$ Average yield on an $(n - 1)$-period bond (preflip)

δ^y_{n-1} Volatility of the yield