After a couple of months, his patriotic zeal got on my nerves so much I began to question whether I agreed with him about communism being evil. I agreed it was a bad idea but no longer felt so sure it would ruin the planet. I began to consider the danger of blind faith in, or blind hatred of, a single idea, any idea.

——Robert Wideman, *Unexpected Prisoner: Memoir of a Vietnam POW*, 2016

CHAPTER 1

The Yield Curve

The main measure of return associated with holding bonds is the *yield to maturity* (YTM) or *gross redemption yield* (GRY). In developed markets there is usually a large number of bonds trading at one time, at different yields and with varying terms to maturity. Investors and traders frequently examine the relationship between the yields on bonds that are in the same class. Plotting yields of bonds that differ only in their term to maturity produces the *yield curve*. The yield curve is an important indicator and knowledge source of the state of a debt capital market.[1] It is sometimes referred to as the *term structure of interest rates*, but strictly speaking this is not correct, as this expression should be reserved for the zero-coupon yield curve only. We shall examine this in detail later.

Much of the analysis and pricing activity that takes place in the bond markets revolves around the yield curve. The yield curve describes the relationship between a particular redemption yield and a bond's maturity. We should be aware that the GRY of a bond is only ever the actual yield one receives during the period one holds the bond if certain specific, and generally unrealistic, conditions are met. However, we will leave the discussion of this for later.

Plotting the yields of bonds along the maturity term structure will give us our yield curve. It is very important that only bonds from the same class of issuer or with the same degree of liquidity are used when plotting the yield curve. For example, a curve may be constructed for UK gilts or for AA-rated sterling Eurobonds, but not a mixture of both, because gilts and Eurobonds are bonds from different class issuers. The primary yield curve

[1] The author was tickled to read this description from someone obviously as excited about the yield curve as he is; Ryan (1997) writes:

"The future . . . of the global economy . . . may well rest on the success of how we finance the [yield] curve. God bless the Treasury Yield Curve!"

in any domestic capital market is the government bond yield curve, so for example, in the US market it is the US Treasury yield curve. With the advent of the euro currency in 11 (ultimately 19) countries of the European Union, in theory any euro-currency government bond can be used to plot a default-free euro yield curve. In practice, only bonds from the same government are used, as for various reasons different bonds within euro-currency countries trade at different yields. Outside the government, bond markets yield curves are plotted for Eurobonds, money market instruments, off-balance sheet instruments, in fact virtually all debt market instruments. Therefore it is always important to remember to compare like-for-like when analysing yield curves across markets.

In this chapter, we look at the yield to maturity yield curve as well as other types of yield curves that may be constructed. We also consider how to derive spot and forward yields from a current redemption yield curve. The main emphasis though, is on interpreting the shape of the yield curve, and explaining why it assumes the shapes it does. Later in the book we examine more advanced techniques for fitting, analysing, and interpreting the yield curve.

First though, we introduce the yield curve for beginners, of course experienced practitioners and graduate students may skip this part.

THE YIELD CURVE FOR BEGINNERS

This section is a summary of the importance and application of the yield curve. It was originally written with private investors in mind, so market practitioners may wish to skip this part.

What is the Yield Curve?

The yield curve is a graph that plots the yield of various bonds against their term to maturity. In other words, it is a snapshot of the current level of yields in the market. It is not an historical graph, that is, it does not show the level of yields over time. That would be an historical price (or yield) chart.

Yield curves are like football . . . it is very easy to grasp the basics, but difficult to become expert at (to continue the football analogy, akin to being good enough to play on a Sunday-morning park football team and being good enough to play on a team that included David Beckham, Steven Gerrard, and Michael Owen). Let us imagine that we looked up gilt yields in the *Financial Times* on a day in August 2002 and saw the following:

TABLE 1.1 Gilt yields

Gilt	Red Yield[2]
Tr 8% 03	3.79
Tr 5% 04	4.00
Tr 7.25% 07	4.62
Tr 5% 12	4.70
Tr 8.75% 17	4.74
Tr 8% 21	468
Tr 4.25% 32	4.52

Table 1.1 shows the yields for gilts of 1-, 2-, 5-, 10-, 15-, 20-, and 30-year maturity. (The 8% 2021 gilt is slightly under 20 years maturity but it will do for our purposes – there is no gilt that matures in 2022 at the time we are looking at this.)

We open up Microsoft Excel[3] and write down two columns, one for "maturity" and one for "yield". The years to maturity column forms the x-axis of the graph, while the yield forms the y-axis. Then we use the Excel "chart wizard" and it plots our graph for us! The result is as shown in Figure 1.1.

This curve looks about right. Intuitively, we expect that yields increase the greater the maturity. If we lend one person an amount of money for one year and another person the same amount of money for 10 years, we would not charge them both the same rate of interest (assuming both had the same credit risk) – we would most likely charge a higher rate to the 10-year borrower, for two reasons:

i. inflation will erode the value of the loan over the longer term, and;
ii. while the longer-dated borrower may be the same credit quality as the short-dated borrower, there are other risks. For example, the long-dated borrower may not be around in 10 years' time. Therefore, as a lender, we need a higher return to compensate for the greater risk the further out into the future we lend money.

[2] "Red" means redemption yield.
[3] The author has no hesitation in endorsing Microsoft Excel as a superior product. He remembers having to use Lotus 1-2-3 in DOS when he first started work in the City . . .

FIGURE 1.1 Creating a yield curve in Excel.

So this gives us the *positively sloping* yield curve we see in Figure 1.1. However, if that is the case, why does the curve not continue to slope upwards, all the way to the 30-year mark? The rate of upward movement declines after the five-year mark, and then actually decreases to the 30-year point. This is a peculiarity of many markets – the 30-year bond, commonly called the *long bond*, is usually in such great demand among institutional investors, such as pension funds, that this demand outstrips supply. As a result, the price of this bond is forced upwards, and this moves the yield down to below what it should be.

Constructing a yield curve in the wholesale markets is a little more involved than what we have described above, and a ever-so-slightly complicated branch of mathematics is employed to derive the models used to fit yield curves. We consider this in later chapters.

Using the Yield Curve

The yield curve tells us where the bond market is trading now. It also implies the level of trading for the future, or at least what the market thinks will be happening in the future. In other words, it is a good indicator of the future level of the market. It is also a much more reliable indicator than any other used by private investors, and we can prove this empirically.

Let us consider the main uses of the yield curve. All participants in the debt capital markets have an interest in the current shape and level of the yield curve, as well as what this information implies for the future. The main uses are introduced below:

- **Setting the yield for all debt market instruments.** The yield curve essentially fixes the cost of money over the maturity term structure.

The yields of government bonds from the shortest-maturity instrument to the longest set the benchmark for yields for all other debt instruments in the market, around which all debt instruments are analysed. Issuers of debt (and their underwriting banks) therefore use the yield curve to price bonds and all other debt instruments. Generally the zero-coupon yield curve is used to price new issue securities, rather than the redemption yield curve.

- **Acting as an indicator of future yield levels.** As we discuss later in this chapter, the yield curve assumes certain shapes in response to market expectations of the future interest rates. Bond market participants analyse the present shape of the yield curve in an effort to determine the implications regarding the future direction of market interest rates. This is perhaps one of the most important functions of the yield curve, and it is as much an art as a science. The yield curve is scrutinised for its information content, not just by bond traders and fund managers, but also by corporate financiers as part of project appraisal. Central banks and government treasury departments also analyse the yield curve for its information content, not just regarding forward interest rates, but also with regard to expected inflation levels.
- **Measuring and comparing returns across the maturity spectrum.** Portfolio managers use the yield curve to assess the relative value of investments across the maturity spectrum. The yield curve indicates the returns that are available at different maturity points and is therefore very important to fixed income fund managers, who can use it to assess which point of the curve offers the best return relative to other points.
- **Indicating relative value between different bonds of similar maturity.** The yield curve can be analysed to indicate which bonds are cheap or dear to the curve. Placing bonds relative to the zero-coupon yield curve helps to highlight which bonds should be bought or sold either outright or as part of a bond spread trade.
- **Pricing interest rate derivative securities.** The price of derivative securities revolves around the yield curve. At the short end, products such as Forward Rate Agreements are priced off the futures curve, but futures rates reflect the market's view on forward three-month cash deposit rates. At the longer end, interest rate swaps are priced off the yield curve, while hybrid instruments that incorporate an option feature such as convertibles and callable bonds also reflect current yield curve levels. The "risk-free" interest rate, which is one of the parameters used in option pricing, is the T-bill rate or short-term government repo rate, both constituents of the money market yield curve.

YIELD TO MATURITY YIELD CURVE

The most commonly occurring yield curve is the yield to maturity yield curve. The equation used to calculate the yield to maturity is given in the Appendix. The curve itself is constructed by plotting the yield to maturity against the term to maturity for a group of bonds of the same class. Three different examples are shown at Figure 1.2. Bonds used in constructing the curve will only rarely have an exact number of whole years to redemption, however, it is often common to see yields plotted against whole years on the x-axis. This is because once a bond is designated the *benchmark* for that term, its yield is taken to be the representative yield. For example, the then 10-year benchmark bond in the UK gilt market, the 5.75% Treasury 2009, maintained its benchmark status throughout 1999 and into 2000, even as its term to maturity fell below 10 years. The yield to maturity yield curve is the most commonly observed curve simply because yield to maturity is the most frequent measure of return used. The business sections of daily newspapers, where they quote bond yields at all, usually quote bond yields to maturity.

As we might expect, given the source data from which it is constructed, the yield to maturity yield curve contains some inaccuracies. We have already come across the main weakness of the yield to maturity measure, which is the assumption of a constant rate for coupon reinvestment during the bond's life at the redemption yield level. Since market rates fluctuate over time, it is not possible to achieve this (a feature known as *reinvestment risk*). Only zero-coupon bondholders avoid reinvestment risk as no coupon is paid during the life of a zero-coupon bond.

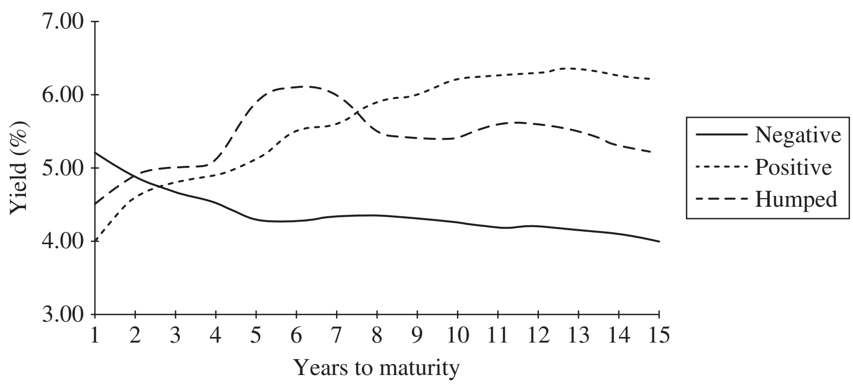

FIGURE 1.2 Yield to maturity yield curves.

Users of the Bloomberg system can call up a large number of yield curves, for different sectors of the market and different instruments. The main screen is IYC, which presents a menu of government bond redemption yield curves. For example, Figure 1.3 shows screen IYC for the UK gilt curve, as at 18 June 2003. The curves are for benchmark bonds and for gilt zero-coupon bonds or strips. Figure 1.4 compares the gilt curve and the sterling swap curve. Figure 1.5 shows the corresponding curves for French and German government bonds, on the same screen, and Figure 1.6 shows the benchmark curve for US Treasuries.

Figure 1.7 shows the screen FMC, which presents a menu of different government and corporate bond yield curves. Users select the curve required from this menu. Selecting curves 664, 670, and 673 (see Figure 1.7) gives us composite yield curves for AAA-, A- and BBB-rated corporate bonds. These curves are shown at Figure 1.8.

Finally, Figure 1.9 shows screen SWCV, the interest rate swap curve. We show the screen as selected for sterling interest rate swaps.

The yield to maturity yield curve does not distinguish between different payment patterns that may result from bonds with different coupons, that is, the fact that low-coupon bonds pay a higher portion of their cash

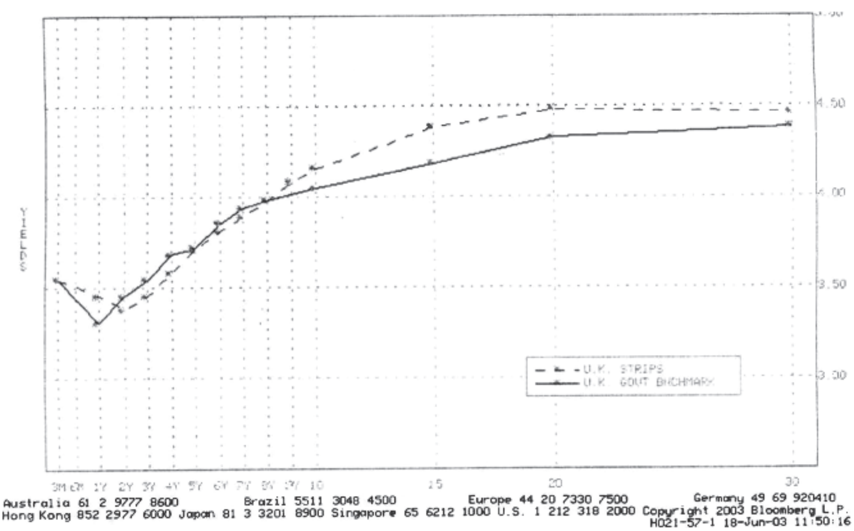

FIGURE 1.3 UK gilt yield curves as at 18 June 2003.
Source: © Bloomberg L.P. Reproduced with permission.

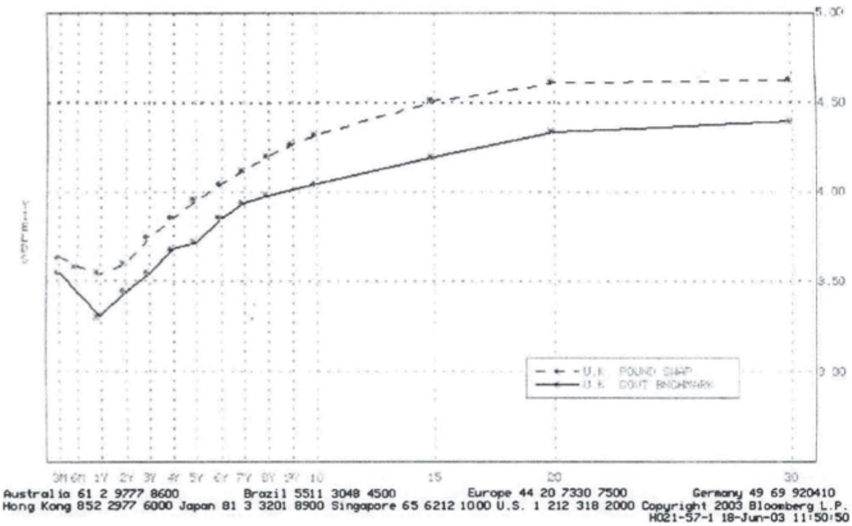

FIGURE 1.4 UK gilt yield curve and sterling interest rate swap curve as at 18 June 2003.
Source: © Bloomberg L.P. Reproduced with permission.

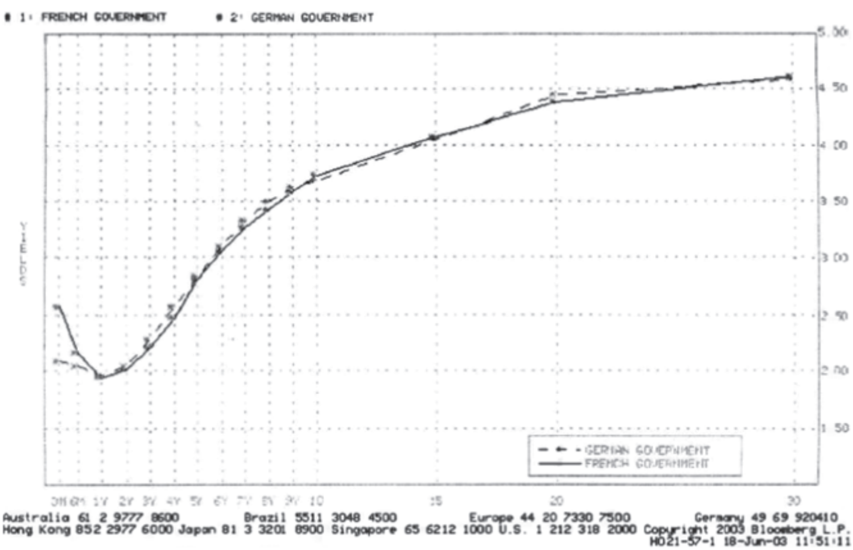

FIGURE 1.5 French and German government yield curves, 18 June 2003.
Source: © Bloomberg L.P. Reproduced with permission.

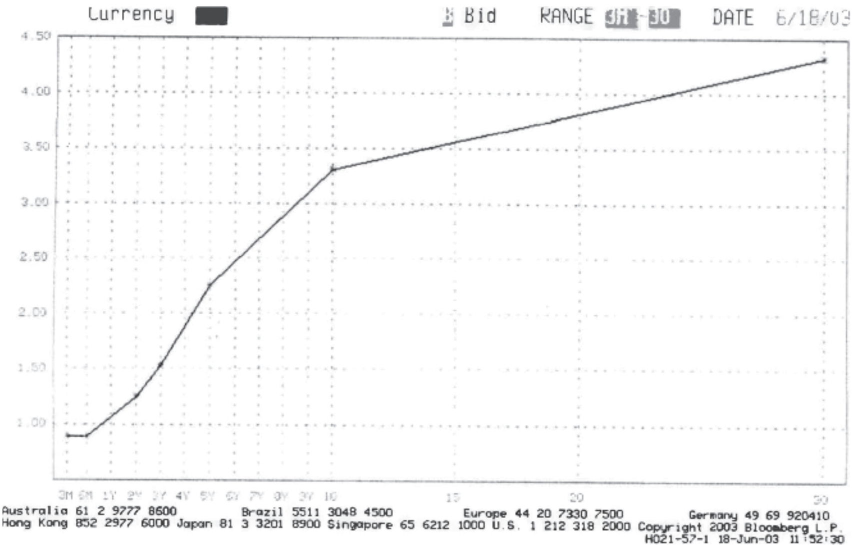

FIGURE 1.6 Treasury yield curve as at 18 June 2003.
Source: © Bloomberg L.P. Reproduced with permission.

FIGURE 1.7 Screen FMC, menu of corporate and government bond yield curves.
Source: © Bloomberg L.P. Reproduced with permission.

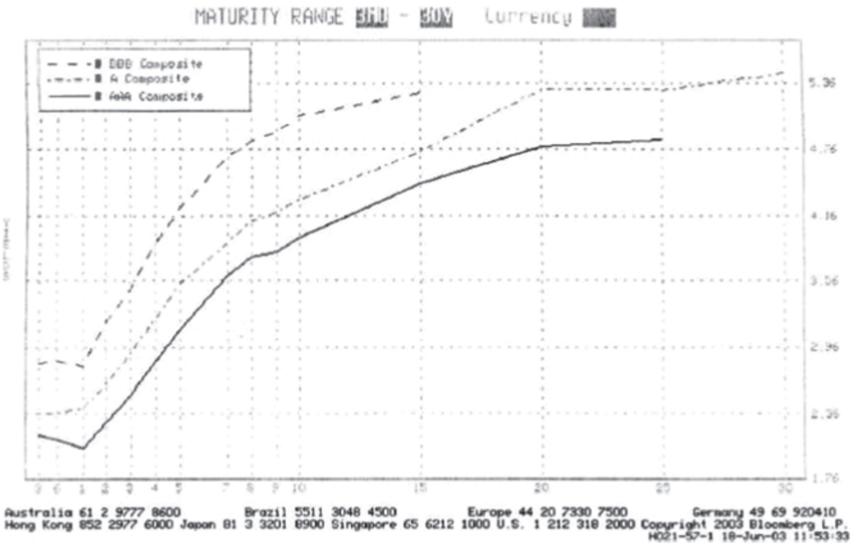

FIGURE 1.8 Screen FMC, AAA-, A-, and BBB-rated corporate yield curves as at 18 June 2003.
Source: © Bloomberg L.P. Reproduced with permission.

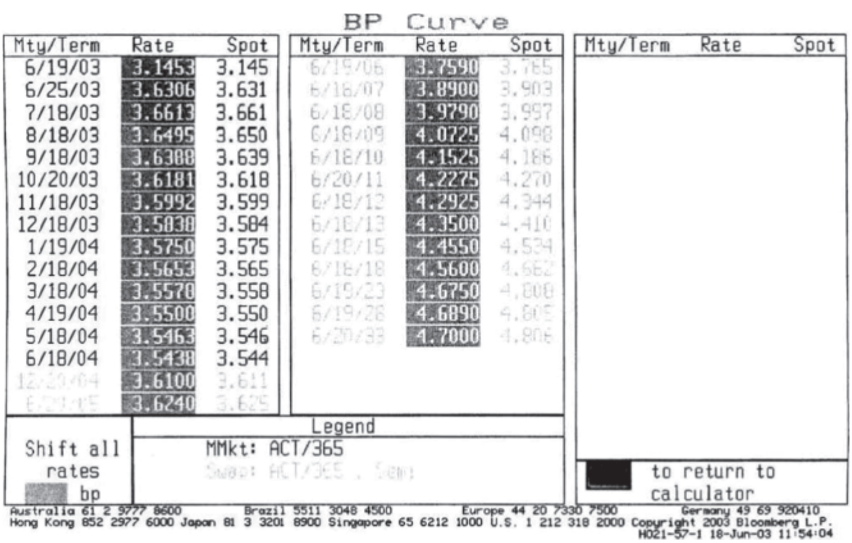

FIGURE 1.9 Screen SWCV, selected for sterling swap curve as at 18 June 2003.
Source: © Bloomberg L.P. Reproduced with permission.

flows at a later date than high-coupon bonds of the same maturity. The curve also assumes an even cash flow pattern for all bonds. Therefore in this case, cash flows are not discounted at the appropriate rate for the bonds in the group being used to construct the curve. To compensate for this, bond analysts may sometimes construct a *coupon yield curve*, which plots yield to maturity against term to maturity for a group of bonds with the same coupon. This may be useful when a group of bonds contains some with very high coupons – high-coupon bonds often trade "cheap to the curve", that is they have higher yields, than corresponding bonds of the same maturity but lower coupon. This is usually because of reinvestment risk and, in some markets (including the UK), for tax reasons.

Now is probably an appropriate time to remind readers that the YTM, and hence the YTM curve, is really just an "assumed" yield. It isn't a true term structure of interest rates. This is because, for the investor to receive the YTM during the period the bond is held, the bond must have been (i) purchased at par, and (ii) all the coupons received during the holding period must have been reinvested at the same YTM. Clearly, these two sets of conditions are rarely, if ever, met during the holding period. So whilst YTM is a useful measure for initial purchase and comparison purposes, one should remember that it is rarely, if ever, the return that is actually received.

THE COUPON YIELD CURVE

The *coupon yield curve* is a plot of the yield to maturity against term to maturity for a group of bonds with the same coupon. If we construct such a curve, we can see that in general high-coupon bonds trade at a discount (have higher yields) relative to low-coupon bonds, because of reinvestment risk and for tax reasons. In the UK, for example, on gilts the coupon is taxed as income tax, while any capital gain is exempt from capital gains tax. Even in jurisdictions where capital gain on bonds is taxable, this can often be deferred whereas income tax cannot. It is frequently the case that yields vary considerably with coupon for the same term to maturity, and with term to maturity for different coupons. Put another way, usually we observe different coupon curves, not only at different levels, but also with different shapes. Distortions arise in the yield to maturity curve if no allowance is made for coupon differences. For this reason, bond analysts frequently draw a line of "best fit" through a plot of redemption yields, because the coupon effect in a group of bonds will produce a curve with humps and troughs. Figure 1.10 shows a hypothetical set of coupon yield curves, however, since in any group of bonds it is unusual to observe bonds with the same coupon along the entire term structure, this type of curve is relatively rare.

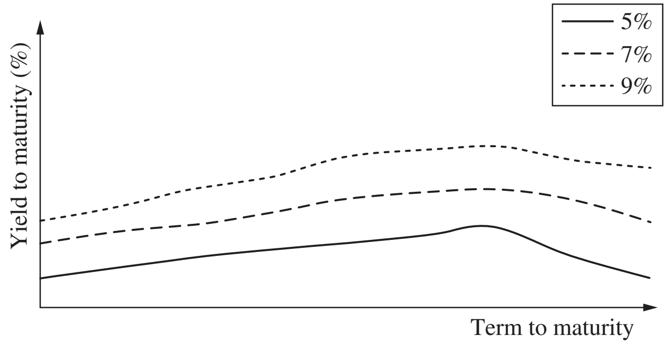

FIGURE 1.10 Coupon yield curves.

THE PAR YIELD CURVE

The *par yield curve* is not usually encountered in secondary market trading, however, it is often constructed for use by corporate financiers and others in the new issues or *primary* market. The par yield curve plots yield to maturity against term to maturity for current bonds trading at par.[4] The par yield is therefore equal to the coupon rate for bonds priced at par or near to par, as the yield to maturity for bonds priced exactly at par is equal to the coupon rate. Those involved in the primary market will use a par yield curve to determine the required coupon for a new bond that is to be issued at par. This is because investors prefer not to pay over par for a new-issue bond, so the bond requires a coupon that will result in a price at or slightly below par.

The par yield curve can be derived directly from bond yields when bonds are trading at or near par. If bonds in the market are trading substantially away from par, then the resulting curve will be distorted. It is then necessary to derive it by iteration from the spot yield curve. As we would observe at almost any time, it is rare to encounter bonds trading at par for any particular maturity. The market therefore uses actual non-par vanilla bond yield curves to derive *zero-coupon yield curves* and then constructs hypothetical par yields that would be observed were there any par bonds being traded.

[4] Par price for a bond is almost invariably 100%. Certain bonds have par defined as 1,000 per 1,000 nominal of paper.

THE ZERO-COUPON (OR SPOT) YIELD CURVE

The *zero-coupon* (or *spot*) *yield curve* plots zero-coupon yields (or spot yields) against term to maturity. A zero-coupon yield is the yield prevailing on a bond that has no coupons. In the first instance, if there is a liquid zero-coupon bond market we can plot the yields from these bonds if we wish to construct this curve. However, it is not necessary to have a set of zero-coupon bonds in order to construct the curve, as we can derive it from a coupon or par yield curve. In fact, in many markets where no zero-coupon bonds are traded, a spot yield curve is derived from the conventional yield to maturity yield curve. This is of course a *theoretical zero-coupon* (spot) yield curve, as opposed to the *market* or *observed* spot curve that can be constructed using the yields of actual zero-coupon bonds trading in the market.[5]

Interest Rates: Basic Concepts

Spot yields must comply with Equation (1.1). This equation assumes annual coupon payments and that the calculation is carried out on a coupon date so that accrued interest is zero:

$$\begin{aligned} P_d &= \sum_{n=1}^{N} \frac{C}{(1+rs_n)^n} + \frac{M}{(1+rs_N)^N} \\ &= \sum_{n=1}^{N} C \times df_n + M \times df_N \end{aligned} \tag{1.1}$$

where

rs_n is the spot or zero-coupon yield on a bond with n years to maturity

$df_n = 1/(1+rs_n)^n$ = the corresponding *discount factor*.

In (1.1) rs_1 is the current one-year spot yield, rs_2 the current two-year spot yield, and so on. Theoretically, the spot yield for a particular term to

[5] It is common to see the terms "spot rate" and "zero-coupon rate" used synonymously. However, the spot rate is a theoretical construct and cannot be observed in the market. The definition of the spot rate, which is the rate of return on a single cash flow that has been dealt today and is received at some point in the future, comes very close to that of the yield on a zero-coupon bond, which can be observed directly in the market. Zero-coupon rates can therefore be taken to be spot rates in practice, which is why the terms are frequently used interchangeably.

maturity is the same as the yield on a zero-coupon bond of the same maturity, which is why spot yields are also known as zero-coupon yields. This last is an important result, as spot yields can be derived from redemption yields that have been observed in the market.

As with the yield to redemption yield curve, the spot yield curve is commonly used in the market. It is viewed as the true term structure of interest rates because there is no reinvestment risk involved – the stated yield is equal to the actual annual return. That is, the yield on a zero-coupon bond of n years maturity is regarded as the true n-year interest rate. Because the observed government bond redemption yield curve is not considered to be the true interest rate, analysts often construct a theoretical spot yield curve. Essentially this is done by breaking down each coupon bond being observed into its constituent cash flows, which become a series of individual zero-coupon bonds. For example, £100 nominal of a 5% two-year bond (paying annual coupons) is considered equivalent to £5 nominal of a one-year zero-coupon bond and £105 nominal of a two-year zero-coupon bond.

Let us assume that in the market there are 30 bonds all paying annual coupons. The first bond has a maturity of one year, the second bond of two years, and so on out to thirty years. We know the price of each of these bonds, and we wish to determine what the prices imply about the market's estimate of future interest rates. We expect interest rates to vary over time, but that all payments being made on the same date are valued using the same rate. For the one-year bond, we know its current price and the amount of the payment (comprised of one coupon payment and the redemption proceeds) we will receive at the end of the year, therefore we can calculate the interest rate for the first year – assume the one-year bond has a coupon of 5%. If the bond is priced at par and we invest £100 today, we will receive £105 in one year's time, hence the rate of interest is apparent and is 5%. For the two-year bond, we use this interest rate to calculate the future value of its current price in one year's time: *this is how much we would receive if we had invested the same amount in the one-year bond*. However, the two-year bond pays a coupon at the end of the first year and if we subtract this amount from the future value of the current price, the net amount is what we should be giving up in one year in return for the one remaining payment. From these numbers we can calculate the interest rate in year two.

Assume that the two-year bond pays a coupon of 6% and is priced at £99.00. If the £99.00 is invested at the rate we calculated for the one-year bond (5%), it will accumulate £103.95 in one year, made up of the £99.00 investment and interest of £4.95. On the payment date in one year's time, the one-year bond matures and the two-year bond pays a coupon of 6%. If everyone expected that at this time the two-year bond would be priced at more than 97.95 (which is 103.95 minus 6.00), then no investor would buy the one-year bond, since it would be more advantageous to buy the two-year bond and sell it after one year for a greater return. Similarly, if the price is less than 97.95, no investor would buy the two-year bond, as it would be cheaper to buy the shorter bond and then buy the longer-dated bond with the proceeds received when the one-year bond matures. Therefore the two-year bond must be priced at exactly 97.95 in 12 months' time. For this £97.95 to grow to £106.00 (the maturity proceeds from the two-year bond, comprising the redemption payment and coupon interest), the interest rate in year two must be 8.22%. We can check this using the present value formula covered earlier. At these two interest rates, the two bonds are said to be in equilibrium.

This is an important result and shows that (in theory) there can be no arbitrage opportunity along the yield curve – using interest rates available today, the return from buying the two-year bond must equal the return from buying the one-year bond and rolling over the proceeds (or *reinvesting)* for another year. This is known as the *breakeven principle,* a law of no-arbitrage.

Using the price and coupon of the three-year bond, we can calculate the interest rate in year three in exactly the same way. Using each of the bonds in turn, we can link together the *implied one-year rates* for each year up to the maturity of the longest-dated bond. The process is known as *boot-strapping*. The "average" of the rates over a given period is the spot yield for that term. In the example given above, the rate in year one is 5%, and in year two is 8.20%. An investment of £100 at these rates would grow to £113.61. This gives a total percentage increase of 13.61% over two years, or 6.588% per annum. (The average rate is not obtained by simply dividing 13.61 by 2, but – using our present value relationship again – by calculating the square root of "1 plus the interest rate" and then subtracting 1 from this number). Thus the one-year yield is 5% and the two-year yield is 8.20%.

In real-world markets it is not necessarily as straightforward as this, for instance, on some dates there may be several bonds maturing, with different coupons, and on some dates there may be no bonds maturing. It is most unlikely that there will be a regular spacing of bond redemptions exactly one year apart. For this reason, it is common for analysts to use a software model to calculate the set of implied spot rates which best fits the market prices of the bonds that do exist in the market. For instance, if there are several one-year bonds, each of their prices may imply a slightly different rate of interest. We choose the rate that gives the smallest average price error. In practice, all bonds are used to find the rate in year one, all bonds with a term longer than one year are used to calculate the rate in year two, and so on. The zero-coupon curve can also be calculated directly from the coupon yield curve using a method similar to that described above. In this case, the bonds would be priced at par and their coupons set to the par yield values.

The zero-coupon yield curve is ideal to use when deriving implied forward rates, which we consider next, and defining the term structure of interest rates. It is also the best curve to use when determining the *relative value*, whether cheap or dear, of bonds trading in the market, and when pricing new issues, irrespective of their coupons. However, it is not an absolutely accurate indicator of average market yields because most bonds are not zero-coupon bonds.

Zero-Coupon Discount Factors

Having introduced the concept of the zero-coupon curve in the previous paragraph, we can illustrate more formally the mathematics involved. When deriving spot yields from redemption yields, we view conventional bonds as being made up of an annuity, which is the stream of fixed coupon payments, and a zero-coupon bond, which is the redemption payment on maturity. To derive the rates we can use (1.1), setting $P_d = M = 100$ and $C = 100 \cdot rm_N$ as shown in (1.2) below. This has the coupon bonds trading at par, so that the coupon is equal to the yield:

$$\begin{aligned} 100 &= 100 \times rm_N \times \sum_{n=1}^{N} df_n + 100 \times df_N \\ &= 100 \times rm_N \times A_N + 100 \times df_N \end{aligned} \tag{1.2}$$

where rm_N is the par yield for a term to maturity of N years, where the discount factor df_N is the fair price of a zero-coupon bond with a par value of £1 and a term to maturity of N years, and where:

$$A_N = \sum_{n=1}^{N} df_n = A_{N-1} + df_N \tag{1.3}$$

is the fair price of an annuity of £1 per year for N years (with $A_0 = 0$ by convention). Substituting (1.3) into (1.2) and rearranging will give us the expression below for the N-year discount factor, shown at (1.4):

$$df_N = \frac{1 - rm_N \times A_{N-1}}{1 + rm_N} \tag{1.4}$$

If we assume one-year, two-year, and three-year redemption yields for bonds priced at par to be 5%, 5.25%, and 5.75% respectively, we will obtain the following solutions for the discount factors:

$$df_1 = \frac{1}{1+0.05} = 0.95238$$
$$df_2 = \frac{1-(0.0525)(0.95238)}{1+0.0525} = 0.90261$$
$$df_3 = \frac{1-(0.0575)(0.95238+0.90261)}{1+0.0575} = 0.84476$$

We can confirm that these are the correct discount factors by substituting them back into equation (1.2). This gives us the following results for the one-year, two-year, and three-year par value bonds (with coupons of 5%, 5.25%, and 5.75% respectively):

$$100 = 105 \times 0.95238$$
$$100 = 5.25 \times 0.95238 + 105.25 \times 0.90261$$
$$100 = 5.75 \times 0.95238 + 5.75 \times 0.90261 + 105.75 \times 0.84476.$$

Now that we have found the correct discount factors it is relatively straightforward to calculate the spot yields using equation (1.1), and this is shown below:

$$df_1 = \frac{1}{(1+rs_1)} = 0.95238 \text{ which gives } rs_1 = 5.0\%$$
$$df_2 = \frac{1}{(1+rs_2)^2} = 0.90261 \text{ which gives } rs_2 = 5.269\%$$
$$df_3 = \frac{1}{(1+rs_3)^3} = 0.84476 \text{ which gives } rs_3 = 5.778\%$$

Equation (1.1) discounts the n-year cash flow (comprising the coupon payment and/or principal repayment) by the corresponding n-year spot yield. In other words, rs_n is the *time-weighted rate of return* on a n-year bond. Thus, as we said in the previous section, the spot yield curve is the correct method for pricing or valuing any cash flow, including an irregular cash flow, because it uses the appropriate discount factors. That is, it matches each cash flow to the discount rate that applies to the time period in which the cash flow is paid. Compare this to the approach for the yield to maturity procedure discussed earlier, which discounts all cash flows by the same yield to maturity. This illustrates neatly why the N-period zero-coupon interest rate is the true interest rate for an N-year bond.

The expressions above are solved algebraically in the conventional manner, although those wishing to use a spreadsheet application such as Microsoft Excel® can input the *constituents of each* equation into individual cells and solve using the "Tools" and "Goal Seek" functions.

EXAMPLE 1.1 ZERO-COUPON YIELDS

- Consider the following zero-coupon market rates:

One-year (1y)	5.000%
2y	5.271%
3y	5.598%
4y	6.675%
5y	7.213%

Calculate the zero-coupon discount factors and the prices and yields of:

a. a 6% two-year bond, and;
b. a 7% five-year bond.

Assume both are annual coupon bonds.
The zero-coupon discount factors are:

1y:	$1/1.05 = 0.95238095$
2y:	$1/(1.05271)^2 = 0.90236554$
3y:	$1/(1.05598)^3 = 0.84924485$
4y:	$1/(1.06675)^4 = 0.77223484$
5y:	$1/(1.07213)^5 = 0.70593182$

The price of the 6% two-year bond is then calculated in the normal fashion using present values of the cash flows:

$$(6 \times 0.95238095) + (106 \times 0.90236554) = 101.365.$$

The yield to maturity is 5.263%, obtained using the iterative method, with a spreadsheet function such as Microsoft Excel® "Goal Seek" or a Hewlett Packard (HP) calculator.

The price of the 7% five-year bond is:

$$\begin{aligned}&(7 \times 0.95238095) + (7 \times 0.90236554)\\&+(7 \times 0.84924485) + (7 \times 0.77223484)\\&+(107 \times 0.70593182) = 99.8690\end{aligned}$$

The yield to maturity is 7.032%.

FORMULA SUMMARY

Example 1.1 illustrates that if the zero-coupon discount factor for n years is df_n and the par yield for N years is rp, then the expression at (1.5) is always true:

$$\begin{aligned}&(rp \times df_1) + (rp \times df_2) + \cdots + (rp \times df_N) + (1 \times df_N) = 1 \\ &\Rightarrow rp \times (df_1 + df_2 + \cdots + df_N) = 1 - df_N \\ &\Rightarrow rp = \frac{1 - df_N}{\sum_{n=1}^{N} df_n}\end{aligned} \qquad (1.5)$$

USING SPOT RATES IN BOND ANALYSIS

The convention in the markets is to quote the yield on a non-government bond as a certain *spread* over the yield on the equivalent maturity government bond, usually using the gross redemption yields. Traders and investment managers will assess the relative merits of holding the nongovernment bond based on the risk associated with the bond's issuer and the magnitude of its yield spread. For example, in the UK at the beginning of 2002, companies such as National Grid, Severn Trent Water, Abbey National plc, and Tesco plc issued sterling-denominated bonds, all of which paid a certain spread over the equivalent gilt bond.[6] Figure 1.11 shows the average yield spreads of corporate bonds over gilts in the UK market through 2001–2002.

Traditionally, investors will compare the redemption yield of the bond they are analysing with the redemption yield of the equivalent government bond. Just as with the redemption yield measure, however, there is a flaw with this measure, in that the spread quoted is not really comparing like-for-like, as the yields do not reflect the true term structure given by the spot rate curve. There is an additional flaw if the cash flow streams of the two bonds do not match, which in practice they will do only rarely.

Therefore the correct method for assessing the yield spread of a corporate bond is to replicate its cash flows with those of a government bond, which can be done in theory by matching the cash flows of the corporate bond with a package of government zero-coupon bonds of the same nominal value. If no zero-coupon bond market exists, the cash flows can be matched synthetically by valuing a coupon bond's cash flows on a zero-coupon basis.

[6] The spread is, of course, not fixed and fluctuates with market conditions and supply and demand.

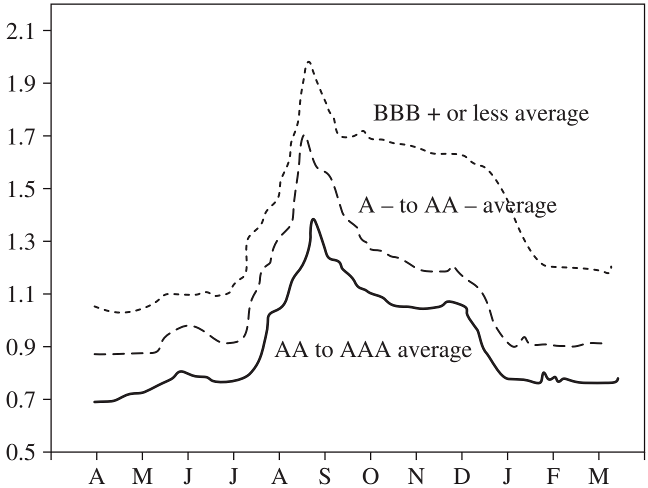

FIGURE 1.11 Average yield spreads of UK corporate bonds versus gilts, 2001–2002.
Source: Bank of England. Reproduced with permission.

The corporate bond's price is of course the sum of the present value of all its cash flows, which should be valued at the spot rates in place for each cash flow's maturity. It is the yield spread of each individual cash flow over the equivalent maturity government spot rate that is then taken to be the true yield spread.

This measure is known in US markets as the *zero-volatility spread* or *static spread,* and it is a measure of the spread that is realised over the government spot rate yield curve if the corporate bond is held to maturity. It is therefore a different measure to the traditional spread, as it is not taken over one point on the (redemption yield) curve but over the whole term to maturity. The zero-volatility spread is that spread which equates the present value of the corporate bond's cash flows to its price, where the discount rates are each relevant government spot rate. The spread is found through an iterative process, and it is a more realistic yield spread measure than the traditional one.

THE FORWARD YIELD CURVE

Forward Yields

Most transaction in the market are for immediate delivery, which is known as the *cash* market, although some markets also use the expression *spot* market, which is more common in foreign exchange. Cash market transactions are settled straight away, with the purchaser of a bond being entitled to interest from the settlement date onwards.[7] There is a large market in *forward* transactions, which are trades carried out today for a forward settlement date. For financial transactions that are forward transactions, the parties to the trade agree today to exchange a security for cash at a future date, but at a price agreed today. So the *forward rate* applicable to a bond is the spot bond yield as at the forward date. That is, it is the yield of a zero-coupon bond that is purchased for settlement at the forward date. It is derived today, using data from a present day yield curve, so it is not correct to consider forward rates to be a prediction of the spot rates as at the forward date.

Forward rates can be derived from spot interest rates. Such rates are then known as *implied* forward rates, since they are implied by the current range of spot interest rates. The *forward* (or *forward-forward*) *yield curve* is a plot of forward rates against term to maturity. Forward rates satisfy expression (1.6):

$$
\begin{aligned}
P_d &= \frac{C}{(1+{}_0rf_1)} + \frac{C}{(1+{}_0rf_1)(1+{}_1rf_2)} + \cdots + \frac{M}{(1+{}_0rf_1)\ldots(1+{}_{N-1}rf_N)} \\
&= \sum_{n=1}^{N} \frac{C}{\prod_{i=1}^{n}(1+{}_{i-1}rf_i)} + \frac{M}{\prod_{i=1}^{N}(1+{}_{i-1}rf_i)}
\end{aligned}
\tag{1.6}
$$

where ${}_{n-1}rf_n$ is the implicit forward rate (or forward-forward rate) on a one-year bond maturing in year *n*.

As a forward or forward-forward yield is implied from spot rates, the forward rate is a forward zero-coupon rate. Comparing (1.1) and (1.6), we see that the spot yield is the *geometric mean* of the forward rates, as shown below:

$$
(1+rs_n)^n = (1+{}_0rf_1)(1+{}_1rf_2)\ldots(1+{}_{n-1}rf_n). \tag{1.7}
$$

This implies the following relationship between spot and forward rates:

$$
\begin{aligned}
(1+{}_{n-1}rf_n) &= \frac{(1+rs_n)^n}{(1+rs_{n-1})^{n-1}} \\
&= \frac{df_{n-1}}{df_n}.
\end{aligned}
\tag{1.8}
$$

[7] We refer to "immediate" settlement, although of course there is a delay between trade date and settlement date, which can be anything from one day to seven days, or even longer in some markets. The most common settlement period is known as "spot" and is two business days.

Using the spot yields we calculated previously, we can derive the implied forward rates from (1.8). For example, the two-year and three-year forward rates are given by:

$$(1+{}_1rf_2)=\frac{(1+0.05269)^2}{(1+0.05)}=5.539\%$$

$$(1+{}_2rf_3)=\frac{(1+0.05778)^3}{(1+0.05269)^2}=6.803\%.$$

Using our expression gives us ${}_0rf_1$, equal to 5%, ${}_1rf_2$ equal to 5.539%, and ${}_2rf_3$ as 6.803%. This means, for example, that given current spot yields, which we calculated from the one-year, two-year, and three-year bond redemption yields (which were priced at par), the market is expecting the yield on a bond with one year to mature in three years' time to be 6.803% (that is, the three-year one-period forward-forward rate is 6.803%).

The relationship between the par yields, spot yields, and forward rates is shown in Table 1.2.

Figure 1.12 highlights our results for all three yield curves graphically. This illustrates another important property of the relationship between the three curves, in that as the original coupon yield curve is positively sloping, so the spot and forward yield curves lie above it. The reasons behind this will be considered later in the chapter.

Let us now consider the following example. Suppose that a two-year bond with cash flows of £5.25 at the end of year 1 and £105.25 at the end of year 2 is trading at par, hence it has a redemption yield (indeed a par yield) of 5.25% (this is the bond in Table 1.2). As we showed in the section on zero-coupon yields and the idea of the break-even principle, in order to be regarded as equivalent to this, a pure zero-coupon bond or discount bond making a lump sum payment at the end of year 2 only (so with

TABLE 1.2 Coupon, spot, and forward yields

Year	Coupon yield (%)	Zero-coupon yield (%)	Forward rate (%)
1	5.000	5.000	5.000
2	5.250	5.269	5.539
3	5.750	5.778	6.803

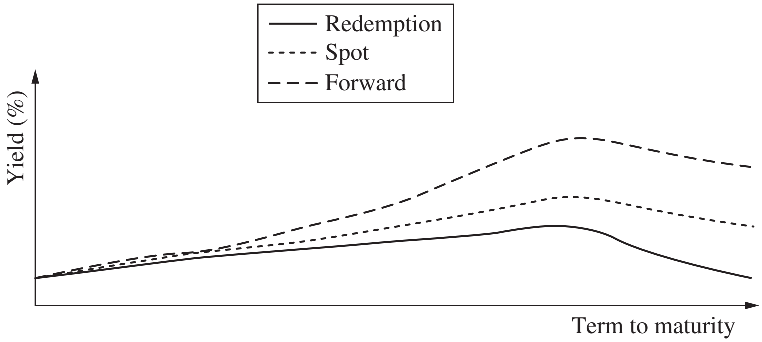

FIGURE 1.12 Redemption, spot, and forward yield curves: traditional analysis.

no cash flow at the end of year 1), requires a rate of return of 5.269%, which is the spot yield. That is, for the same investment of £100, the maturity value would have to be £110.82 (this figure is obtained by multiplying 100 by $(1+0.05269)^2$).

This illustrates why the zero-coupon curve is important to corporate financiers involved in new bond issues. If we know the spot yields, then we can calculate the coupon required on a new three-year bond that is going to be issued at par in this interest rate environment by making the following calculation:

$$100 = \frac{C}{(1.05)} + \frac{C}{(1.05269)^2} + \frac{C+100}{(1.05778)^3}.$$

This is solved in the conventional algebraic manner to give C equal to 5.75%.

The relationship between spot yields and forward rates was shown at (1.8). We can illustrate it as follows. If the spot yield is the *average return*, then the forward rate can be interpreted as the *marginal return*. If the marginal return between years 2 and 3 increases from 5.539% to 6.803%, then the average return increases from 5.269% up to the three-year spot yield of 5.778% as shown below:

$$\left((1.05269)^2 (1.06803)\right)^{1/3} - 1 = 0.05777868$$

or 5.778%, as shown in Table 1.2.

FORMULA SUMMARY

- The forward zero-coupon rate from interest period a to period b is given by (1.9):

$$_a rf_b = \left(\frac{(1 + rs_b)^b}{(1 + rs_a)^a} \right)^{1/(b-a)} - 1 \tag{1.9}$$

where rs_a and rs_b are the a and b period spot rates respectively.

- The forward rate from interest period a to period $(a+1)$ is given by (1.10):

$$_a rf_{a+1} = \left(\frac{(1 + rs_{a+1})^b}{(1 + rs_a)^a} \right)^{a+1} - 1 \tag{1.10}$$

Calculating Spot Rates From Forward Rates

The previous section showed the relationship between spot and forward rates. Just as we have derived forward rates from spot rates based on this mathematical relationship, it is possible to reverse this and calculate spot rates from forward rates. If we are presented with a forward yield curve, plotted from a set of one-period forward rates, we can use this to construct a spot yield curve. Equation (1.7) states the relationship between spot and forward rates, rearranged as (1.11) to solve for the spot rate:

$$rs_n = \left((1 + {_0rf_1}) \times (1 + {_1rf_2}) \times (1 + {_2rf_3}) \times \cdots \times (1 + {_nrf_{n+1}}) \right)^{1/n} - 1 \tag{1.11}$$

where ${_1rf_1}$, ${_2rf_1}$, ${_3rf_1}$ are the one-period versus two-period, two-period versus three-period forward rates up to the $(n-1)$ period versus n-period forward rates.

Remember to adjust (1.11) as necessary if dealing with forward rates relating to a deposit of a different interest period. If we are dealing with the current six-month spot rate and implied six-month forward rates, the relationship between these and the n-period spot rate is given by (1.11) in the same way as if we were dealing with the current one-year spot rate and implied one-year forward rates.

EXAMPLE

Example 1.2(i)

- The one-year cash market yield is 5.00%. Market expectations have priced one-year rates in one year's time at 5.95% and in two years' time at 7.25%. What is the current three-year spot rate that would produce these forward rate views?

  To calculate this we assume an investment strategy dealing today at forward rates, and calculate the return generated from this strategy. The return after a three-year period is given by the future value relationship, which in this case is $1.05 \times 1.0595 \times 1.0725 = 1.1931$.

  The three-year spot rate is then obtained by:

$$\left(\frac{1.1931}{1}\right)^{1/3} - 1 = 6.062\%$$

Example 1.2(ii)

- Consider the following six-month implied forward rates, when the six-month spot rate is 4.0000%:

${}_0rf_1$	4.0000%
${}_2rf_3$	4.4516%
${}_3rf_4$	5.1532%
${}_4rf_5$	5.6586%
${}_5rf_6$	6.0947%
${}_6rf_7$	7.1129%

An investor is debating deciding between purchasing a three-year zero-coupon bond at a price of £72.79481 per £100 nominal or buying a six-month zero-coupon bond and then rolling over their investment every six months for the three-year term. If the investor is able to reinvest the proceeds every six months at the actual forward rates

(*Continued*)

in place today, what would the proceeds be at the end of the three-year term?

An investment of £72.79481 at the spot rate of 4% and then reinvested at the forward rates in our table over the next three years would yield a terminal value of:

$$72.79481 \times (1.04)(1.044516)(1.051532)(1.056586)$$
$$\times (1.060947)(1.071129) = 100.$$

This reflects our spot and forward rates relationship, in that if all the forward rates are realised, our investor's £72.79 will produce a terminal value that matches the investment in a three-year zero-coupon bond priced at the three-year spot rate. This illustrates the relationship between the three-year spot rate, the six-month spot rate, and the implied six-month forward rates. So what is the three-year zero-coupon bond trading at? Using (1.11), the solution to this is given by:

$$rs_6 = ((1.04)(1.044516)(1.051532)(1.056586)$$
$$\times (1.060947)(0.071129))^{1/6} - 1 = 5.4346\%$$

which solves our three-year spot rate rs_6 as 5.4346%. Similarly, we could have solved for rs_6 using the conventional price/yield formula for zero-coupon bonds, however, the calculation above illustrates the relationship between spot and forward rates.

ANALYSING AND INTERPRETING THE YIELD CURVE

From observing yield curves in different markets at any time, we notice that a yield curve can adopt one of four basic shapes, which are:

- *normal* or *conventional*: in which yields are at "average" levels and the curve slopes gently upwards as maturity increases;
- *upward-sloping* or *positive* or *rising*: in which yields are at historically low levels, with long rates substantially greater than short rates;
- *downward-sloping* or *inverted* or *negative*: in which yield levels are very high by historical standards, but long-term yields are significantly lower than short rates;
- *humped*: where yields are high with the curve rising to a peak in the medium-term maturity area, and then sloping downwards at longer maturities.

Sometimes yield curves will incorporate a mixture of the above features.

A great deal of effort is expended by bond analysts and economists analysing and interpreting yield curves. There is often a considerable information content associated with any curve at any time. For example, Figure 1.13 shows the UK gilt redemption yield curve at three different times in the ten years from June 1989 to June 1999. What does the shape of each curve tell us about the UK debt market, and the UK economy at each particular time?

In this section, we consider the various explanations that have been put forward to explain the shape of the yield curve at any one time. None of the theories can adequately explain everything about yield curves and the shapes they assume at any time, so generally observers seek to explain specific curves using a combination of the accepted theories. This subject is a large one, indeed we could devote several books to it, therefore at this stage we will introduce the main ideas, reserving a more detailed investigation for later in the book.

The existence of a yield curve itself indicates that there is a cost associated with funds of different maturities, otherwise we would observe a flat yield curve. The fact that we very rarely observe anything approaching a flat yield curve suggests that investors require different rates of return depending on the maturity of the instrument they are holding. In this section, we review the main theories that have been put forward to explain the shape of the yield curve, which all have fairly long-dated antecedents.

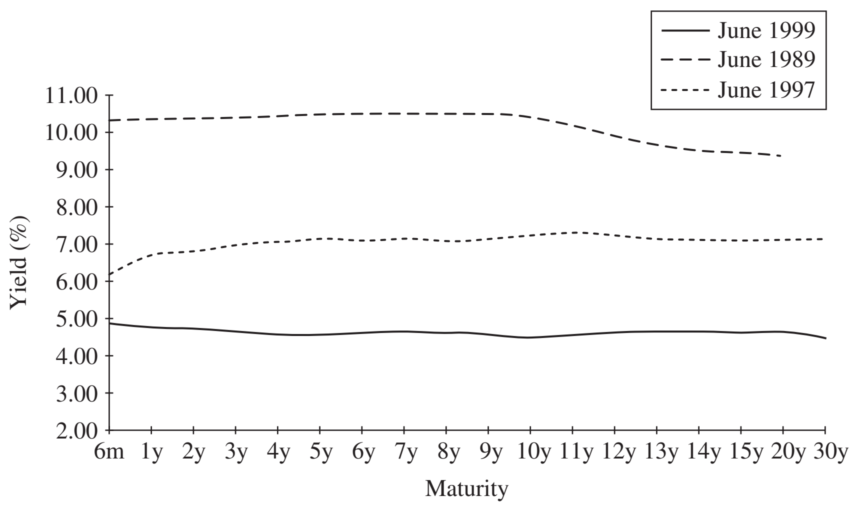

FIGURE 1.13 UK gilt redemption yield curves.
Source: Bloomberg L.P. Reproduced with permission.

An excellent account of the term structure is given in *Theory of Financial Decision Making* by Jonathan Ingersol (1987), Chapter 18. In fact it is worth purchasing this book just for Chapter 18 alone. Another quality account of the term structure is by Shiller (1990). In the following section we provide an introductory review of the research on this subject to date.

The Expectations Hypothesis

The expectations hypothesis suggests that bondholders' expectations determine the course of future interest rates. There are two main competing versions of this hypothesis, the *local expectations hypothesis* and the *unbiased expectations hypothesis*. The *return to maturity expectations* hypothesis and *yield to maturity expectations* hypothesis are also quoted (see Ingersoll 1987). The local expectations hypothesis states that all bonds of the same class, but differing in term to maturity, have the same expected holding period rate of return. This suggests that a six-month bond and a twenty-year bond will produce the same rate of return, on average, over the stated holding period. So if we intend to hold a bond for six months, we will receive the same return no matter what specific bond we buy. The author feels that this theory is not always the case nor relevant, despite being mathematically neat, however, it is worth spending a few moments discussing it and related points. Generally, holding period returns from longer-dated bonds are on average higher than those from short-dated bonds. Intuitively we would expect this, with longer-dated bonds offering higher returns to compensate for their higher price volatility (risk). The local expectations hypothesis would not agree with the conventional belief that investors, being risk averse, require higher returns as a reward for taking on higher risk. In addition, it does not provide any insight about the shape of the yield curve. An article by Cox, Ingersoll and Ross (1981) showed that the local expectations hypothesis best reflected equilibrium between spot and forward yields. This was demonstrated using a feature known as Jensen's inequality, which is described in Appendix 1.2 of the first edition. Robert Jarrow (1996) states:

> *". . . in an economic equilibrium, the returns on . . . similar maturity zero-coupon bonds cannot be too different. If they were too different, no investor would hold the bond with the smaller return. This difference could not persist in an economic equilibrium".*
>
> (Jarrow 1996, p. 50)

This reflects economic logic, but in practice other factors can impact on holding period returns between bonds that do not have similar maturities.

For instance, investors have restrictions as to which bonds they can hold, for example, banks and building societies are required to hold short-dated bonds for liquidity purposes. In an environment of economic disequilibrium, these investors still have to hold shorter-dated bonds, even if the holding period return is lower.

So although it is economically "neat" to expect that the return on a long-dated bond is equivalent to rolling over a series of shorter-dated bonds, it is often observed that longer-term (default-free) returns exceed annualised short-term default-free returns. Therefore an investor that continually rolled over a series of short-dated zero-coupon bonds would most likely receive a lower return than if they had invested in a long-dated zero-coupon bond. Rubinstein (1999) gives an excellent, accessible explanation of why this is so. The reason is that compared to the theoretical model, in reality future spot rates are not known with certainty. This means that short-dated zero-coupon bonds are more attractive to investors for two reasons; first, they are more appropriate instruments to use for hedging purposes, and secondly they are more liquid instruments, in that they may be more readily converted back into cash than long-dated instruments. With regard to hedging, consider an exposure to rising interest rates. If the yield curve shifts upwards at some point in the future, the price of long-dated bonds will fall by a greater amount. This is a negative result for holders of such bonds, whereas the investor in short-dated bonds will benefit from rolling over their funds at the (new) higher rates. With regard to the second issue, Rubinstein (1999) states:

> *". . . it can be shown that in an economy with risk-averse individuals, uncertainty concerning the timing of aggregate consumption, the partial irreversibility of real investments (longer-term physical investments cannot be converted into investments with earlier payouts without sacrifice), [and] . . . real assets with shorter-term payouts will tend to have a 'liquidity' advantage."*
>
> (Rubinstein 1999, pp. 84–85)

Therefore the demand for short-term instruments is frequently higher, and hence short-term returns are often lower than long-term returns.

The *pure* or *unbiased expectations hypothesis* is more commonly encountered and states that current implied forward rates are unbiased estimators of future spot interest rates.[8] It assumes that investors act in a

[8] For original discussion, see Lutz (1940) and Fisher (1986) (although he formulated his ideas earlier).

way that eliminates any advantage of holding instruments of a particular maturity. Therefore if we have a positively sloping yield curve, the unbiased expectations hypothesis states that the market expects spot interest rates to rise. Equally, an inverted yield curve is an indication that spot rates are expected to fall. If short-term interest rates are expected to rise, then longer yields should be higher than shorter ones to reflect this. If this were not the case, investors would only buy the shorter-dated bonds and roll over the investment when they matured. Likewise, if rates are expected to fall, then longer yields should be lower than short yields. The unbiased expectations hypothesis states that the long-term interest rate is a geometric average of expected future short-term rates. This was in fact the theory that was used to derive the forward yield curve using (1.5) and (1.7) previously. This gives us:

$$(1 + rs_N)^N = (1 + rs_1)(1 + {}_1rf_2)\ldots(1 + {}_{N-1}rf_N) \tag{1.12}$$

or:

$$(1 + rs_N)^N = (1 + rs_{N-1})^{N-1}(1 + {}_{N-1}rf_N) \tag{1.13}$$

where rs_N is the spot yield on a N-year bond and ${}_{n-1}rf_n$ is the implied one-year rate n years ahead. For example, if the current one-year spot rate is $rs_1 = 5.0\%$ and the market is expecting the one-year rate in a year's time to be ${}_1rf_2 = 5.539\%$, then the market is expecting a £100 investment in two one-year bonds to yield:

$$£100(1.05)(1.05539) = £110.82$$

after two years. To be equivalent to this, an investment in a two-year bond has to yield the same amount, implying that the current two-year rate is $rs_2 = 5.7\%$, as shown below:

$$£100(1 + rs_2)^2 = £110.82$$

which gives us rs_2 equal to 5.27%, and gives us the correct future value as shown below:

$$£100(1.0527)^2 = £110.82$$

This result must be so, to ensure no arbitrage opportunities exist in the market and in fact we showed as much earlier in the chapter when we considered forward rates. According to the unbiased expectations hypothesis, therefore, the forward rate ${}_0rf_2$ is an unbiased predictor of the spot rate ${}_1rs_1$ observed one period later. On average the forward rate should equal the subsequent spot rate. The hypothesis can be used to explain any shape in the yield curve.

A rising yield curve is therefore explained by investors expecting short-term interest rates to rise, that *is* $_1rf_2 > rs_2$. A falling yield curve is explained by investors expecting short-term rates to be lower in the future. A humped yield curve is explained by investors expecting short-term interest rates to rise and long-term rates to fall. Expectations, or views on the future direction of the market, are a function mainly of the expected rate of inflation. If the market expects inflationary pressures in the future, the yield curve will be positively shaped, while if inflation expectations are inclined towards disinflation, then the yield curve will be negative. However, several empirical studies including one by Fama (1976) have shown that forward rates are essentially biased predictors of future spot interest rates, and often overestimate future levels of spot rates. The unbiased hypothesis has also been criticised for suggesting that investors can forecast (or have a view on) very long-dated spot interest rates, which might be considered slightly unrealistic. As yield curves in most developed country markets exist to a maturity of up to 30 years or longer, such criticisms may have some substance. Are investors able to forecast interest rates 10, 20, or 30 years into the future? Perhaps not, nevertheless this is indeed the information content of say, a 30-year bond. Since the yield on the bond is set by the market, it is valid to suggest that the market has a view on inflation and future interest rates for up to 30 years forward.

The expectations hypothesis is stated in more than one way – we have already encountered the local expectations hypothesis. Other versions include the *return to maturity* expectations hypothesis, which states that the total return from holding a zero-coupon bond to maturity is equal to the total return that is generated by holding a short-term instrument and continuously rolling it over the same maturity period. A related version, the *yield to maturity* hypothesis, states that the periodic return from holding a zero-coupon bond is equal to the return from rolling over a series of coupon bonds, but refers to the annualised return earned each year rather than the total return earned over the life of the bond. This assumption enables a zero-coupon yield curve to be derived from the redemption yields of coupon bonds. The unbiased expectations hypothesis states that forward rates are equal to the spot rates expected by the market in the future. The Cox–Ingersoll–Ross article suggests that only the local expectations hypothesis describes a model that is purely arbitrage-free, as under the other scenarios it is possible to employ certain investment strategies that would produce returns in excess of what was implied by today's yields. Although it has been suggested[9] that the differences between the local and the unbiased hypotheses are not material, a model that describes such a scenario does not reflect investors' beliefs, which is why further research is ongoing in this area.

[9] For example, see Campbell (1986) and Livingstone (1990).

The unbiased expectations hypothesis does not by itself explain all the shapes of the yield curve or the information content contained within it, so it is often tied in with other explanations, including the liquidity preference theory.

Mathematical Description of Expectations Hypothesis

Simply put, the *expectations hypothesis* states that the slope of the yield curve reflects the market's expectations about future interest rates. There are in fact four main versions of the hypothesis, each distinct from the other and each not compatible with the others. The expectations hypothesis has a long history, first being described in 1896 by Fisher and later developed by Hicks (1946), among others.[10] As Shiller (1990) describes, the thinking behind it probably stems from the way market participants discuss their views on future interest rates when assessing whether to purchase long-dated or short-dated bonds. For instance, if interest rates are expected to fall, investors will purchase long-dated bonds in order to "lock in" the current high long-dated yield. If all investors act in the same way, the yield on long-dated bonds will decline as prices rise in response to demand, and this yield will remain low as long as short-dated rates are expected to fall, and will revert to a higher level only once the demand for long-term rates is reduced. Therefore, downward-sloping yield curves are an indication that interest rates are expected to fall, while an upward-sloping curve reflects market expectations of a rise in short-term interest rates.

Let us briefly consider the main elements of the discussion. The *unbiased expectations hypothesis* states that current forward rates are unbiased predictors of future spot rates. Let $f_t(T, T+1)$ be the forward rate at time t for the period from T to $T+1$. If the one-period spot rate at time T is r_T, then according to the unbiased expectations hypothesis:

$$f_t(T, T+1) = E_t[r_T] \tag{1.14}$$

which states that the forward rate $f_t(T, T+1)$ is the expected value of the future one-period spot rate given by r_T at time T.

The *return to maturity expectations hypothesis* states that the return generated from an investment of term t to T by holding a $(T-t)$-period

[10] See the footnote on page 644 of Shiller (1990) for a fascinating historical note on the origins of the expectations hypothesis. An excellent overview of the hypothesis itself is contained in Ingersoll (1987, pp. 389–392).

bond is equal to the expected return generated by a holding a series of one-period bonds and continually rolling them over on maturity. More formally, this can be expressed as (1.15):

$$\frac{1}{P(t,T)} = E_t\left[(1+r_t)(1+r_{t+1})\ldots(1+r_{T-1})\right]. \qquad (1.15)$$

The left-hand side of Equation (1.15) represents the return received by an investor holding a zero-coupon bond to maturity, which is equal to the expected return associated with rolling over £1 from time t to time T by continually reinvesting one-period maturity bonds, each of which has a yield of the future spot rate r_t. A good argument for this hypothesis is contained in Jarrow (1996, p. 52), which states that, essentially, in an environment of *economic equilibrium*, the returns on zero-coupon bonds of similar maturity cannot be significantly different, otherwise investors would not hold the bonds with the lower return. A similar argument can be put forward in relation to coupon bonds of differing maturities. Any difference in yield would not therefore disappear as equilibrium was re-established. However, there are a number of reasons why investors will hold shorter-dated bonds, irrespective of the yield available on them, so it is possible for the return to maturity version of the hypothesis not to apply. In essence, this version represents an equilibrium condition in which expected *holding period returns* are equal, although it does not state that this return is the same from different bond-holding strategies.

From (1.14) and (1.15), we can determine that the unbiased expectations hypothesis and the return to maturity hypothesis are not compatible with each other, unless there is no correlation between future interest rates. As Ingersoll (1987) notes, although it would be both possible and interesting to model such an economic environment, it is not related to reality, as interest rates are highly correlated. Given positive correlation between rates over a period of time, bonds with maturity terms longer than two periods will have a higher price under the unbiased expectations hypothesis than under the return to maturity version. Bonds of exactly two-period maturity will have the same price.

The *yield to maturity expectations hypothesis* is described in terms of yields. It is given by:

$$\left[\frac{1}{P(t,T)}\right]^{\frac{1}{T-t}} = E_t\left[\left\{(1+r_t)(1+r_{t+1})\ldots(1+r_{T-1})\right\}\frac{1}{T-t}\right] \qquad (1.16)$$

where the left-hand side of the equation specifies the yield to maturity of the zero-coupon bond at time *t*. In this version, the expected holding period *yield* on continually rolling over a series of one-period bonds is equal to the yield that is guaranteed by holding a long-dated bond until maturity.

The *local expectations hypothesis* states that all bonds will generate the same expected rate of return if held over a small term. It is given by:

$$\frac{E_t\left[P(t+1,T)\right]}{P(t,T)} = 1 + r_t \tag{1.17}$$

This version of the hypothesis is the only one that is consistent with no-arbitrage, because the expected rates of return on all bonds are equal to the risk-free interest rate. For this reason, the local expectations hypothesis is sometimes referred to as *the risk-neutral* expectations hypothesis.

Liquidity Preference Theory

Intuitively we might feel that longer maturity investments are more risky than shorter ones. An investor lending money for a five-year term will usually demand a higher rate of interest than if they were to lend the same customer money for a five-week term. This is because the borrower may not be able to repay the loan over the longer time period as they may for instance, have gone bankrupt in that period. For this reason, longer-dated yields should be higher than short-dated yields, to compensate the lender for the higher risk exposure during the term of the loan.[11]

We can consider this theory in terms of inflation expectations as well. Where inflation is expected to remain roughly stable over time, the market anticipates a positive yield curve. However, the expectations hypothesis cannot by itself explain this phenomenon, as under stable inflationary conditions one would expect a flat yield curve. The risk inherent in longer-dated investments, or the *liquidity preference theory*, seeks to explain a positively shaped curve. Generally, borrowers prefer to borrow over as long a term as possible, while lenders will wish to lend over as short a term as possible. Therefore, as we first stated, lenders have to be compensated for lending over the longer term. This compensation is considered a premium for a loss in *liquidity* for the lender. The premium is increased the further the investor lends across the term structure, so that the longest-dated investments will, all else being equal, have the highest yield. So the liquidity preference theory states that the yield curve should almost always be upward-sloping, reflecting bondholders' preference for the liquidity and lower risk of shorter-dated bonds.

[11] For original discussion, see Hicks (1946).

An inverted yield curve could still be explained by the liquidity preference theory when it is combined with the unbiased expectations hypothesis. A *humped* yield curve might be viewed as a combination of an inverted yield curve together with a positively sloping liquidity preference curve.

The difference between a yield curve explained by unbiased expectations and an actual observed yield curve is sometimes referred to as the *liquidity premium*. This refers to the fact that in some cases short-dated bonds are easier to transact in the market than long-term bonds. It is difficult to quantify the effect of the liquidity premium, which in any cases is not static and fluctuates over time. The liquidity premium is so-called because, in order to encourage investors to hold longer-dated securities, the yields on such securities must be higher than those available on short-dated securities, which are more liquid and may be converted into cash more easily. The liquidity premium is the compensation required for holding less liquid instruments. If longer-dated securities then provide higher yields, as is suggested by the existence of the liquidity premium, they should generate on average, higher total returns over an investment period. This is not consistent with the local expectations hypothesis. More formally we can write:

$$0 = L_1 < L_2 < L_3 < \cdots < L_n$$

and

$$(L_2 - L_1) > (L_3 - L_2) > \cdots < (L_n - L_{n-1})$$

where L_n is the premium for a bond with term to maturity of n years, which states that the premium increases as the term to maturity rises and that an otherwise flat yield curve will have a positively sloping curve, with the degree of slope steadily decreasing as we extend along the yield curve. This is consistent with the observation of yield curves under "normal" conditions.

The expectations hypothesis assumes that forward rates are equal to the expected future spot rates, that is as shown in (1.18):

$$_{n-1}rf_n = E\left(_{n-1}rs_n\right) \tag{1.18}$$

where $E(\ )$ is the expectations operator for the current period. This assumption implies that the forward rate is an unbiased predictor of the future spot rate, as we suggested in the previous paragraph. Liquidity preference theory on the other hand, recognises the possibility that the forward rate may contain an element of liquidity premium which declines over time as the period approaches, given by (1.19):

$$_{n-1}rf_n > E\left(_{n-1}rs_n\right) \tag{1.19}$$

If there is uncertainty in the market about the future direction of spot rates and hence where the forward rate should lie, (1.19) is adjusted to give the reverse inequality.

An illustration of how liquidity preference will impact the "clean" term structure suggested by the expectations hypothesis is given at Figure 1.14.

Segmentation Hypothesis

The capital markets are made up of a wide variety of users, each with different requirements. Certain classes of investors will prefer dealing at the short end of the yield curve, while others concentrate on the longer end of the market. The *segmented markets* theory suggests that activity is concentrated in certain specific areas of the market, and that there are no interrelationships between these parts of the market. The relative amounts of funds invested in each of the maturity spectrum causes differentials in supply and demand, which results in humps in the yield curve. That is, the shape of the yield curve is determined by supply and demand for certain specific maturity investments, each of which has no reference to any other part of the curve.

For example, banks and building societies concentrate a large part of their activity at the short end of the curve, as part of daily cash management (known as *asset and liability management*) and for regulatory purposes (known as *liquidity* requirements). Fund managers such as pension funds and insurance companies are active at the long end of the market. Few institutional investors, however, have any preference for medium-dated bonds. This behaviour on the part of investors leads to high prices (low yields) at both the short and long ends of the yield curve and lower prices (higher yields) in the middle of the term structure.

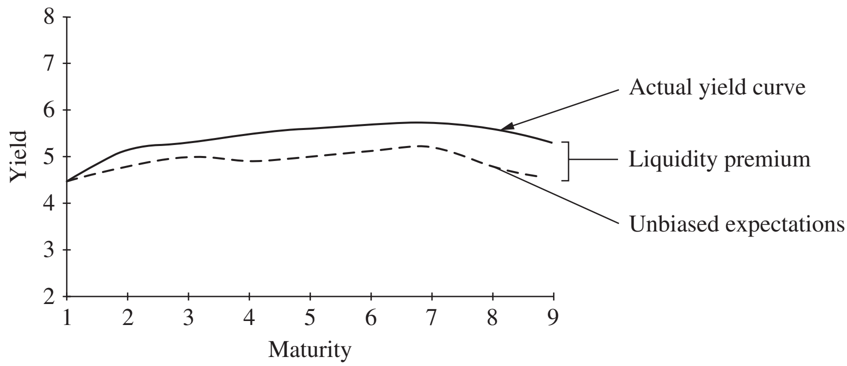

FIGURE 1.14 Yield curve explained by expectations hypothesis and liquidity preference theory.

According to the segmented markets hypothesis, a separate market exists for specific maturities along the term structure, and interest rates for these maturities are set by supply and demand.[12] Where there is no demand for a particular maturity, the yield lies above other segments. Market participants do not hold bonds in any other area of the curve outside their area of interest,[13] so that short-dated and long-dated bond yields exist independently of each other. The segmented markets theory is usually illustrated by reference to banks and life companies. Banks and building societies hold their funds in short-dated instruments, usually no longer than five years in maturity. This is because of the nature of retail banking operations, with a large volume of instant access funds being deposited at banks, and also for regulatory purposes. Holding short-term, liquid bonds enables banks to meet any sudden or unexpected demand for funds from customers. The classic theory suggests that as banks invest their funds in short-dated bonds, the yields on these bonds are driven down. When they then liquidate part of their holding, perhaps to meet higher demand for loans, the yields are driven up and prices of the bonds fall. This affects the short end of the yield curve, but not the long end.

The segmented markets theory can be used to explain any particular shape of the yield curve, although it fits best perhaps with positively sloping curves. However, it cannot be used to interpret the yield curve whatever shape it may be, and therefore offers no information content during analysis. By definition, the theory suggests that for investors, bonds with different maturities are not perfect substitutes for each other. This is because different bonds would have different holding period returns, making them imperfect substitutes of one another.[14] As a result of bonds being imperfect substitutes, markets are segmented according to maturity.

The segmentations hypothesis is a reasonable explanation of certain features of a conventional positively sloping yield curve, but by itself is not sufficient. There is no doubt that banks and building societies have a requirement to hold securities at the short end of the yield curve, as much for regulatory purposes as for yield considerations, however, other investors are probably more flexible and will place funds where value is deemed to exist. Nonetheless the higher demand for benchmark securities does drive down yields along certain segments of the curve.

---

[12] See Culbertson (1957).

[13] For example, retail and commercial banks hold bonds in the short dates, while life assurance companies hold long-dated bonds.

[14] *Ibid.*

A slightly modified version of the market segmentation hypothesis is known as the *preferred habitat theory*. This suggests that different market participants have an interest in specified areas of the yield curve, but can be encouraged to hold bonds from other parts of the maturity spectrum if there is sufficient incentive. Hence banks may at certain times hold longer-dated bonds once the price of these bonds falls to a certain level, making the return on the bonds worth the risk involved in holding them. Similar considerations may persuade long-term investors to hold short-dated debt. So higher yields will be required to make bond holders shift out of their usual area of interest. This theory essentially recognises the flexibility that investors have, outside regulatory or legal requirements (such as the terms of an institutional fund's objectives), to invest in whatever part of the yield curve they identify value.

Humped Yield Curves

When plotting a yield curve of all the bonds in a certain class, it is common to observe humped yield curves. These usually occur for a variety of reasons. In line with the unbiased expectations hypothesis, humped curves are observed when interest rates are expected to rise over the next several periods and then decline. On other occasions, humped curves can result from skewed expectations of future interest rates. This is when the market believes that fairly constant future interest rates are likely, but also believes that there is a small probability for lower rates in the medium term. The other common explanation for humped curves is the preferred habitat theory.

The Combined Theory

The explanation for the shape of the yield curve at any time is more likely to be described by a combination of the pure expectations hypothesis and the liquidity preference theory, and possibly one or two other theories. Market analysts often combine the unbiased expectations hypothesis with the liquidity preference theory into an "eclectic" theory. The result is consistent with any shape of yield curve, and is also a predictor of rising interest rates. In the combined theory, the forward interest rate is equal to the expected future spot rate, together with a quantified liquidity premium. This is shown at (1.20):

$$ {}_0rf_i = E\left({}_{i-1}rs_1 \right) + L_i \tag{1.20} $$

where L_i is the liquidity premium for a term to maturity of i. The size of the liquidity premium is expected to increase with increasing maturity.[15] An illustration is given at Example 1.3.

[15] So that $L_i > L_{i-1}$.

EXAMPLE 1.3 POSITIVE YIELD CURVE WITH CONSTANT EXPECTED FUTURE INTEREST RATES

- Consider the interest rates structure in Table 1.3.

TABLE 1.3 Positive yield curve with constant expected future interest rates

Period n	0	1	2	3	4	5
E(rs)		4.50%	4.50%	4.50%	4.50%	4.50%
Forward rate ${}_0rf_n$		5.00%	5.50%	6.00%	6.50%	7.50%
Spot rate rs_n	5.00%	5.30%	5.80%	6.20%	6.80%	7.00%

The current term structure is positively sloping since the spot rates increase with increasing maturity. However, the market expects future spot rates to be constant at 4.5%. The forward and spot rates are also shown, however, the forward rate is a function of the expected spot rate and the liquidity premium. This premium is equal to 0.50% for the first year, 1.0% in the second, and so on.

The combined theory is consistent with an inverted yield curve. This applies even when the liquidity premium is increasing with maturity, for example, where the expected future spot interest rate is declining. Typically this would be where there was a current curve of falling yields along the term structure. The spot rates might be declining where the fall in the expected future spot rate exceeds the corresponding increase in the liquidity premium.

The Flat Yield Curve

The conventional theories do not seek to explain a flat yield curve. Although it is rare to observe flat yield curves in a market, certainly for any length of time, at times they do emerge in response to peculiar economic circumstances. In the conventional thinking, a flat curve is not tenable because investors should in theory have no incentive to hold long-dated bonds over shorter-dated bonds when there is no yield premium, so that as they sell off long-dated paper, the yield at the long end should increase, producing an upward-sloping curve. In previous circumstances of a flat curve, analysts have produced different explanations for their existence. In November 1988, the US Treasury yield curve was flat relative to the recent past. Researchers contended that this was the result of the market's view that long-dated yields would fall as bond prices rallied upwards.[16] One recommendation is to buy longer maturities when the yield curve is flat, in anticipation of lower long-term interest rates, which is the direct opposite to the view that a flat curve is a signal to sell long bonds. In the case of the US market in 1988, long bond yields did in fact fall by approximately 2% in the following 12 months. This indicates that our view of future long-term rates should be behind the decision to buy or sell long bonds, rather than the shape of the yield curve itself. A flat curve may well be more heavily influenced by supply and demand factors than anything else, with the majority opinion eventually winning out and forcing a change in the curve to a more conventional shape.

Further Views on the Yield Curve

In this and the previous chapter, our discussion of present values, spot, and forward interest rates assumed an economist's world of the *perfect market* (also sometimes called the *frictionless* financial market). Such a perfect capital market is characterised by:

- perfect information;
- no taxes;
- bullet maturity bonds;
- no transaction costs.

Of course in practice markets are not completely perfect. However, assuming perfect markets makes the discussion of spot and forward rates and the term structure easier to handle. When we analyse yield curves for their information content, we have to remember that the markets that they represent are not perfect, and that frequently we observe anomalies that are not explained by the conventional theories.

At any one time it is probably more realistic to suggest that a range of factors contributes to the yield curve being one particular shape. For instance, short-term interest rates are greatly influenced by the availability of funds in the money market. The slope of the yield curve (usually defined as the ten-year yield minus the three-month interest rate) is also a measure of the degree of tightness of government monetary policy. A low, upward-sloping curve is often thought to be a sign that an environment of cheap money, due to a more loose monetary policy, is to be followed by a period of higher inflation and higher bond yields. Equally, a high downward-sloping curve is taken to mean that a situation of tight credit, due to more strict monetary policy, will result in falling inflation and lower bond yields. Inverted

[16] See Levy (1999).

yield curves have often preceded recessions; for instance, The *Economist* in an article from April 1998 remarked that in the US every recession since 1955, except one, had been preceded by a negative yield curve. The analysis is the same – if investors expect a recession, they also expect inflation to fall, so the yields on long-term bonds will fall relative to short-term bonds. So therefore the conventional explanation for an inverted yield curve is that the markets and the investment community expect either a slow-down of the economy, or an outright recession.[17] In this case, one would expect the monetary authorities to ease the money supply by reducing the base interest rate in the near future – hence an inverted curve. At the same time, a reduction of short-term interest rates will affect short-dated bonds and these are sold off by investors, further raising their yield.

While the conventional explanation for negative yield curves is an expectation of economic slow-down, on occasion other factors are involved. In the UK in the period July 1997–June 1999, the gilt yield curve was inverted.[18] There was no general view that the economy was heading for recession, however. In fact the new Labour government led by Tony Blair inherited an economy believed to be in good health. Instead, the explanation behind the inverted shape of the gilt yield curve focused on two other factors: firstly, the handing of responsibility for setting interest rates to the Monetary Policy Committee (MPC) of the Bank of England, and secondly the expectation that the UK would, over the medium term, abandon sterling and join the euro currency. The yield curve at this time suggested that the market expected the MPC to be successful and keep inflation at a level around 2.5% over the long term (its target is actually a 1% range either side of 2.5%), and also that sterling interest rates would need to come down over the medium term as part of *convergence* with interest rates in euroland. These are both medium-term expectations, however, and in the author's view not logical at the short end of the yield curve. In fact, the term structure moved to a positively sloped shape up to the six- to seven-year area, before inverting out to the long end of the curve, in June 1999. This is a more logical shape for the curve to assume, but it was short-lived and returned to being inverted after the two-year term.

There is therefore significant information content in the yield curve, and economists and bond analysts will consider the shape of the curve as part of their policy-making and investment advice. The shape of parts of the

[17] A recession is formally defined as two successive quarters of falling output in the domestic economy.

[18] Although the curve briefly went positively sloped out to seven to eight years in July 1999, it very quickly reverted to being inverted throughout the term structure, and remained so until summer 2001. It subsequently reverted to a conventional positively sloping curve.

curve, whether the short end or long end, as well that of the entire curve, can serve as useful predictors of future market conditions. As part of an analysis, it is also worthwhile considering the yield curves across several different markets and currencies. For instance, the interest rate swap curve, and its position relative to that of the government bond yield curve, is also regularly analysed for its information content. In developed country economies, the swap market is invariably as liquid as the government bond market, if not more liquid, and so it is common to see the swap curve analysed when making predictions about, say, the future level of short-term interest rates.

Government policy influences the shape and level of the yield curve, including policy on public sector borrowing, debt management, and open-market operations. The market's perception of the size of public sector debt influences bond yields. For example, an increase in the level of debt can lead to an increase in bond yields across the maturity range. Open-market operations, which refers to the daily operation by the Bank of England to control the level of the money supply (to which end the Bank purchases short-term bills and also engages in repo dealing), can have a number of effects. In the short-term, it can tilt the yield curve both upwards and downwards. In the longer-term, changes in the level of the base rate will affect yield levels. An anticipated rise in base rates can lead to a decrease in prices for short-term bonds, because short-term bond yields will be expected to rise and this can lead to a temporary inverted curve. Finally, debt management policy influences the yield curve. (In the UK this is the responsibility of the Debt Management Office.) Most government debt is rolled over as it matures, but the maturity of the replacement debt can have a significant influence on the yield curve in the form of humps in the market segment in which the debt is placed, if the debt is priced by the market at a relatively low price and hence high yield.

AN INTRODUCTION TO FITTING THE YIELD CURVE

When graphing a yield curve, we plot a series of discrete points of yield against maturity. Similarly for the term structure of interest rates, we plot spot rates for a fixed time period against that time period. The yield curve itself, however, is a smooth curve drawn through these points. Therefore we require a method that allows us to fit the curve as accurately as possible, known as *yield curve modelling* or *estimating the term structure*. There are several ways to model a yield curve, which we introduce in this section. We will return to the subject in greater depth later in the book.

Ideally, the fitted yield curve should be a continuous function, with no gaps in the curve, while passing through the observed yield vertices. The curve also needs to be "smooth", as "kinks" in the curve will produce sudden sharp jumps in derived forward rates. We have stated how it is possible to calculate a set of discrete discount factors or a continuous discount function. It has been shown that the discount function, par yield curve, spot rate yield curve, and forward rate curve are all related mathematically, such that if one knows any one of these, the other three can be derived. In practice, in many markets it is generally not possible to observe the curves directly, hence they need to be derived from coupon bond prices and yields. In attempting to model a yield curve from bond yields, we need to consider the two fundamental issues introduced above. Firstly, there is the problem of gaps in the maturity spectrum, as in reality there will not be a bond maturing at regular intervals along the complete term structure. For example, in the UK gilt market at the time of writing of the first edition, there was no bond at all maturing between 2017 and 2021, or between 2021 and 2028. Secondly, as we have seen, the term structure is formally defined in terms of spot or zero-coupon interest rates, but in many markets there is no actual zero-coupon bond market. In such cases spot rates cannot be inferred directly but must be implied from coupon bonds. Where zero-coupon bonds are traded, for example, in the US and UK government bond markets, we are able to observe zero-coupon yields directly in the market.

Further problems in fitting the curve arise from these two issues:

- How is the gap in maturities to be tackled? Analysts need to choose between "smoothness" and "responsiveness" of the curve estimate. Most models opt for a smooth fitting, however, enough flexibility should be retained to allow for true movements in the term structure where indicated by the data.
- Should the yield curve be estimated from the discount function or say, the par yield curve?

There are other practical factors to consider as well, such as the effect of withholding tax on coupons, and the size of bond coupons themselves. We will consider the issues connected with estimating the yield curve in a later chapter. At this point, however, we confine ourselves to introducing the main methods.

Interpolation

The simplest method that is employed to fit a curve is *linear interpolation*, which involves drawing a straight line joining each pair of yield vertices. To calculate the yield for one vertex we use (1.21):

$$rm_t = rm_i + \frac{t - n_i}{n_{i+1} - n_i} + \left(rm_{i+1} - rm_i\right) \tag{1.21}$$

where rm_t is the yield being estimated and n is the number of years to maturity for yields that are observed. For example, consider the following redemption yields:

1 month:	4.00%
2 years:	5.00%
4 years:	6.50%
10 years:	6.75%

If we wish to estimate the six-year yield, we calculate it using (1.21), which is:

$$\begin{aligned} rm_{6y} &= 6.50\% + \frac{6-4}{10-6} \times 6.75\% - 6.50\% \\ &= 6.5833\% \end{aligned}$$

The limitations of using linear interpolation are that first, the curve can have sharp angles at the vertices where two straight lines meet, resulting in unreasonable jumps in the derived forward rates. Second, and more fundamentally, being a straight-line method, it assumes the yield between two vertices should automatically be rising (in a positive yield curve environment) or falling. This assumption can lead to gross inaccuracies in the fitted curve.

Another approach is to use *logarithmic interpolation*, which involves applying linear interpolation to the natural logarithms of the corresponding discount factors. Therefore, given any two discount factors we can calculate an intermediate discount factor using (1.22):

$$\ln\left(df_t\right) = \ln\left(df_{n_i}\right) + \frac{t-n}{n_{i-1} - n_i} \times \left(\ln\left(df_{n_{i+1}}\right) - \ln\left(df_{n_i}\right)\right) \tag{1.22}$$

To calculate the six-year yield from the same yield structure above, we use the following procedure:

- calculate the discount factors for years 4 and 10 and then take the natural logarithms of these discount factors;
- perform a linear interpolation on these logarithms;
- take the anti-log of the result, to get the implied interpolated discount factor;
- calculate the implied yield in this discount factor.

Using (1.22) we obtain a six-year yield of 6.6388%. The logarithmic interpolation method reduces the sharpness of angles on the curve and so makes it smoother, but it retains the other drawbacks of the linear interpolation method.

Polynomial Models[19]

The most straightforward method for estimating the yield curve involves fitting a single polynomial in time. For example, a model might use an F-order polynomial, illustrated with (1.23):

$$rm_i = \alpha + \beta_1 N_i + \beta_2 N_i^2 + \cdots + \beta_F N_i^F + u_i \tag{1.23}$$

where

rm_i is the yield to maturity of the i-th bond;

N_i is the term to maturity of the i-th bond;

α, β_1, are coefficients of the polynomial;

u_i is the residual error on the i-th bond.

To determine the coefficients of the polynomial we minimise the sum of the squared residual errors, given by:

$$\sum_i^T u_i^2 \tag{1.24}$$

where T is the number of bonds used. This is represented graphically in Figure 1.15.

The type of curve that results is a function of the order of the polynomial F. If F is too large, the curve will not be smooth but will be in effect too "responsive", such that the curve runs through every point, known as being "over-fitted". The extreme of this is given when $F = T - 1$. If F is too small, the curve will be an over-estimation.

The method described above has been supplanted by a more complex method, which fits different polynomials over different, but overlapping, terms to maturity. The fitted curves are then spliced together to produce a single smooth curve over the entire term structure. This is known as a *spline* curve and is the one most commonly encountered in the markets. For an accessible introduction to spline techniques, see James and Webber (2000) and Choudhry *et al.* (2001).

The limitation of the polynomial method is that a "blip" in the observed series of vertices, for instance, a vertex which is out of line with others in

[19] These are standard econometric techniques. For an excellent account see Campbell *et al.* (1997).

the series, produces a "wobbled" shape, causing wild oscillations in the corresponding forward yields. This can result in the calculation of negative long-dated forward rates.

Cubic Splines

The cubic spline method involves connecting each pair of yield vertices by fitting a unique cubic equation between them. This results in a yield curve where the whole curve is represented by a chain of cubic equations, instead of a single polynomial. This technique adds some "stiffness" to the yield curve, while at the same time preserving its smoothness.[20]

Using the same example as before, we wish to fit the yield curve from 0 to 10 years.

There are four observed vertices, so we require three cubic equations, $rm_{(i,t)}$ each one connecting two adjacent vertices n_i and n_{i+1} as follows:

$rm_{(0,t)} = a_0 n^3 + b_0 n^2 + c_0 n + d_0$, which connects vertex n_0 with n_1,

$rm_{(1,t)} = a_1 n^3 + b_1 n^2 + c_1 n + d_1$, which connects vertex n_1 with n_2, and

$rm_{(2,t)} = a_2 n^3 + b_2 n^2 + c_2 n + d_2$, which connects vertex n_2 with n_3,

where $a, b, c,$ and d are unknowns. The equations each contain four unknowns (the coefficients a to d), and there are three equations so we require a total of twelve conditions to solve the system. The cubic spline method imposes certain conditions on the curves, which makes it possible to solve the system. The solution for this set is summarised in Appendix 1.3 of the first edition.

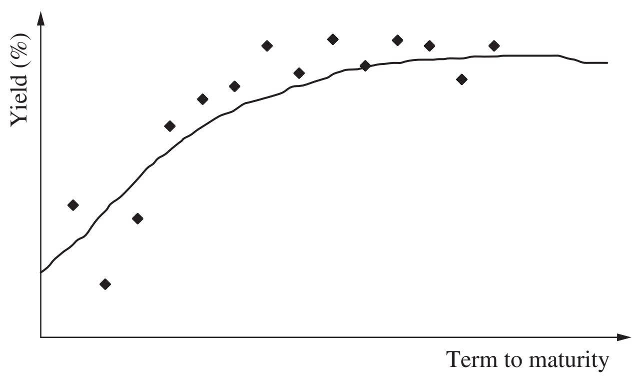

FIGURE 1.15 Polynomial curve-fitting.

[20] In case you're wondering, a spline is a tool used by a carpenter in order to draw smooth curves.

The three cubic equations for the data in this example are:

$rm_{(0,t)} = 0.022 \times n^3 + 0.413 \times n + 4.000$ for vertices $n_0 - n_1$,

$rm_{(1,t)} = -0.047 \times n^3 + 0.411 \times n^2 - 0.410 \times n + 4.548$ for vertices$n_0 - n_2$, and

$rm_{(2,t)} = 0.008 \times n^3 - 0.249 \times n^2 + 2.230 \times n + 1.029$ for vertices $n_2 - n_3$,

Using a cubic spline produces a smoother curve for both the spot rates and the forward rates, while the derived forward curve will have fewer "kinks" in it.

To calculate the estimated yield for the six-year maturity, we apply the third cubic equation, which spans the four to ten year vertices, which is:

$$rm_{(2,t)} = 0.008 \times 6^3 - 0.249 \times 6^2 + 2.230 \times 6 + 1.029 = 7.173\%$$

From Appendix 1.3 of the first edition, it is clear that simply to fit a four-vertex spline requires the inversion of a fairly large matrix. In practice, more efficient mathematical techniques, known as basis splines or *B-splines*, are typically used when there is a larger number of observed yield vertices. This produces results that are very close to what we would obtain by simple matrix inversion.

Regression Models

A variation on polynomial fitting is regression analysis. In this method, bond prices are used as the dependent variable, with the coupon and maturity cash flows of the bonds being the independent variables. This is given by (1.25):

$$P_i = \beta_1 C_{1i} + \beta_2 C_{2i} + \cdots + \beta_N \left(C_{Ni} + M \right) + u_i \qquad (1.25)$$

where

P_d is the dirty price of the i-th bond;

C_{ni} is the coupon of the i-th bond in period n;

β_n is the coefficient of the regression equation;

u_i, is the residual error in the i-th bond.

In fact, the coefficient in (1.25) is an estimate of the discount factor, as shown by (1.26) and can be used to generate the spot interest rate curve.

$$\beta_n = df_n = \frac{1}{\left(1 + rs_n\right)^n}. \qquad (1.26)$$

In the form shown, (1.25) cannot be estimated directly. This is because individual coupon payment dates differ across different bonds, and in a semi-annual coupon market there are more coupons than bonds available.

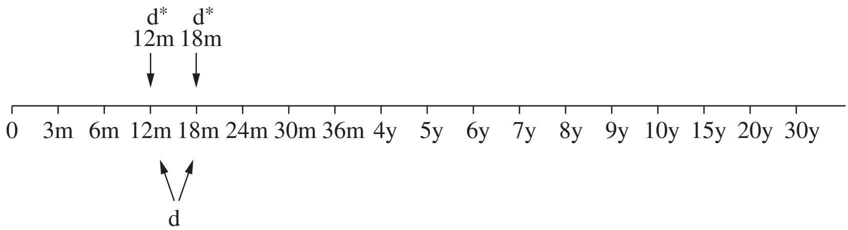

FIGURE 1.16 Grid point allocation in regression analysis.

In practice therefore, the term structure is divided into specific dates, known as *grid points*, along the entire maturity term; and coupon payments are then allocated between two grid points. The allocation between two points is done in such a way so that the present value of the coupon is not altered. This is shown in Figure 1.16.

Note that there are more grid points at the short end of the term structure, with progressively fewer points as we reach the longer end. This is because the preponderance of the data is invariably at the shorter end of the curve, which makes yield curve fitting more difficult. At the long end, however, the shortage of data, due to the relative lack of issues, makes curve estimation more inaccurate.

The actual regression equation that is used in the analysis is given at (1.27) where d^*_{ni} represents the grid points:

$$P_{di} = \beta_1 d^*_{1i} + \beta_2 d^*_{2i} + \cdots + \beta_N \left(d^*_{Ni} + M \right) + u_i \tag{1.27}$$

The two methodologies described above are the most commonly encountered in the market. Models used to estimate the term structure generally fall into two distinct categories, these being the ones that estimate the structure using the par yield curve, and those that fit it using a discount function. We shall examine these in greater detail in a later chapter.

SPOT AND FORWARD RATES IN THE MARKET

Using Spot Rates

The concepts discussed in this chapter are important and form a core part of debt markets analysis. It may appear that the content is largely theoretical, especially since many markets do not trade zero-coupon instruments and so spot rates are therefore not observable in practice, however, the concept of the spot rate is an essential part of bond (and other instruments') pricing. In the first instance, we are already aware that bond redemption yields do not reflect a true interest rate for that maturity, for which we use the spot rate. For relative value purposes, traders and portfolio managers frequently compare a bond's actual market price to its theoretical price, calculated using specific zero-coupon yields for each cash flow, and determine whether the bond is "cheap" or "dear". Even where there is some misalignment between the theoretical price of a bond and the actual price, the decision to buy or sell may be based on judgemental factors, since there is often no zero-coupon instrument against which to effect an arbitrage trade. In a market where no zero-coupon instruments are traded, the spot rates used in the analysis are theoretical and are not represented by actual market prices. Traders therefore often analyse bonds in terms of relative value against each other, and the redemption yield curve, rather than against their theoretical zero-coupon based price.

What considerations apply where a zero-coupon bond market exists alongside a conventional coupon-bond market? In such a case, theoretically arbitrage trading is possible if a bond is priced above or below the price suggested by zero-coupon rates. For example, a bond priced above its theoretical price could be sold, and zero-coupon bonds that equated its cash flow stream could be purchased – the difference in price is the arbitrage profit. Similarly, a bond trading below its theoretical price could be purchased and its coupons "stripped" and sold individually as zero-coupon bonds. The proceeds from the sale of the zero-coupon bonds would then exceed the purchase price of the coupon bond. In practice, often the existence of both markets equalises prices between both markets so that arbitrage is no longer possible, although opportunities will still occasionally present themselves.

Using Forward Rates

Newcomers to the markets frequently experience confusion when first confronted with forward rates. Do they represent the market's expectation of where interest rates will actually be when the forward date arrives? If forward rates are a predictor of future interest rates, exactly how good are they at making this prediction? Empirical evidence[21] suggests that in fact forward rates are not accurate predictors of future interest rates, frequently overstating them by a considerable margin. If this is the case, should we attach any value or importance to forward rates?

The value of forward rates does not lie, however, in its track record as a market predictor, but moreover in its use as a hedging tool. As we illustrate in Example 1.3, the forward rate is calculated on the basis that if we are to price say, a cash deposit with a forward starting date, but we wish to deal today, the return from the deposit will be exactly the same as if we invested for a start date today and rolled over the investment at the forward date. The forward rate allows us to lock in a dealing rate now. Once we have dealt today, it is irrelevant what the actual rate pertaining on the forward date is – we have already dealt. Therefore forward rates are often called *hedge* rates, as they allow us to lock in a dealing rate for a future period, thus removing uncertainty.

---

[21] Including Fama (1976).

The existence of forward prices in the market also allows us to make an investment decision, based on our view compared to the market view. The forward rate implied by say, government bond prices is, in effect, the market's view of future interest rates. If we happen not to agree with this view, we will deal accordingly. We are effectively comparing our view on future interest rates with that of the market, and making our investment decision based on this comparison.

- **Understanding forward rates**
  Spot and forward rates that are calculated from current market rates follow mathematical principles to establish what the market believe the arbitrage-free rates for dealing *today* at rates that are effective at some point in the future. As such forward rates are *not* a type of market prediction of where interest rates will be (or should be!) in the future. As we have already noted, forward rates are not a prediction of future rates. It is important to be aware of this distinction. If we were to plot the forward rate curve for the term structure in three months' time, and then compare it in three months with the actual term structure prevailing at the time, the curves would almost certainly not match. When we calculate forward rates, we use the current term structure. The current term structure incorporates all known information, both economic and political, and reflects the market's views. This is exactly the same as when we say that a company's share price reflects all that is known about the company and all that is expected to happen with regard to the company in the near future, including expected future earnings. The term structure of interest rates reflects everything the market knows about relevant domestic and international factors. It is this information that goes into the forward rates calculation. In three months' time, however, there will be new developments that will alter the market's view and therefore alter the current term structure. These developments and events are (by definition, as we cannot know what lies in the future!) not known at the time we calculated and used the three-month forward rates. This is why rates actually turn out to be different from what the term structure predicted at an earlier date. However, for dealing today we use today's forward rates, which reflect everything we know about the market today. In essence then, forward rates and spot rates are two sides of the same coin: if we wish to deal today for a value date in the future, the forward rate is the rate given by the no-arbitrage concept. It is, in this respect, equivalent to the spot rate: it's just the spot rate for forward date value if we wish to execute the trade today.

THE INTEREST-RATE SWAP CURVE AND THE SOVEREIGN BOND CURVE

We have noted at the outset the importance of interpreting the yield curve so one can understand as fully as possible what it is telling us about current and future expected market conditions. This is not necessarily a straightforward exercise even when all the accepted orthodoxies are in place; namely,

- positive interest rates;
- a positively sloping yield curve;
- domestic sovereign bond yields exhibiting the lowest rates in that currency.

All of the above conventions are commonly not evident in many of today's bond markets. We consider negative rates in a subsequent chapter. For now, consider Figure 1.17 which is the US Treasury and US dollar interest-rate swap curves as at 30 April 2018.

We note that the swap curve trades through the Treasury curve at the long end. Even this is not particularly extreme given the post-2008 environment, when in recent years one has observed the swap curve sitting below the Treasury curve at shorter tenors, including the five- and ten-year points. Given that the local market sovereign bond yield is generally regarded as the "risk-free" rate in that economy (we accept that such a position is not "risk free", this is just an expression for the lowest risk investment in that currency), how does one interpret a curve showing the swap rate lower than the sovereign bond yield?

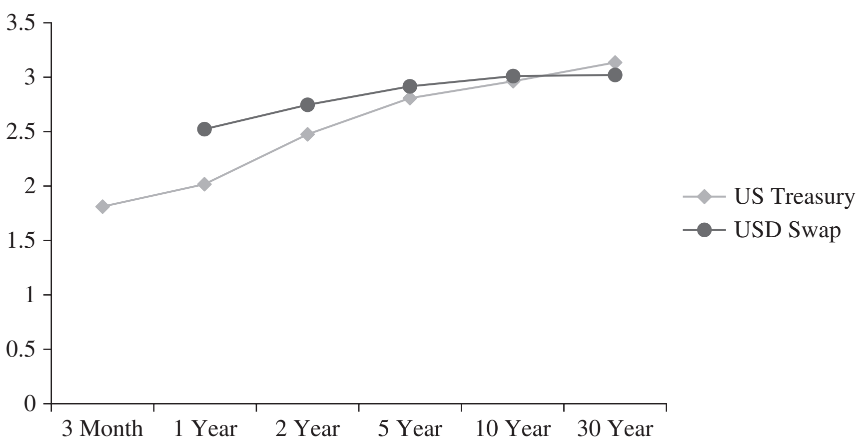

FIGURE 1.17 US Treasury and US dollar interest rate swap curves, 30 April 2018.
Rates source: YieldCurve.com, InterestRateSwapsToday.com.

Compared to the orthodoxy prevailing before 2008, we should note first that interest-rate swap markets, like all interbank derivatives markets now cleared through centralised clearing counterparties (CCPs), are collateralised. The counterparty risk is removed as a result of collateral passed to the CCP by the party that is negative mark-to-market on the swap. (Ignore margin period of risk in this instance.) This is on top of the initial margin (IM) that all CCP parties place with the clearing house or their agent bank. As such, one can view a swap transaction between banks as "risk free" in the same way that one views a US Treasury holding as credit risk free. This would explain the swap curve being very close to the Treasury curve.

But in our example, which we noted is by no means the most extreme manifestation of this phenomenon, the swap rate is *lower* than the sovereign rate. In other words, it is perceived as *lower* risk than the Treasury. What explanation can we offer for this behaviour?

In the first instance, there is supply and demand. Banks and non-bank financial institutions have a sizeable demand to hedge long-dated fixed-rate receivables risk, which means they need to pay fixed in the swap. The excess demand for swaps pushes the rate down at the market-making banks (who drive what the swap rate is). Then there is liquidity: the swap market is very liquid in a way that is not always available in the Treasury market, since the advent of more onerous capital rules for banks who run Trading Books (as opposed to the Banking Book).

Finally, there is the conventional feature of government bonds where the longest-dated yield, in a positively sloping environment, often drops below the prior tenor point (thereby becoming downward sloping). This isn't the case with the curve shown at Figure 1.17, and the explanation for that is the expectation that base interest rates will continue to rise in the short and medium term (as indeed, the Federal Reserve was indicating at that time). Thus this feature combined with the other two factors gives us the two curves we are discussing.

APPENDIX: CUBIC SPLINE INTERPOLATION

There are four observed vertices in the example quoted in the main text, which requires three cubic equations, $m_{(i,t)}$, each one connecting two adjacent vertices n_i and n_{i+1}, as follows:

$$m_{(0,t)} = a_0 n^3 + b_0 n^2 + c_0 n + d_0, \text{ connecting vertex } n_0 \text{ with } n_1,$$

$$m_{(1,t)} = a_1 n^3 + b_1 n^2 + c_1 n + d_1, \text{ connecting vertex } n_1 \text{ with } n_2,$$

$$m_{(2,t)} = a_2 n^3 + b_2 n^2 + c_2 n + d_2, \text{ connecting vertex } n_3 \text{ with } n_4,$$

where a, b, c, and d are unknowns. The three equations require 12 conditions in all. The cubic spline method imposes the following set of conditions on the curves. Each cubic equation must pass through its own pair of vertices. Thus, for the first equation:

$$a_0 n_0^3 + b_0 n_0^2 + c_0 n_0 + d_0 = 4.00$$
$$a_0 n_1^3 + b_0 n_1^2 + c_1 n_1 + d_0 = 5.00$$

For the second and third equations:

$$a_1 n_1^3 + b_1 n_1^2 + c_1 n_1 + d_1 = 5.00$$
$$a_1 n_2^3 + b_1 n_2^2 + c_1 n_2 + d_1 = 6.50$$
$$a_2 n_2^3 + b_2 n_2^2 + c_2 n_2 + d_2 = 6.50$$
$$a_2 n_3^3 + b_2 n_3^2 + c_2 n_3 + d_2 = 6.75$$

The resulting yield curve should be smooth at the point where one cubic equation joins with the next one. This is achieved by requiring the slope and the convexity of adjacent equations to be equal at the point where they meet, ensuring a smooth rollover from one equation to the next. Mathematically, the first and second derivatives of all adjacent equations must be equal at the point where the equations meet.

Thus, at vertex n_1:

$$3a_0 n_1^2 + 2b_0 n_1 + 1 = 3a_1 n_1^2 + 2b_1 n_1 + 1 \quad \text{(the first derivative)}$$
$$6a_0 n_1 + 2 = 6a_1 n_1 + 2 \quad \text{(the second derivative).}$$

And at vertex n_2:

$$3a_1 n_2^2 + 2b_1 n_2 + 1 = 3a_2 n_2^2 + 2b_2 n_2 + 1 \quad \text{(the first derivative)}$$
$$6a_1 n_2 + 2 = 6a_2 n_2 + 2 \quad \text{(the second derivative).}$$

Finally, we may impose the condition that the splines tail off flat at the end vertices, or more formally, we state mathematically that the second derivatives should be zero at the end points:

$$6a_0 n_0 + 2 = 0 \text{(first spline starts flat)}$$
$$6a_2 n_3 + 2 = 0 \text{(second spline ends flat).}$$

These constraints together give us a system of 12 equations from which we can solve for the 12 unknown coefficients. The solution is usually using matrices, where the equations are expressed in matrix form. This is shown at Figure 1.18.

$$\begin{bmatrix} n_0^3 & n_0^2 & n_0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ n_1^3 & n_1^2 & n_1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & n_1^3 & n_1^2 & n_1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & n_2^3 & n_2^2 & n_2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & n_2^3 & n_2^2 & n_2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & n_3^3 & n_3^2 & n_3 & 1 \\ 3n_1^2 & 2n_1 & 1 & 0 & -3n_1^2 & -2n_1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 6n_1 & 2 & 0 & 0 & -6n_1 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3n_2^2 & 2n_2 & 1 & 0 & -3n_2^2 & -2n_1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 6n_2 & 2 & 0 & 0 & -6n_2 & -2 & 0 & 0 \\ 6n_0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 6n_3 & 2 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} a_0 \\ b_0 \\ c_0 \\ d_0 \\ a_1 \\ b_1 \\ c_1 \\ d_1 \\ a_2 \\ b_2 \\ c_2 \\ d_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6.5 \\ 6.75 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

FIGURE 1.18 Cubic spline interpolation matrix.

In matrix notation we have $[n]\times[\text{Coefficients}]=[rm]$ therefore the solution is $[\text{Coefficients}]=[n]^{-1}\times[rm]$. Inverting the matrix *n* and then pre-multiplying *rm* with the resulting inverse, we obtain the array of required coefficients:

$$\text{Coefficients} = \begin{array}{l}[0.022, 0.000, 0.413, 4.000, -0.047, 0.411, -0.410, \\ 4.548, 0.008, -0.249, 2.230, 1.029]\end{array}$$

So the three cubic equations are specified as:

$$rm_{(0,t)} = 0.022\times n^3 + 0.413\times n + 4.000 \text{ for vertices } n_0 - n_1,$$
$$rm_{(1,t)} = -0.047\times n^3 + 0.411\times n^2 - 0.410\times n + 4.548 \text{ for vertices } n_1 - n_2,$$
$$rm_{(2,t)} = 0.008\times n^3 - 0.249\times n^2 + 2.230\times n + 1.029 \text{ for vertices } n_2 - n_3.$$

SELECTED BIBLIOGRAPHY AND REFERENCES

Anderson, N. and Sleath, J., *Bank of England Quarterly Bulletin,* November 1999.
Campbell, J., "A Defence of Traditional Hypotheses about the Term Structure of Interest Rates", *Journal of Finance,* March 1986, pp. 183–193.
Campbell, J., Lo, A. and MacKinlay, A., *The Econometrics of Financial Markets*, Princeton UP, 1997, Chapters 10–11.

Choudhry, M., "The information content of the United Kingdom gilt yield curve", unpublished MBA assignment, Henley Management College, 1998.

Choudhry, M., Joannas, D., Pereira, R. and Pienaar, R., *Capital Markets Instruments,* FT Prentice Hall, 2001.

Cox, J., Ingersoll, J.E. and Ross, S.A., "A re-examination of traditional hypothesis about the term structure of interest rates", *Journal of Finance* 36, September 1981, pp. 769–799.

Culbertson, J.M., "The term structure of interest rates", *Quarterly Journal of Economics* 71, November 1957, pp. 485–517.

Fama, E.F., "Forward rates as predictors of future spot interest rates", *Journal of Financial Economics,* Vol. 3, No. 4, October 1976, pp. 361–377.

Fama, E.F., "The information in the term structure", *Journal of Financial Economics* 13, December 1984, pp. 509–528.

Fisher, I., "Appreciation of Interest", *Publication of the American Economic Association,* August 1986, pp. 23–39.

Heath, D., Jarrow, R. and Morton, A., "Bond pricing and the term structure of interest rates", *Journal of Financial and Quantitative Analysis* 25, 1990, pp. 419–440.

Hicks, J., *Value and Capital*, OUP, 1946.

Ingersoll, J., *Theory of Financial Decision Making*, Rowman & Littlefield, 1987, Chapter 18.

James, J. and Webber, N., *Interest Rate Modelling*, John Wiley and Sons, 2000.

Jarrow, R, "Liquidity premiums and the expectations hypothesis", *Journal of Banking and Finance* 5(4), 1981, pp. 539–546.

Jarrow, R., *Modelling Fixed Income Securities and Interest Rate Options*, McGraw-Hill, 1996

Jarrow, R., and Turnbull, S., *Derivative Securities*, South-Western Publishing, 2000.

Kessel, R.A., "The cyclical behaviour of the term structure of interest rates", *Essays in Applied Price Theory*, University of Chicago, 1965.

Kolb, R., *Futures, Options and Swaps*, Blackwell, 2000.

Levy, H., *Introduction to Investments*, 2nd edition, South-Western College Publishing, 1999.

Livingstone, M., *Money and Capital Markets*, Prentice-Hall, 1990, pp. 254–256.

Lutz, F., "The Structure of Interest Rates", *Quarterly Journal of Economics*, November 1940, pp. 36–63.

Mastronikola, K., "Yield curves for gilt-edged stocks: a new model", *Bank of England Discussion Paper* (Technical Series), No. 49, 1991.

McCulloch, J.H., "An estimate of the liquidity premium", *Journal of Political Economy* 83, Jan–Feb 1975, pp. 95–109.

Meiselman, D., *The Term Structure of Interest Rates*, Prentice Hall, 1962.

Ryan, R. (ed.), *Yield Curve Dynamics*, Glenlake Publishing Company, 1997.

Rubinstein, M., *Rubinstein on Derivatives*, RISK, 1999, pp. 84–85.

Shiller, R. "The Term Structure of Interest Rates", in Friedman, B., Hahn, F. (ed.), *Handbook of Monetary Economics*, North-Holland, 1990, Chapter 13.

Windas, T., *An Introduction to Option-Adjusted Spread Analysis*, Bloomberg Publishing, 1993.