1. Discount Factors and Interest Rates

Zero-coupon bonds

The zero-coupon bond is the simplest form of debt, and as such perhaps the most important, since it forms the foundation on which other pricing is built. The instrument itself is merely a promise to pay a certain amount N at some future time T, called the maturity of the bond. As with many things in finance, N goes by several different names but is often called the face amount or par amount of the bond. The most general term for this quantity is the notional amount of the instrument.

These instruments are common and in virtually all cases have short maturities. Governments issue a great deal of zero-coupon debt; all US Treasuries with maturities of one year or less, for example, are of this variety. Large companies issue short-term debt in the commercial paper market that is most commonly of this form. Short-term collateralized lending in the repo market (short for "repurchase agreement") is also usually structured this way, as are other very short-term unsecured obligations such as bankers' acceptances. Of course, in daily life most lending takes place this way as well: Informal personal loans often take the form of a promise to repay in a lump sum at a single future time.

The reason for the prevalence of this debt only at short maturities is that longer-term lenders tend to prefer intermediate payments on the loan amount. Intermediate payments of interest—called coupons—mitigate some risks of the instrument but also complicate the question of pricing somewhat. The zero-coupon instrument is named by contrast with more common coupon-bearing debt.

Factors affecting the market price of a zero-coupon bond

Suppose we are considering the purchase of a particular zero-coupon bond with defining features N and T as above, and that the bond is not actively traded. We wish to determine what the present value V of the future cash flow is to us. For the purposes of this analysis, we will disregard the possibility that the bond may not pay as promised.

Even if we are certain of receipt of the cash flow, the value of it to us today is under normal circumstances not the same as the bond's face amount. Purchasing the bond (making a loan) involves the loss of use of V over the bond's life. Unless the bond itself is indistinguishable from currency in terms of its liquidity characteristics—namely, it can be offered in trade for any consumption, investment, or tax needs—then the buyer of the bond requires compensation for loss of liquidity. Further, the effect of inflation means that the face amount N received at T will have less purchasing power than it would at present. These effects taken together are usually referred to colloquially as the time value of money.

Thus it is virtually always the case that a zero-coupon bond trades at a discount to its face amount, making $V < N$. As the maturity of the bond approaches, we would expect this difference to close—a phenomenon sometimes referred to as the "pull to par," generating “roll down” return on the investment—until when $T = 0$ the value of the bond and its face amount coincide.

This process is often not smooth, and various factors may influence market participants’ beliefs about what the present value of a particular future cash flow is. In order to assign a price to a future cash flow, it seems sensible to obtain market information about the current state of those beliefs. The benchmark asset selected for this purpose, which forms the basis of valuation, is referred to as the risk-free asset.

The law of one price

Suppose that in the case above, where we wish to determine the fair price of a zero-coupon bond that is not actively traded, we are able to find a traded risk-free zero-coupon bond with the same maturity T. It seems intuitively clear that the price of the risk-free asset should at the very least influence the prices of all other zero-coupon bonds with the same maturity.

This basic guiding principle of our approach will be that two different instruments that produce exactly the same cash flows under all circumstances must also have the same present value. This simple notion goes by the somewhat grandiose name of “the law of one price,” and like most model-based approaches, it relies on an idealization of how the world works.

In learning about this approach—arbitrage-free pricing—it is important not to become too enamored of the result or too dismissive of it, either. While finance is not a science in the pure sense, nevertheless comparisons with natural sciences may be useful in understanding the utility of the idea. Students of physics are commonly exposed, almost from the start, to Newton’s law of inertia, which in essence states that the “natural” motion of objects is to maintain a constant velocity.

This qualifies as an insight that is at once brilliant and utterly ludicrous. Any child quickly learns that objects tend to come to rest in time and stay that way; the idea that some special intervention is required to prevent an object from moving in a straight line at the same speed forever is contrary to observed experience in virtually every way. However, this idealization turns out to be quite productive since it offers a basis on which the causes of departure from this motion—what we call “forces”—can be studied.

Similarly, the idea that arbitrage does not exist in markets, under a set of very stylized assumptions, does not pretend to be a description of observed reality. Nor, it should be noted, is it a description of some ideal condition to which markets aspire, or inevitably tend. Failure to appreciate this fact has led to some fairly spectacular mishaps even in the relatively short history of “quant” investment.

Rather, arbitrage-free pricing offers a way of identifying and, one hopes, better understanding the nature and behavior of individual instruments, specific assets and market segments, and the idea of markets generally. By offering a straightforward approach to determining how instruments’ prices ought to move in relation to one another, it can not only make predictive statements of value but can also provide a tool for quantifying departures from the ideal.

In employing this approach, the main assumptions we will make initially are that market assets are available to be traded in any quantity (there are no limitations on liquidity or lot sizes), and that all market participants can go short or long market assets at the same price (there is no bid-ask spread). In a nod to physics, a market like this with no position limits or transaction costs is sometimes called a "frictionless" market.

Along the way, we may make other assumptions in our analyses and will always endeavor to identify them as they are made. The good news is that results obtained under a particular set of assumptions may be found to hold when some are relaxed or removed, and in other cases the same general approach can be extended to accommodate more realistic views of market operation.

Replication and hedging
Our model of an arbitrage-free market immediately suggests a method for pricing any instrument: If traded market instruments can somehow be used to replicate it, then it must be that the price of the instrument and the cost to replicate it are the same. Otherwise, if the instrument's price differs from its replication cost, then under the assumptions outlined above we can always buy the instrument via the cheaper method and sell it back at the same time via the more expensive one.

In this case, the question of how to replicate our zero-coupon bond is of course quite simple. We have assumed that the instrument has no risk of failing to pay, and a risk-free market instrument is available with the same maturity. Since we can trade any amount of it we like, we should be able to purchase the exact quantity that would pay N at T. Since the bond of interest can be precisely replicated using the risk-free asset, its price is therefore the same as the purchase price of a proper quantity of the risk-free asset.

Some prefer to think of such problems in terms of hedging instead, which is really just the other side of the same coin. Suppose we hold a position in the instrument whose price we seek to learn and form a portfolio by adding some hedge to it. If the result is a portfolio whose price we know, and if furthermore we know the value of the hedge, then we can easily obtain the price of the instrument as well.

In this simple case, we suppose we are long the zero-coupon bond whose price we wish to obtain. We could take a short position in the risk-free market asset paying N at T. The resulting portfolio is the simplest imaginable: At T, we receive N from the long bond position and pay the same amount due to our short position in the market asset, meaning that the portfolio involves no net cash flow at any time in the future.

It seems clear that the present value of such a portfolio must be zero, since under the assumption of no arbitrage we should not be able to generate cash flow by buying or selling anything that has no future cash flows. The result is that we are long an asset of unknown price, short an asset of known price, and taken together the portfolio's value is zero. Naturally, this must mean that the two bonds' prices are equal.

Discount factors

The observation that a risk-free market asset can be used to value any future risk-free cash flow leads us to a fundamental quantity used in pricing. If a risk-free bond paying N at time T has present value V, then we define the discount factor at this maturity P_T by:

$$P_T = \frac{V}{N}$$

This unitless quantity provides an easy means of determining the present value of any future cash flow c occurring at time T:

$$V = cP_T$$

As we have already seen, the nature of debt dictates that in virtually all cases, $V < N$, and thus $P_T < 1$, except of course P_0, which is 1. However, it is not impossible for discount factors to be greater than 1, and all of the results we will discuss here hold in this rare instance.

We may also use the discount factor as a measure of the growth of an investment. If we invest some amount V today in a risk-free asset with maturity T, then the asset likewise must repay us the amount N at T as long as the asset has no other cash flows or complicating features. That is, the asset is equivalent to the market risk-free asset, meaning that it too must obey…

$$V = NP_T$$

…and…

$$N = V\frac{1}{P_T}$$

…so that the reciprocal of the discount factor can thus be thought of as a growth factor for investments maturing at this time.

Although we will later introduce other more intuitive quantities that are used to describe market views of the time value of money, it must be emphasized that the notion of discount factor is the most fundamental and, from a pricing standpoint, convenient of them all. We will usually assume that risk-free zero-coupon bonds are available to be traded to all future times T, meaning that we will always know the complete term structure of discount factors. Nothing more is needed to price future cash flows, and indeed, there is a further interesting consequence of this term structure being completely specified.

Forward discount factors

Suppose we wish to enter into an agreement today with the following terms: At a time t in the future, we will pay some amount V_t in cash; at a time $T > t$, we will receive the amount N. In essence, we are committing to purchase at time t a zero-coupon bond with face amount N whose maturity at that time will be $T - t$. We wish to determine the fair amount V_t as seen today, at time zero.

The instrument we are considering buying is generally referred to as a forward contract and is actually quite simple to price. It consists of two cash flows: one of $-V_t$ at time t, and a second of N at time T. If the amount V_t is appropriately chosen, then we need neither pay nor receive anything today, and thus the net value of the agreement is zero. This means that:

$$NP_T - V_t P_t = 0$$

$$V_t = N\frac{P_T}{P_t}$$

Not only is this an answer to our question, but it also bears an interesting resemblance to an expression we have already seen. Recall that if we purchase a zero-coupon bond today paying N at T, then its value is:

$$V = NP_T$$

Above, we have concluded that the same bond purchased at the later time t has the fair price, as seen today, of:

$$V_t = N\frac{P_T}{P_t} = NP_{t,T}$$

It appears that the ratio of our two discount factors is now serving the same purpose in answering this question that the discount factor we observe today does in the question of what a bond purchased today is worth. Indeed, the value V_t we have determined here is known as the forward price of the bond, and the value $P_{t,\,T}$ the forward discount factor.

Forwards and arbitrage

It is important to understand that the value of the forward discount factor is not a forecast: The price of the bond with maturity $T - t$ may be anything it likes when t arrives, and indeed in practice it is rather unlikely that it will actually be V_t.

Instead, the forward discount factor is a consequence of the requirement that our market be free of arbitrage. One way of phrasing this requirement is that, for any choice of t and T with $t < T$, we must have:

$$P_T = P_t P_{t,T}$$

This amounts to a requirement that it should be impossible to generate different returns today with the combination of an investment maturing at t and a second investment in the forward from t to T than could be generated by a single investment maturing at T.

To illustrate this, consider the contract above where we agree today to purchase at t a bond whose maturity, seen today, is T. Suppose for the sake of illustration that the price attached to this contract is some $V_K > V_t$. Then the forward-starting bond is, by our calculations, too expensive; we ought to sell this contract, meaning that we agree to receive V_K at t in return for paying N at T.

As we indicated above, however, the actual price of the bond when t arrives may be quite different from V_K, and indeed in the meantime the market may have changed in such a way that we incur a loss. In order to ensure a risk-free gain, we must find a way to hedge the exposure.

Since we will owe N at T, a sensible start to forming the hedge is to purchase a zero-coupon bond that will pay this amount at this time; the price of this bond today is NP_T. We wish to hedge at zero net cost, so we must finance this position somehow; we choose to do so by going short a zero-coupon bond maturing at t whose value today is NP_T , the purchase price of the bond maturing at T.

So we have now set up a portfolio, consisting of the forward and the hedge, at zero cost. When t arrives, we receive the amount V_K from the forward contract. At the same time, the short position from our hedge in the bond maturing at t requires us to repay the bond. The amount we owe on this position can be calculated as we indicated above, using the initial amount of the loan and the reciprocal of the discount factor P_t:

$$NP_T \frac{1}{P_t} = N\frac{P_T}{P_t} = NP_{t,T} = V_t$$

That is, using instruments tradeable at time zero, we have synthetically manufactured an instrument with value V_t at t. Our portfolio pays V_K at this time, we owe V_t, and as stated initially, $V_K > V_t$. So we receive a net positive cash flow at this time.

When T arrives, our short position via the forward requires us to pay N, while at the same time our hedge produces N from the long bond position. Net, then, our hedged portfolio produced only a positive cash flow at t with certainty, and furthermore this portfolio cost nothing to set up, resulting in arbitrage. We can produce the same situation if $V_K < V_t$ merely by reversing our positions in the forward in the hedge. We conclude, then, that any choice of V_K other than V_t leads to arbitrage. This is true no matter what the price of the bond whose initial maturity was T actually is when t arrives.

Interest rates

While discount factors can be taken as the fundamental expression of the value of future cash flows, the numbers themselves are not particularly intuitive. For example, it is common (although by no means essential) in debt markets for investors to demand a greater rate of return for investments with longer maturities. Using only discount factors as a reference, however, it can be difficult to discern whether or not a given term structure satisfies this condition. For example, consider the simple term structure:

time (months)	discount factor
0	1
1	0.995850622
2	0.992391664

It is far more natural to evaluate the term structure in terms of a rate of return rather than this, which is essentially raw price information. We imagine investing a unit amount today and receiving the amount implied by each discount factor when maturity arrives. The rate implied by this arrangement is known as an interest rate. More specifically, it is often called a “zero rate,” since it is the rate that matches the price implied by a zero-coupon bond.

There are several different ways, however, in which we might imagine actually realizing this return. The result is that there are several different ways of expressing the interest rate.

Simple interest

We have already seen that, if we invest V_0 today in a zero-coupon bond with maturity T, then the amount we will receive at maturity is $V_T = V_0 / P_T$. The easiest way we can imagine the return accruing is linearly in proportion to the initial investment, meaning that the investment gains a constant dollar amount per unit time. We will call this the rate of simple interest r_s over this period. The equation of the line passing through our initial and final points is thus:

$$V_T = r_s V_0 T + V_0$$

$$V_T = V_0 \left(1 + r_s T\right)$$

$$\frac{V_T}{V_0} = 1 + r_s T$$

$$\frac{1}{P_T} = 1 + r_s T$$

As a result, we can convert back and forth between a rate of simple interest and a discount factor by:

$$P_T = \frac{1}{1 + r_s T} = \left(1 + r_s T\right)^{-1}$$

$$r_s = \frac{1}{T}\left(\frac{1}{P_T} - 1\right)$$

Both expressions seem quite sensible: The discount factor is, as always, the reciprocal of the amount into which an initial unit investment grows over the time period. In the second expression, the quantity inside parentheses is the gain attributable to time value—i.e., the actual amount of interest paid by the investment. Dividing by the maturity spreads this amount evenly over the entire period.

Many interest rates in fixed-income markets are expressed as a rate of simple interest. LIBOR, a common reference rate in transactions, is reported as a rate of simple interest. Furthermore, as we will see later, the rate a coupon-bearing bond pays is expressed (with rare exceptions) as a rate of simple interest.

To apply this to the sample term structure above, we must choose a value T to assign to a one-month maturity. In real fixed-income markets, this turns out not to be a trivial concern in some cases, but here we simply state without complication that each month is 1 / 12 of a year. Under this assumption, the term structure of zero rates becomes:

time(months)	T	discount factor	zero rate (simple interest, rounded to nearest basis point)
0	0	1	n/a
1	0.083333333	0.995850622	0.0500
2	0.166666667	0.992391664	0.0460

As indicated, the rate reported here is rounded to the nearest basis point—10^{-4}, or 1% of 1%. As we can see, this term structure turns out to be unusual in that the rate of return over two months is less than the rate of return over the first month.

Discretely compounded interest
Particularly for longer-term investments, it is not realistic to suppose that the interest amount accrues linearly over its entire term. Most commonly, a debt will make periodic payments of accumulated interest. These periodic payments are then available to be reinvested, giving rise to what you no doubt know is called compound interest.

We can therefore imagine replicating the return of a risk-free zero-coupon bond using an account that accrues linearly as above for a time, but then makes a payment of interest which is then reinvested in the account. The number of compoundings m per year is referred to as the compounding frequency of the account.

If we invest a given amount at time zero in such an account paying the rate r_m, then after a single period (time 1 / m) the value is:

$$V_{\frac{1}{m}} = V_0\left(1+\frac{r_m}{m}\right)$$

This amount then becomes available to be reinvested in the account at the same rate, which after two periods has value…

$$V_{\frac{2}{m}} = V_0\left(1+\frac{r_m}{m}\right)^2$$

…and, in general, after n periods:

$$V_{\frac{n}{m}} = V_0\left(1+\frac{r_m}{m}\right)^n$$

We observe that since each time period is of length 1 / m, this can be expressed in terms of the maturity T in years as:

$$V_T = V_0\left(1+\frac{r_m}{m}\right)^{mT}$$

$$\frac{V_T}{V_0} = \left(1+\frac{r_m}{m}\right)^{mT}$$

$$\frac{1}{P_T} = \left(1+\frac{r_m}{m}\right)^{mT}$$

We arrived at this expression by assuming *T* contained an integer number of compounding periods, but from the above it is apparent that this quantity is well defined for any maturity *T*. Thus, we can convert back and forth between discount factors and rates of compound interest using:

$$P_T = \left(1+\frac{r_m}{m}\right)^{-mT}$$

$$r_m = m\left(P_T^{-\frac{1}{mT}} - 1\right)$$

In the case of our sample term structure, we can express the discount factors using rates of compound interest instead:

T	discount factor	zero rate (simple interest, rounded to nearest bp)	zero rate (compounded annually, to the nearest bp)	zero rate (compounded monthly, to the nearest bp)
0	1	n/a	n/a	n/a
0.083333333	0.995850622	0.0500	0.0512	0.0500
0.166666667	0.992391664	0.0460	0.0469	0.0459

Note that for the 1-month rate, in the case where the time between compoundings matches the maturity, the rates of simple interest and compound interest coincide, as we would expect from the way we initially defined compound interest. When the time between compoundings is less than the maturity, the compound rate required to produce the same discount factor is less, since the compounding effect produces greater return than simple interest does. Conversely, when the time between compoundings is greater than the maturity, the compound rate that produces the same discount factor must be greater than the rate of simple interest. That is, there is no time-independent way of converting between rates of simple and compound interest.

By contrast, however, consider the case of a conversion between rates with different compounding frequencies m_1 and m_2. The most convenient way to perform such conversions is by way of the discount factor:

$$\left(1+\frac{r_{m_1}}{m_1}\right)^{-m_1T} = P_T = \left(1+\frac{r_{m_2}}{m_2}\right)^{-m_2T}$$

$$\left(1+\frac{r_{m_1}}{m_1}\right)^{-m_1} = \left(1+\frac{r_{m_2}}{m_2}\right)^{-m_2}$$

This expression can easily be solved for the rate of interest, but its most important feature is that, unlike conversions to simple interest, the conversion between compound rates with different frequencies is time-invariant. As would be expected, a given rate with more frequent compoundings produces greater returns, meaning that increasing the compounding frequency in which the market rate to a given maturity is expressed decreases the rate reported.

Continuously compounded interest
The observation above suggests the question of what happens when we allow the compounding frequency to become arbitrarily large. While of course it is physically not possible to pay interest continually and continually reinvest it, in theory it can be imagined, and the result is quite useful for further theoretical questions of continuous-time finance.

We recall that the expression for the value at maturity T of an initial amount V_0 invested at rate r with compounding frequency m is…

$$V_T = V_0\left(1+\frac{r}{m}\right)^{mT}$$

…and thus in the theoretical case outlined above, we seek an expression for:

$$V_T = \lim_{m\to+\infty} V_0\left(1+\frac{r}{m}\right)^{mT}$$

Taking account of constants, this becomes…

$$V_T = V_0\left[\lim_{m\to+\infty}\left(1+\frac{r}{m}\right)^{m}\right]^{T}$$

…which, recalling a result familiar from your first calculus course…

$$\lim_{n\to+\infty}\left(1+\frac{x}{n}\right)^{n} = e^{x}$$

…becomes:

$$V_T = V_0 e^{rT}$$

$$\frac{V_T}{V_0} = e^{rT}$$

$$\frac{1}{P_T} = e^{rT}$$

This is referred to as the rate of continuously compounded interest, and as with the other cases above, we can convert back and forth from discount factors quite easily:

$$P_T = e^{-rT}$$

$$r = -\frac{1}{T}\ln P_T$$

In our further discussions where interest rates are involved, unless otherwise specified we will assume that the stated rate is quoted with continuous compounding.

Instantaneous forward rates

Above, we developed the idea of a forward discount factor and saw that, due to no-arbitrage requirements, a given term structure of discount factors uniquely determines all forward discount factors between any two times t and T. Further, we saw that for a specified quotation method, the term structure of discount factors can be expressed as a term structure of zero rates, since after all any discount factor can be uniquely converted into a zero rate for a given maturity. Denote this term structure by r_t, a function of the continuously compounded zero rate with respect to time.

It thus follows that for any pair of times t and T, we can express the forward discount factor $P_{t,T}$ in rate terms using various compounding methods; these rates are referred to as forward rates.

Consider the forward rate beginning at t over the interval Δt (that is, $T = t + \Delta t$) quoted with continuous compounding. Using our previous results, we see that this can be calculated as…

$$r_{t,t+\Delta t} = -\frac{1}{\Delta t}\ln P_{t,t+\Delta t} = -\frac{1}{\Delta t}\ln\frac{P_{t+\Delta t}}{P_t}$$

$$r_{t,t+\Delta t} = -\frac{\ln P_{t+\Delta t} - \ln P_t}{\Delta t}$$

…which is a familiar-looking expression. We define the instantaneous forward rate at time t by:

$$r_t^* = \lim_{\Delta t \to 0} r_{t,t+\Delta t} = -\lim_{\Delta t \to 0}\frac{\ln P_{t+\Delta t} - \ln P_t}{\Delta t} = -\frac{d}{dt}\ln P_t$$

This is interesting, since by the fundamental theorem of calculus we can then say:

$$\ln P_T - \ln P_t = -\int_t^T r_s^* ds$$

$$\ln P_T = \ln P_t - \int_t^T r_s^* ds$$

$$P_T = P_t e^{-\int_t^T r_s^* ds} = P_t P_{t,T}$$

That is, the integral of the instantaneous forward rate over the time from t to T determines what the forward discount factor will be. The corresponding result applies in the case of the zero rate as well, since if we take $t = 0$, then $P_t = 1$ and the result becomes…

$$P_T = e^{-\int_0^T r_s^* ds} = e^{-r_T T}$$

…where r_T, as we said above, is the zero rate to that maturity. It then becomes clear how these zero rates relate to the instantaneous forwards:

$$-\int_0^T r_s^* ds = -r_T T$$

$$r_T = \frac{1}{T}\int_0^T r_s^* ds$$

Namely, the zero rate to each maturity is just the average value of the instantaneous forward rate over that time period.

Interpolation in the term structure of discount factors: linear spot rates
Although we have assumed throughout that a risk-free zero-coupon bond is traded to every maturity T, of course this is physically impossible. Given a discrete set of traded risk-free instruments, to determine how to discount cash flows at intermediate times, we must choose a method of interpolating between the traded points, which here we will refer to as the "knot points" of our term structure. Many different methods can conceivably be used for this purpose; here, we will discuss three of interest.

The simplest is merely to assume that the zero rate is linear between knot points; since zero rates are sometimes referred to as "spot rates" (as opposed to forward rates), the interpolation method is sometimes referred to as linear spot rate interpolation. Take as an example the following term structure of zero rates:

maturity	T	df	r (continuous)
o/n	0	1	0.005
1W	0.019230769	0.999865394	0.007
1M	0.083333333	0.9990005	0.012
3M	0.25	0.996257022	0.015
6M	0.5	0.992031915	0.016
1Y	1	0.980198673	0.02
2Y	2	0.964640293	0.018
3Y	3	0.938943474	0.021

The interest rate at the first point is usually taken from an overnight rate, and is assumed to be the instantaneous rate right now.

If we are interested in the zero rate r_t at a time t between knot point times t_i and t_{i+1}, with zero rates r_i and r_{i+1}, then we can simply calculate…

$$r_t = \frac{r_{i+1} - r_i}{t_{i+1} - t_i}(t - t_i) + r_i$$

…and as always convert to the discount factor:

$$P_t = e^{-r_t t}$$

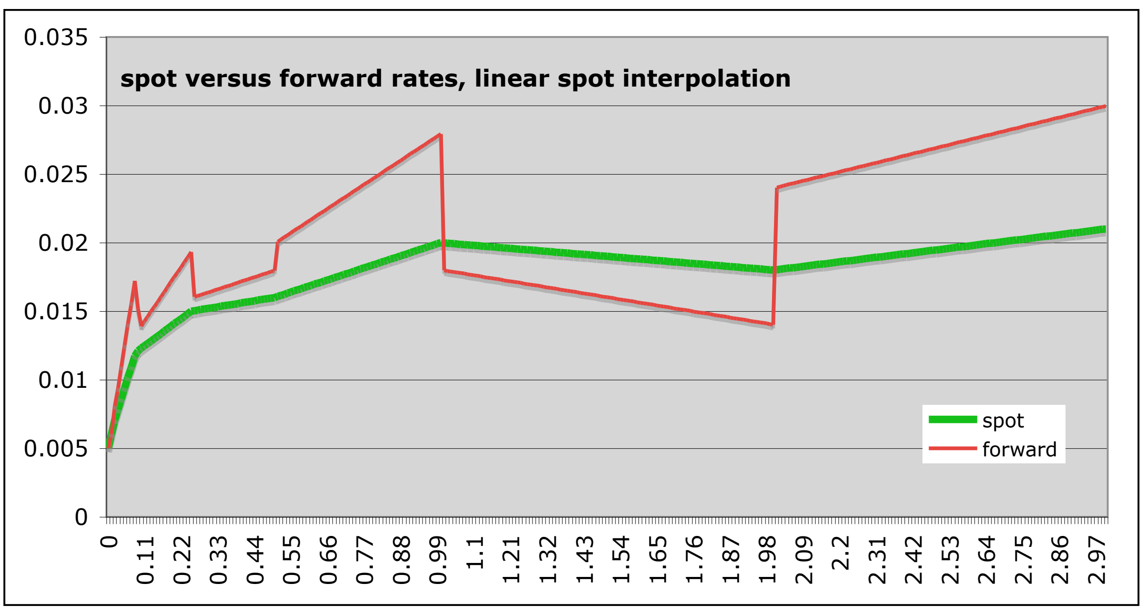
This method is simple and intuitive, and indeed is quite commonly used. However, it does have some drawbacks, particularly in regard to the forward rates it generates. Recall that the instantaneous forward rate can be expressed as:

$$r_t^* = -\frac{d}{dt}\ln P_t = -\frac{d}{dt}\ln e^{-r_t t} = \frac{d}{dt} r_t t$$

In this case, we know that r_t is explicitly of the form $at + b$, calculated as above. This means that on a given interval between knot points…

$$r_t^* = \frac{d}{dt} at^2 + bt = 2at + b$$

…meaning that the forward rates are twice as steeply sloped as the spot rates are; moreover, the instantaneous forward is discontinuous at the knot points. We can see this effect in action by calculating the forward rate over small intervals using the sample data above; here, we show both the zero rates and the (almost) instantaneous forward rates for comparison's sake, showing discontinuities as vertical lines:



Clearly this method is not particularly suitable for the pricing or trading of instruments that are highly sensitive to forward rates. Moreover, it is difficult to imagine what mechanics of interest rates would cause such a term structure of forward rates to arise.

Piecewise constant forward rates

In the absence of any further information, it seems simplest to explain the current term structure of spot rates by assuming that the forward rate between each pair of adjacent knot points is constant. While again this method results in unappealing discontinuities in the forward rate at each knot point, at least it does not impose illogical forward rates within each segment that are purely artifacts.

As saw above in our discussion of instantaneous forward rates, the forward is the opposite of the slope in the log of the discount factor…

$$r_{i,i+1} = -\frac{1}{t_{i+1} - t_i} \ln \frac{P_{t_{i+1}}}{P_{t_i}} = -\frac{\ln P_{t_{i+1}} - \ln P_{t_i}}{t_{i+1} - t_i}$$

…and we calculate the discount factor at times between these two knot points by…

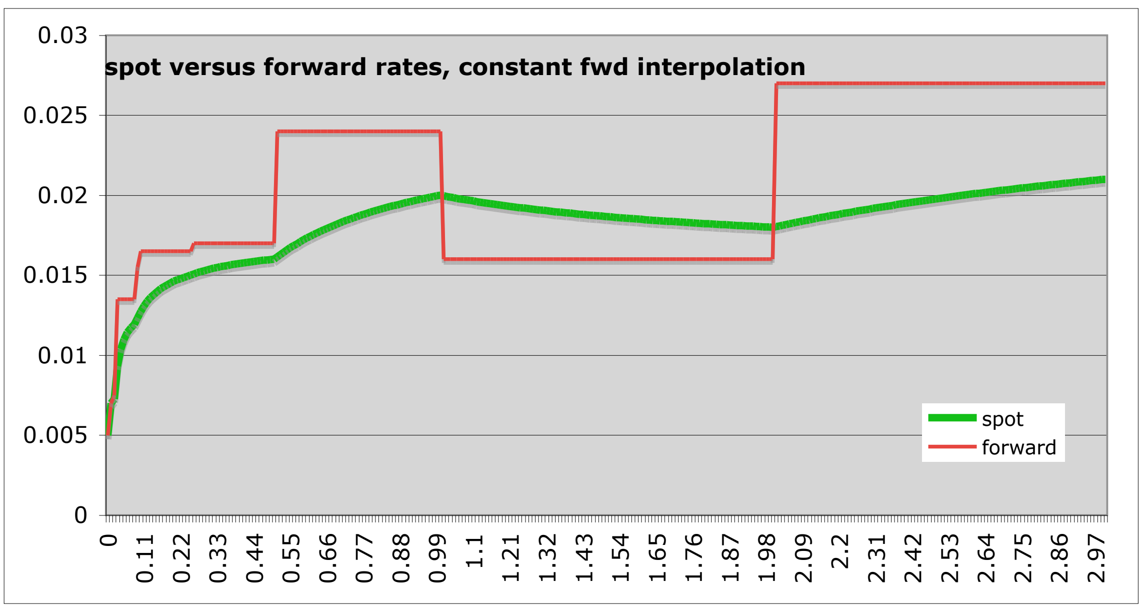
$$P_t = P_{t_i} e^{-r_{i,i+1}(t - t_i)}$$

$$\ln P_t = -r_{i,i+1}(t - t_i) + \ln P_{t_i}$$

$$\ln P_t = \frac{\ln P_{t_{i+1}} - \ln P_{t_i}}{t_{i+1} - t_i}(t - t_i) + \ln P_{t_i}$$

…making the method equivalent to linear interpolation in the log of the discount factor.

As we can see, this method results in a far less odd term structure of forward rates:



The resulting term structure of spot rates, however, is hardly the most attractive. In the end, both of the methods discussed so far suffer due to the fact that the resulting spot rates lack a smooth—or even a continuous—derivative.

Cubic spline interpolation

The solution to this problem is to introduce constraints that cause the resulting interpolated rate curve to be smooth at each knot point. A "spline" is a method whereby a function is specified piecewise to connect the knot points, forming a continuous function. The method we initially used could be thought of as a linear spline in the spot rates.

A common method of creating a smooth function from discrete points is to assume that the curve between knot points is cubic. The particular cubic function used in each segment can then be chosen so that the resulting curve is not only continuous at each knot point, but its first and second derivatives are as well. Here, we will construct a cubic spline in the spot rates; to do this, we will need to solve a linear system of equations.

We will construct a cubic spline through knot points at times $t_0, t_1, \ldots, t_N$ with corresponding spot rates $r_0, r_1, \ldots, r_N$. In this case, we take $t_0 = 0$ and r_0 equal to the overnight rate.

These $N + 1$ knot points delineate N segments, the i^{th} of which we will fit using the cubic polynomial C_i, so that for…

$$t \in (t_{i-1}, t_i]$$

$$r_t = C_i(t) = a_i t^3 + b_i t^2 + c_i t + d_i$$

Since we have N such segments, the unknown coefficients in each segment require us to specify $4N$ constraints to determine the function. There is more than one equivalent way to formulate these constraints, and even a small measure of discretion in how they are formed, depending upon the desired features at each extreme end of the curve. The below is a commonly used set of constraints.

Our N segments must pass through $N - 1$ interior knot points and the two knot points at either extreme end of the curve. We will begin by formulating constraints around the interior points.

At each interior node $i = 1, 2, \ldots, N - 1$, we require that:

$C_i(t_i) = r_i$: The segment to the left of this point passes through it.
$C_i(t_i) = C_{i+1}(t_i)$: The curve is continuous at this point.
$C'_i(t_i) = C'_{i+1}(t_i)$: The derivative of the curve is continuous at this point.
$C''_i(t_i) = C''_{i+1}(t_i)$: The second derivative of the curve is continuous at this point.

With four constraints per interior knot point, we have a total of $4(N - 1)$ specified already. There remain four to specify.

Two are given by the requirement that the curve pass through the extreme points:

$C_1(t_0) = r_0$
$C_N(t_N) = r_N$

This leaves two remaining to be specified. The most common requirement is that the spot curve be linear at either extreme limit, or namely that:

$C''_1(t_0) = 0$
$C''_N(t_N) = 0$

We can express these requirements in the form of the matrix equation $Ax = y$, with the column vector x containing the unknown coefficients and the matrix A and column vector y appropriately chosen to represent the constraints. This equation can then be solved using commonly available methods to determine the coefficients.

The sample data above, which requires us to find the coefficients of seven segments, is too large to represent here. To illustrate the formation of the various components of the equation, we choose the simple case with three knot points at times t_0, t_1, t_2 corresponding to rates r_0, r_1, r_2. This requires us only to find the equations of two cubic segments, making the resulting square matrix A of size eight.

In this case, we would form the matrix…

$$A = \left[\begin{matrix} t_0^3 & t_0^2 & t_0 & 1 & 0 & 0 & 0 & 0 \\ 6t_0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ t_1^3 & t_1^2 & t_1 & 1 & 0 & 0 & 0 & 0 \\ t_1^3 & t_1^2 & t_1 & 1 & -t_1^3 & -t_1^2 & -t_1 & -1 \\ 3t_1^2 & 2t_1 & 1 & 0 & -3t_1^2 & -2t_1 & -1 & 0 \\ 6t_1 & 2 & 0 & 0 & -6t_1 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & t_2^3 & t_2^2 & t_2 & 1 \\ 0 & 0 & 0 & 0 & 6t_2 & 2 & 0 & 0 \end{matrix}\right] \begin{matrix} \leftarrow C_1(t_0) \\ \leftarrow C_1''(t_0) \\ \leftarrow C_1(t_1) \\ \leftarrow C_1(t_1) - C_2(t_1) \\ \leftarrow C_1'(t_1) - C_2'(t_1) \\ \leftarrow C_1''(t_1) - C_2''(t_1) \\ \leftarrow C_2(t_2) \\ \leftarrow C_2''(t_2) \end{matrix}$$

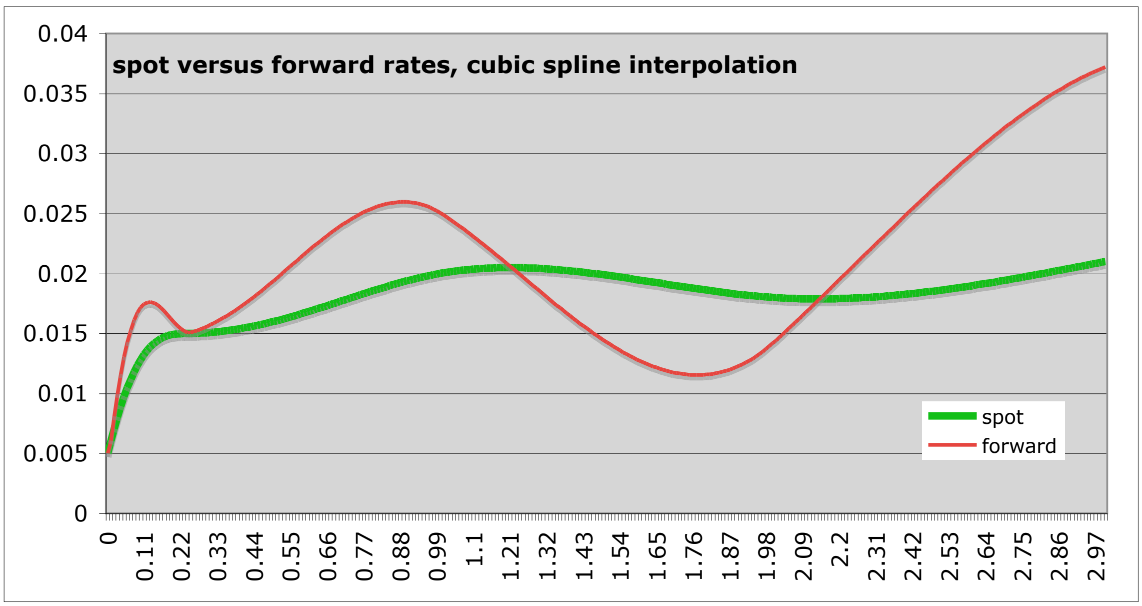
…the vector of unknown coefficients…

$$x = \begin{bmatrix} a_1 \\ b_1 \\ c_1 \\ d_1 \\ a_2 \\ b_2 \\ c_2 \\ d_2 \end{bmatrix}$$

…and the constants:

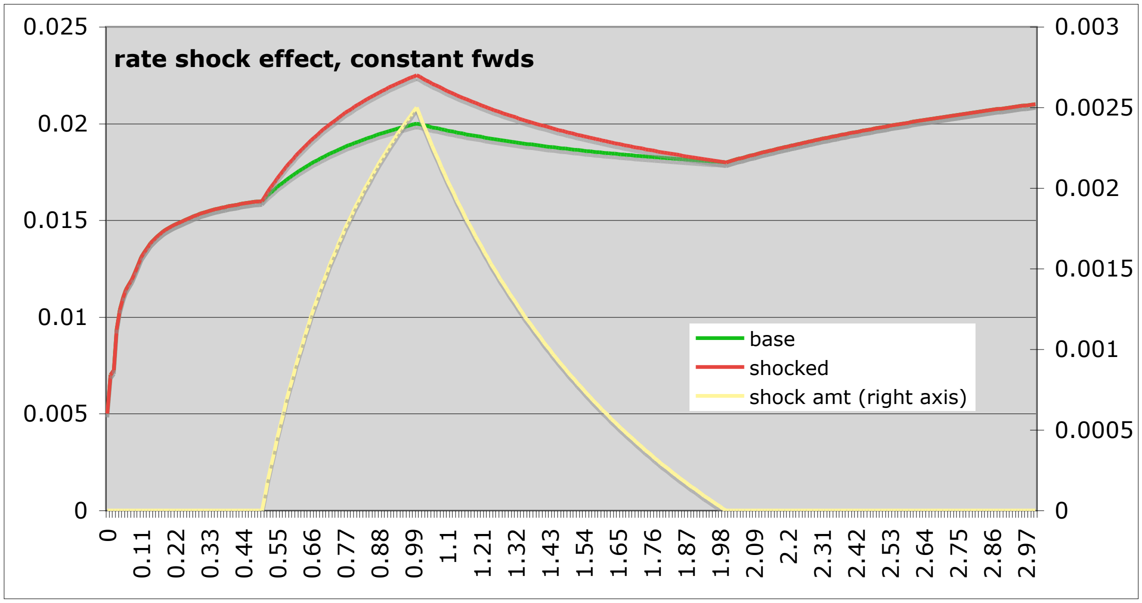
$$y=\begin{bmatrix} r_0 \\ 0 \\ r_1 \\ 0 \\ 0 \\ 0 \\ r_2 \\ 0 \end{bmatrix}\begin{array}{l} \leftarrow C_1(t_0)=r_0 \\ \leftarrow C_1''(t_0)=0 \\ \leftarrow C_1(t_1)=r_1 \\ \leftarrow C_1(t_1)-C_2(t_2)=0 \\ \leftarrow C_1'(t_1)-C_2'(t_2)=0 \\ \leftarrow C_1''(t_1)-C_2''(t_2)=0 \\ \leftarrow C_2(t_2)=r_2 \\ \leftarrow C_2''(t_2)=0 \end{array}$$

In the case of our sample data, the method gives spot and forward rates that look like the following:



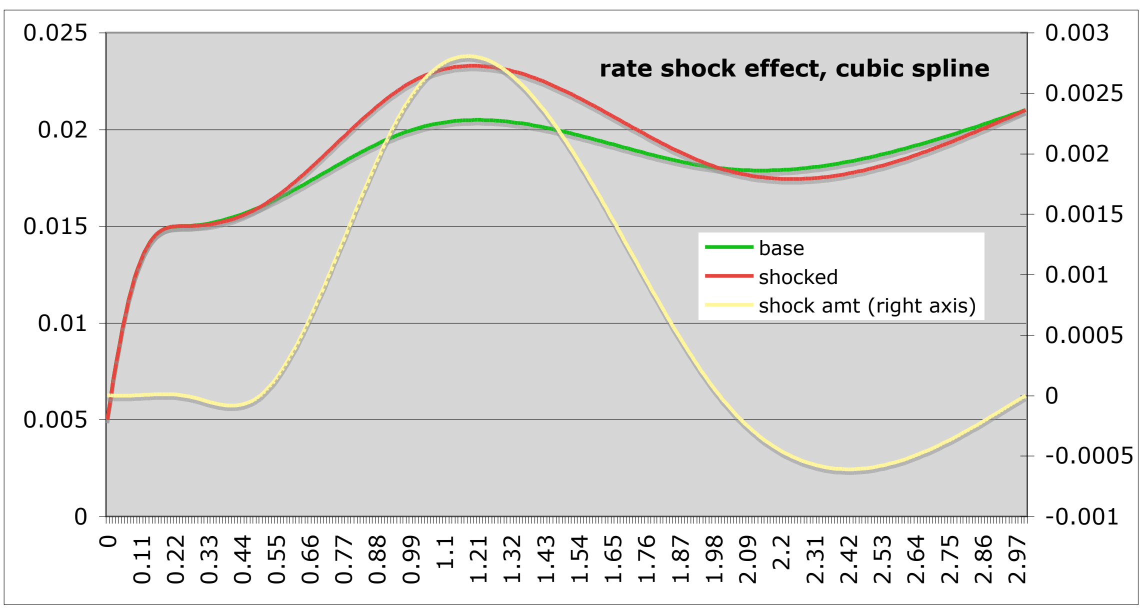
As you can see, this interpolation method does produce far more attractive spot rates, but it does little to address some of the problems we saw with forward rates when doing simple linear interpolation. Indeed, it seems that the smoothness requirement can produce even wilder variations in the forward rates.

Further problems with smoothness
Another consequence of imposing a smoothness requirement on the resulting rates only becomes apparent when we consider the effect of perturbations—shocks—in the base interest rate curve. For example, consider the effect of increasing the one-year zero rate 25 basis points but leaving all other knot points the same. When we interpolate using constant forward rates, the result is:



In this case, the changes in zero rates are confined, as one might expect, to the intervals to either side of the one-year point, where the shock was applied. The interpolation method adjusts the forward rate used on either side of the shocked point but does not change the interpolated zero rate in any other interval. Furthermore, the direction of the change in the interpolated rate is everywhere in the same direction as the applied shock.

By contrast, consider the corresponding result when a cubic spline interpolation method is used:



In this case, the requirement that we must have two continuous derivatives at each interior knot point propagates the shock over the entire term structure of zero rates.

Essentially only the knot points themselves are left undisturbed by this shock, and in the interval between two and three years, there are actually interpolated rates that go down as a result of the upward shock at the one-year point. While the mathematical reasons for this are straightforward, there is no financial motivation for such an observation.

Such effects do have practical consequences. Fixed-income managers are often concerned with their portfolios' sensitivities to individual maturities on the risk-free rate curve. Calculating such sensitivities in conjunction with a spline method can give misleading results by causing value changes in instruments whose cash flows are quite distant from the shocked point. If a method imposing continuity of forward rates is chosen for interest rate interpolation, some adjustment of the naïve method is required to prevent spurious results in this sort of an analysis.