Smooth Interpolation of Zero Curves

Ken Adams

Smoothness is a desirable characteristic of interpolated zero curves; not only is it intuitively appealing, but there is some evidence that it provides more accurate pricing of securities. This paper outlines the mathematics necessary to understand the smooth interpolation of zero curves, and describes two useful methods: cubic-spline interpolation—which guarantees the smoothest interpolation of continuously compounded zero rates—and smoothest forward-rate interpolation—which guarantees the smoothest interpolation of the continuously compounded instantaneous forward rates. Since the theory of spline interpolation is explained in many textbooks on numerical methods, this paper focuses on a careful explanation of smoothest forward-rate interpolation.

Risk and other market professionals often show a keen interest in the smooth interpolation of interest rates. Though smooth interpolation is intuitively appealing, there is little published research on its benefits. Adams and van Deventer's (1994) investigation into whether smooth interpolation affected the accuracy of pricing swaps lends some credence to the intuitive belief that smooth interpolation gives more accurate results than linear interpolation. The authors took swap rates for maturities of one, two, three, five, seven and 10 years together with the six-month zero rate, removed the seven-year swap rate from the data and created the implied zero-rate curve from the remaining data. The resulting zero-rate curve was used to calculate the missing seven-year swap rate which was then compared to the actual seven-year swap rate. The authors found that swap rates calculated with smoothly interpolated zero-rate curves were closer to the actual seven-year swap rate than swap rates calculated with curves that were linearly interpolated.

This paper aims to describe the mathematics and finance theory necessary for an understanding of smooth interpolation. There is a simple mathematical definition of smoothness, namely, **a smooth function** is one that has a continuous differential. Thus, any zero curve that can be represented by a function with a continuous first derivative is necessarily smooth. However, interpolated zero curves are not necessarily smooth; for example, curves that are created by the well-known technique of linear interpolation of a set of yields are not smooth, since the first derivative is discontinuous. This article extends the simple mathematical definition of smoothness, and then describes the implications of the extended definition for the smooth interpolation of zero curves.

The mathematics of zero curves

The mathematics of zero curves is derived from the prices of discount bonds; a **discount bond** being a security that pays, with certainty, a unit amount at maturity. The following axioms define the discount bond market:

- The market trades continuously over its trading horizon: it extends from the current time to some distant future time such that the maturities of all the instruments to be valued fall between now and the trading horizon.

- The market is frictionless: no transaction costs or taxes are incurred in trading, there are no restrictions on trade (legal or otherwise) such as margin requirements on short sales, and the goods in the market are infinitely divisible.
- The market is competitive: every trader can buy and sell as many bonds as desired without changing the market price.
- The market is efficient: information is available to all traders simultaneously, and every trader makes use of all the available information.
- The market is complete: any desired cash flow can be obtained from a suitable self-financing strategy based on a portfolio of discount bonds.
- There are no arbitrage opportunities: the price of a portfolio is the sum of its constituent parts.
- All traders in the market act to maximize their profits: they are rational and prefer more to less.

These axioms are necessary for the development of the mathematics of zero curves. However, they may not apply to real markets. For example, one conclusion that can be drawn from the axioms is that the market prices equal the intrinsic value of the bonds; that is, there is no "noise" in the market prices. Further, the market completeness axiom implies that all points on the zero curve are known. In fact, the zero curve is not defined by an infinite set of values, but rather by a discrete, finite set.

Discount bonds are often called zero coupon bonds, in contrast to **coupon bonds** that make more than one cash payment to their owner. It is usual to distinguish between yield curves derived from the periodically compounded yields to maturity of coupon-bearing bonds and yield curves derived from the continuously compounded yields to maturity of zero coupon (discount) bonds. This article is concerned only with the latter, which are called **zero curves**. In addition, the term **zero yield** is used to refer to the continuously compounded yield to maturity of a zero coupon bond. Note that all rates are continuously compounded. It is an easy matter to convert to and from periodic compounding, and the use of continuous compounding enables the expression of the mathematics of zero curves in a particularly simple and elegant form, which greatly simplifies the discussion.

Prices and yields

Consider a bond which is sold now, at time t, and is due to mature at time x, where $t \le x < x_\infty$. The trading horizon, x_∞, is much greater than zero and is longer than the maturity of any bond. Suppose the price of the bond is denoted by $P(t,x)$. Since the bond pays a unit amount at maturity, we must have $P(x,x) = 1$. When $t < x$ the bond sells at a discount and $P(t,x) < 1$. Thus, in general, we have $P(t,x) \le 1$.

Now, define the zero yield in terms of the bond price. The zero yield, as seen at time t, of a bond that matures at time x, $t \le x < x_\infty$, is denoted by $y(t, x)$ and is defined, for $t < x$, by the relationship

$$P(t,x) = exp(-(x-t)y(t, x)) \qquad (1)$$

Equation 1 states that the price of the bond at time t is equal to its discounted value. Note that this relationship does not define $y(t,t)$ since $P(T, T) = 1$ for all T; so, for a discount bond maturing at $t = T$, both sides of the equation are equal to unity, irrespective of the value of $y(t,t)$. The zero yield in terms of the price of a bond is obtained by rearranging Equation 1:

$$y(t,x) = \frac{-ln(P(t,x))}{x-t} \qquad (2)$$

provided $t < x$.

Forward rates

Suppose that at time t we enter into a forward contract to deliver at time x_1 a bond that will mature at time x_2. Let the forward price of the bond be denoted by $P(t, x_1, x_2)$. At the same time, a bond that matures at time x_1 is purchased; the price of this bond is $P(t, x_1)$. Further, again at time t, a bond that matures at time x_2 is bought; the price of this bond is $P(t, x_2)$. Note that the complete-market axiom guarantees that these bonds exist. In addition, the axiom specifying that there are no arbitrage opportunities implies that the price of the bond maturing at time x_2 must be equal to the product of the price of the bond maturing at time x_1 and the forward price:

$$P(t,x_2) = P(t,x_1)P(t,x_1,x_2) \tag{3}$$

Let the implied forward rate, as seen at time t, for the period x_1 to x_2 be $f(t, x_1, x_2)$, defined by:

$$P(t,x_1,x_2) = exp(-(x_2 - x_1)f(t,x_1,x_2)) \tag{4}$$

Note the similarity between the definition of the implied forward rate (as defined in Equation 4) and the zero rates (as defined in Equation 1). On substituting Equation 1 and Equation 4 into Equation 3 we obtain

$$\begin{aligned} exp(-(x_2 - t)y(t,x_2)) &= exp(-(x_1 - t)y(t,x_i)) \\ &\times exp(-(x_2 - x_1)f(t,x_1,x_2)) \end{aligned} \tag{5}$$

Rearranging Equation 5 gives

$$f(t, x_1, x_2) = \frac{(x_2 - t)y(t,x_2) - (x_1 - t)y(t,x_1)}{x_2 - x_1} \tag{6}$$

The forward rate $f(t, x_1, x_2)$, defined in Equation 6, is the **period forward rate**. However, the *instantaneous* forward rate is of much greater importance in the theory of the term structure. The **instantaneous forward rate** for time x, as seen at time t, is denoted by $f(t,x)$ and is the continuously compounded rate defined by

$$f(t,x) = \lim_{h \to 0} f(t,x,x+h) = y(t, x) + (x - t)y_x(t,x) \tag{7}$$

where

$$y_x(t,x) = \frac{\partial}{\partial x}y(t,x)$$

To derive an equation for the instantaneous forward rates in terms of the bond prices, Equation 2 can be rearranged to obtain

$$-ln(P(t,x)) = (x - t)y(t, x) \tag{8}$$

Differentiating Equation 8 with respect to x gives

$$\frac{-P_x(t, x)}{P(t, x)} = y(t, x) + (x - t)y_x(t, x) \tag{9}$$

Finally, by direct comparison of Equation 7 and Equation 9

$$f(t,x) = \frac{-P_x(t,x)}{P(t,x)} \tag{10}$$

It is now possible to define $y(t, t)$, which, as noted, is not defined by Equation 2. First, note that Equation 7 implies $f(t,t) = y(t,t)$. Then, noting that $P(t, t) = 1$, Equation 10 can be used to obtain

$$y(t,t) = \lim_{x \to t} \frac{-P_x(t,x)}{P(t,x)} = \lim_{x \to t} -P_x(t,x)$$

Defining the zero curve

Assume that the prices of all bonds in the market are known; the implication being that the value of $y(t, x)$ for $(t \le x \le x_\infty)$ is known. Then, the current zero curve (i.e., the one seen at time t) comprises the zero yields, as seen at time t, of the zero coupon bonds, which mature between t and x_∞, inclusive; that is, the current zero curve defined by $y(t, x)$ for $(t \le x \le x_\infty)$.

Though the zero curve is defined in terms of the zero yields, it can be defined in terms of the instantaneous forward rates. In this case, the zero curve is defined by the instantaneous forward rates $f(t,x)$ for $t \le x \le x_\infty$. The zero yield in terms of the instantaneous forward rate is obtained by integrating Equation 7:

$$y(t, x) = \frac{1}{x - t}\int_t^x f(t,u)du \tag{11}$$

This completes the essential mathematical theory of zero curves. In the following sections, the relevance of this theory to the interpolation of zero curves is shown, with particular emphasis on smoothest forward-rate interpolation.

The smooth interpolation of zero curves

To construct zero curves from market data, assume that the n data values are

$$\{(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)\}$$

where $0 \le x_1 < x_2 < \ldots < x_n < x_\infty$ are the times to maturity of $n \ge 1$ zero coupon bonds and $y_i = y(0,x_i)$ is the zero rate corresponding to the time to maturity x_i $(i = 1,2,\ldots,n)$. Note that we have implicitly set $t = 0$, as is customary when constructing a zero curve from current market data. This allows the simplification of the notation as follows. Use $y(x)$ and $f(x)$ to denote $y(0,x)$ and $f(0, x)$, respectively. Using this new notation, $y_i = y(x_i)$.

In developing the mathematical theory of zero curves, it is assumed that the value of $y(t, x)$ for $(t \le x \le x_\infty)$ is known. In reality, the current zero curve is not defined by this infinite set of values implied by the complete market axiom, but, rather, by a set of discrete data values $\{(x_i,y_i)\}$, each value comprising a time to maturity and a zero rate. If we wish to use the mathematics of zero curves derived above, the discrete set of values must be extended to an infinite set. This is achieved by defining the current zero curve by a combination of the set of discrete data values and a method for interpolating those values. Given these, the value of $y(x)$ for any value of x in the range $(0 \le x < x_\infty)$ can be found. Note that one consequence of this definition of the zero curve is that changing the interpolation method changes the zero curve.

Interpolation

Interpolation methods provide a means to calculate values of $y(x)$ for times x_i that do not coincide with the given times to maturity, $x_1,x_2,\ldots,x_n$. Though there are many interpolation methods, here, we consider those methods that require knowledge of the data points only. This excludes, for example, Hermitian interpolation, which requires knowledge of the derivative values as well as the values at the data points. In addition, the use of algebraic polynomials, such as Lagrange polynomials, are excluded because the order of the interpolating polynomial must, in general, be $n-1$, which implies that there could be as many as $n-2$ maxima and minima, and this is not a desirable property of a zero curve.

Consequently, we consider only piecewise spline curves. Splines were originally strips of elastic material used by engineering draughtsmen to draw smooth curves through a given set of points, known as knot points. Being elastic, the splines assume the shape that minimizes their strain energy. Unconstrained, this shape is a straight line. However, when the splines are constrained to pass through a set of points, and no other constraints are imposed (e.g., they were not twisted at the ends), an elastic spline assumes a shape that is "as straight as possible."

Analogously, the zero curve is defined by the n data points $\{(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)\}$, where each one of the data values represents a knot point, that is a point at which the (as yet unspecified) spline segments join. If the spline segment function for the interval $x_i \le x < x_{i+1}$ is denoted by S_i, then the zero curve can be defined by the set of functions $\{S_1,S_2,\ldots,S_{n-1}\}$. If it is necessary to extrapolate beyond the end values, (x_1,y_1) and (x_n,y_n), two further spline functions, S_0 and S_n, will be needed; the former being used for the range at the left-hand end, $0 \le x < x_1$, and the latter for the range at the right-hand end, $x_n \le x < x_\infty$. Though nothing has been said about the form of the spline functions, polynomial splines are sufficient for our purpose.

The simplest type of interpolation algorithm is the two-point algorithm, where interpolation of values in the interval $x_i \le x < x_{i+1}$ depends only on the two points (x_i,y_i) and (x_{i+1},y_{i+1}). In contrast, the multi-point algorithm requires knowledge beyond the adjacent knot points. The well-known technique of linear interpolation is a two-point algorithm, and cubic-spline interpolation is a multi-point algorithm.

Smoothness

Recall that a smooth function is one that has a continuous differential. Thus, any zero curve that can be represented by a continuous function is smooth. However, the same is not necessarily true for interpolated zero curves; for example, curves that are created by linear interpolation are not smooth. This observation applies to all two-point interpolation formulae, since, in general, there is a discontinuity in the first derivative at the knot points. For similar reasons, it also applies to any method that does not use the full set of data points in the construction of the spline curve. In general, there must be at least one knot point at which the derivatives are not continuous. Consequently, only those interpolation methods that use all of the data points to construct the spline curve are considered.

The mathematical definition of smoothness does not help to distinguish between different spline functions; in particular, it does not provide a measure of smoothness. To define a measure of smoothness, we begin with a simple idea that makes intuitive sense and then give it a precise meaning. Intuitively, interpolation functions with the smallest number of maxima and minima have the fewest possible "bends," that is, they are as close as possible to a straight line. Recall that straight lines can not be used to join the knot points, because when data sets contain more than two points, linear interpolation is not smooth.

Elastic splines are a mechanical equivalent of our intuitive idea to draw a curve with the fewest possible bends, as well as a means of defining a precise measure of smoothness. The strain energy depends on the shape, $g(x)$, assumed by the spline ($g(x)$ is used to avoid confusion with $f(x)$, which is used to denote forward rates). It can be shown (see, for example, Schwarz (1989)) that the strain energy over the interval $[a,b]$ is related to the quantity

$$\int_a^b \left(\frac{\partial^2}{\partial x^2} g(x)\right)^2 dx$$

Though the theory of strain energy is not discussed, note that the smaller the quantity, the less the strain energy or, intuitively, the smaller the "bending" of the elastic spline. This quantity is taken as the measurement of smoothness for the splines: the smaller this measure of smoothness, the smoother the interpolating curve. The measure of smoothness shall be used to determine the "best" interpolation methods for zero curves.

A second property of elastic splines is useful. Since the only constraints are that the splines must pass through the given points, the parts of the splines that project beyond the curve defined by the specified points are not subject to any constraints. This implies that the portions of the splines beyond the ends of the specified points are linear (since this minimizes the strain energy in those parts of the splines). Thus, the second derivatives of the splines beyond the given data points are zero.

The determination of the smoothest possible interpolation method depends on whether we want to find the smoothest zero curve or the smoothest forward-rate curve (specifically, the smoothest continuously compounded forward-rate curve). Thus, two interpolation methods are considered. Cubic-spline interpolation guarantees the smoothest zero curve, and smoothest forward-rate interpolation guarantees the smoothest continuously compounded forward-rate curve. Since cubic-spline interpolation is a standard technique dealt with in many books on numerical methods, it is reviewed only briefly here.

Smoothest zero-rate interpolation

The cubic-spline interpolation method produces smooth zero curves. Moreover, it can be shown that (see, e.g., Burden and Faires (1997)), for the measure of smoothness defined by the strain energy, no interpolating function passing through the given data values is smoother than the cubic spline passing through the same points. The most commonly used cubic spline is the **natural cubic spline**, which is constructed so that the second derivatives at both end points are zero. This is analogous to allowing the ends of the splines to be unconstrained so that the free ends are linear.

Values beyond the two end points are calculated using linear extrapolation. However, if the gradient of this line is too steep, the extrapolated values may be unacceptably high (positive gradient) or negative (negative gradient). If our aim is to create a zero curve with a better shape, it is possible to constrain the cubic spline so that the gradient at the right-hand end is zero; the constrained spline behaves like zero curves that tend to flatten at longer maturities. Then linear extrapolation gives a smooth curve, albeit a horizontal one for maturities longer than the maturities in the data set.

A **financial cubic spline** denotes a cubic spline that is constrained so that its derivative at its right-hand end is zero, and its second derivative at the left-hand end is also zero. These additional constraints mean that it will have a slightly different shape than the natural cubic spline passing through the same set of points, so it will not be the smoothest curve to interpolate those points (the natural cubic spline is). However, no other interpolation function that is subject to the same constraints as the financial cubic spline, and which fits the given data, is smoother than the financial cubic spline that interpolates that data.

A further property of cubic-spline interpolation is worthy of mention. The general equation for a cubic is $y = a + bx + cx^2 + dx^3$, suggesting the need for four coefficients (a, b, c and d) for each section of the spline curve. However, it is also possible to define a cubic spline in terms of the given x and y values and the second derivatives at the knot points. Thus, to create a cubic-spline zero curve, it is necessary to find only the second derivatives at the knot points, which leads to a set of tri-diagonal linear equations. A standard algorithm for solving sets of linear equations is LU decomposition and back-substitution. When the set of equations is tri-diagonal, the algorithm takes a particularly efficient form so that the implementation of cubic-spline interpolation is computationally efficient. These algorithms are described in standard texts on numerical methods (see, e.g., Burden and Faires (1997)).

Smoothest forward-rate interpolation

Though smooth zero curves are desirable, practitioners often state a preference for zero curves that have the smoothest forward rates. The interpolating function that guarantees the smoothest continuously compounded instantaneous forward-rate curve is a quartic spline. Recall that the data consist of zero rates, not forward rates. Therefore, the quartic spline that is constructed does not pass through the data points. In this section, the equations of the smoothest forward-rate interpolating function are given, and the linear equations to be solved to find the coefficients defining that function are specified.

We wish to construct a quartic spline for each of the $n-1$ sections between the knot points. The ith spline segment can be expressed as

$$S_i(x) = a + bx + cx^2 + dx^3 + ex^4 \qquad (12)$$

where $S_i(x)$ represents the function $f(x)$ over the range $x_i \le x < x_{i+1}$.

Each spline segment is a quartic polynomial and there are $n-1$ segments; $5(n-1) = 5n-5$ unknown coefficients must be found.

The system of linear equations is defined by the following constraints:

- the n original zero rates must be recoverable from the zero curve
- there must be continuity:
    - at the $n-2$ interior knot points
    - of the first derivatives at the $n-2$ interior knot points
    - of the second derivatives at the $n-2$ interior knot points
    - of the third derivatives at the $n-2$ interior knot points.

Thus far, $n + 4(n-2) = 5n-8$ conditions have been defined. The additional three conditions imposed are $f'(x_n) = 0$, $f''(x_1) = 0$ and $f''(x_n) = 0$. The first condition, $f'(x_n) = 0$, ensures that the right-hand end of the curve is flat. The last two of these conditions constrain the second derivatives of the forward curve at both the left-hand and right-hand ends to be zero. Together, these $(5n-5)$ conditions ensure that the interpolating spline has the smoothest instantaneous forward rates.

Each of these constraints is considered in constructing the linear equations that will be solved to find the coefficients.

Recovering the yields

Recall from Equation 11 that:

$$(x-t)y(t,x) = \int_t^x f(t,u)du \tag{13}$$

Then, since $t = 0$, Equation 13 takes the form

$$xy(x) = \int_0^x f(u)du$$

This must apply at each knot point. Thus, at the knot point (x_i, y_i)

$$x_{i+1}y_{i+1} = \int_0^{x_i} f(u)du + \int_{x_i}^{x_{i+1}} f(u)du = x_iy_i + \int_{x_i}^{x_{i+1}} f(u)du \tag{14}$$

This is true for $1 \le i \le n-1$. Substituting Equation 12 for $f(x)$ in Equation 14 over this range and integrating gives

$$\begin{aligned} x_{i+1}y_{i+1} - x_iy_i &= a_ix_{i+1} - x_i + \frac{b_i}{2}x_{i+1}^2 - x_i^2 \\ &+ \frac{c_i}{3}(x_{x+1}^3 - x_i^3) + \frac{d_i}{4}(x_{i+1}^4 - x_i^4) \\ &+ \frac{e_i}{5}(x_{i+1}^5 - x_i^5) \end{aligned} \tag{15}$$

We consider two cases. In the first case, the spot rate at time zero is known; in the second case, it is unknown.

Equation 15 can be used to match all yields except the first. If the spot rate at $t = 0$ is known ($x_1 = 0$), then $f(x_1) = f(0) = y_1$ and

$$y_1 = S_1(x_1) = S_1(0) = a_1$$

In the second case, where $0 < x_1$, the spot rate at time zero is unknown. Again, from Equation 11

$$\int_0^{x_1} f(u)du = x_1y_1 \tag{16}$$

In addition, the functional form of $f(x)$ must be specified. Since $S_1''(x_1) = 0$, an obvious choice is to use linear extrapolation for the range $0 \le x \le x_1$; this corresponds to the straight line that an unconstrained elastic spline would take. Thus, the equation for the left-hand extrapolation is:

$$f(x) = f_1 + m(x - x_1) \tag{17}$$

where

$$f_1 = a_1 + b_1x_1 + c_1x_1^2 + d_1x_1^3 + e_1x_1^4 \tag{18}$$

$$m = b_1 + 2c_1x_1 + 3d_1x_1^2 + 4e_1x_1^3 \tag{19}$$

The values of m and f_1 are derived directly from Equation 12. Substituting Equation 17 into Equation 16, we obtain

$$xy = \int_0^x (f_1 + m(u - x_1))du$$

Hence, the extrapolated zero rate is

$$y = f_1 + \frac{mx}{2} - mx_1 \tag{20}$$

The validity of Equation 20 can be checked by noting that $y(0) = f_1 - mx_1 = f(0)$, as expected. Therefore,

$$y(0) = y_0 = f_1 - \frac{mx_1}{2} = a_1 + \frac{b_1x_1}{2} - \frac{d_1x_1^3}{2} - e_1x_1^4 \tag{21}$$

Equation 21 is also valid for $x_1 = 0$; so it suffices whether or not the spot rate at time zero is known.

First, note that the validity of the extrapolation backward from x_1 to time zero depends on the value of x_1. Here, x_1 is the shortest maturity; so the larger its value, the greater the time to maturity over which extrapolation is linear. Thus, the smaller x_1, the better; if possible, the overnight rate should be used.

Continuity constraints

Continuity at the interior knot points is equivalent to requiring that $S_i(x_{i+1}) = S_{i+1}(x_{i+1})$ for $0 \le i \le n-2$, so that, on substituting the equations for S_i and S_{i+1}

$$a_i + b_i x_{i+1} + c_i x_{i+1}^2 + d_i x_{i+1}^3 + e_i x_{i+1}^4$$
$$= a_{i+1} + b_{i+1} x_{i+1} + c_{i+1} x_{i+1}^2 + d_{i+1} x_{i+1}^3 + e_{i+1} x_{i+1}^4$$

Continuity of the first derivatives at the interior knot points is equivalent to $S'_i(x_{i+1}) = S'_{i+1}(x_{i+1})$ for $1 \le i \le n-1$, so that

$$b_i + 2c_i x_{i+1} + 3d_i x_{i+1}^2 + 4e_i x_{i+1}^3$$
$$= b_{i+1} + 2c_{i+1} x_{i+1} + 3d_{i+1} x_{i+1}^2 + 4e_{i+1} x_{i+1}^3$$

Continuity of the second derivatives at the interior knot points is equivalent to

$S''(x_{i+1}) = S''_{i+1}(x_{i+1})$ for $1 \le i \le n-1$, so that

$$c_i + 3d_i x_{i+1} + 6e_i x_{i+1}^2 = c_{i+1} + 3d_{i+1} x_{i+1} + 6e_{i+1} x_{i+1}^2$$

Continuity of the third derivative at the interior knot points is equivalent to the constraint $S'''_i(x_{i+1}) = S'''_{i+1}(x_{i+1})$ for $1 \le i \le n-1$, so that

$$d_i + 4e_i x_{i+1} = d_{i+1} + 4e_{i+1} x_{i+1}$$

Additional constraints

The first additional constraint, $f'(x_n) = 0$, implies

$$b_{n-1} + 2c_{n-1} x_n + 3d_{n-1} x_n^2 + 4e_{n-1} x_n^3 = 0$$

The second additional constraint, $f''(x_n) = 0$, implies

$$c_{n-1} + d_{n-1} x_n + e_{n-1} x_n^2 = 0$$

The third additional constraint, $f'''(x_1) = 0$, implies

$$c_1 + 3d_1 x_1 + 6e_1 x_1^2 = 0$$

Solving the system of equations

The linear equations defined above are arranged into blocks of five equations each. If the blocks are numbered $1, 2, \ldots, n-1$ to correspond to the spline segments $S_1, S_2, \ldots, S_{n-1}$, then the rows in block i ($1 \le i < n-1$) are, in order,

- Match y_{i+1}
- Ensure continuity at internal knot points: $S_i(x_{i+1}) = S_{i+1}(x_{i+1})$
- Ensure continuity of first derivative at internal knot points: $S'_i(x_{i+1}) = S'_{i+1}(x_{i+1})$
- Ensure continuity of second derivative at internal knot points: $S''_i(x_{i+1}) = S''_{i+1}(x_{i+1})$
- Ensure continuity of third derivative at internal knot points: $S'''_i(x_{i+1}) = S'''_{i+1}(x_{i+1})$

The last block, $i = n-1$, contains the rows

- Match y_n
- Set $S'_{n-1}(x_n) = 0$
- Set $S''_{n-1}(x_n) = 0$
- Match y_1
- Set $S''_1(x_1) = 0$

Note that the second, third and fifth lines contain the additional constraints.

Unlike cubic splines, the system of linear equations describing quartic splines cannot be represented by a tri-diagonal matrix; standard diagonal decomposition and back-substitution must be used to solve the system of equations. (For a description of this standard method, see Burden and Faires (1997)).

Once the coefficients are known, it is possible to interpolate and extrapolate values from the zero curve, such that these values correspond to the smoothest forward-rate curve.

Interpolating values from the zero curve

In order to interpolate the value of $f(x)$ for $x_1 < x \le x_n$, first determine the index i such that

$x_i \le x < x_{i+1}$. Then, from Equation 11

$$\begin{aligned} y &= \frac{1}{x}\left(\int_0^{x_i} f(u)du + \int_{x_i}^{x} f(u)du\right) = \frac{1}{x}\left(y_i x_i + \int_{x_i}^{x} f(u)du\right) \\ &= \frac{1}{x}\Big(x_i y_i + a_i(x - x_i) + \frac{b_i}{2}(x^2 - x_i^2) \\ &\quad + \frac{c_i}{3}(x^3 - x_i^3) + \frac{d_i}{4}(x^4 - x_i^4) + \frac{e_i}{5}(x^5 - x_i^5)\Big) \end{aligned}$$

The value $x = x_1$ has been excluded to avoid problems with division by zero when $x_1 = 0$.

Extrapolating values from the zero curve

It has already been shown that, if $0 \le x \le x_1$, the forward rates can be extrapolated using Equation 17. Now, the value of the zero rate at $x = 0$ is given by

$$y(0) = f(0) = f_1 - m x_1$$

where f_1 and m are defined by Equations 18 and 19.

Two cases must be considered, $0 \le x \le x_1$ and $x_n < x$.

The integral of the forward rate is

$$\int_0^x f(u)du = \int_0^x (f_1 + m(u - x_1))du = x\left(f_1 + m\left(\frac{x}{2} - x_1\right)\right)$$

where $0 \le x < x_1$. Therefore,

$$y = f_1 + m\left(\frac{x}{2} - x_1\right)$$

Finally, consider extrapolating zero rates for values of x such that $x_n < x$. Recall that the additional constraints stipulate that the first and second derivatives of the right-hand end of the forward-rate curve be zero, that is

$$S'_{n-1}(x_n) = 0$$

$$S''_{n-1}(x_n) = 0$$

Thus, it is possible to extrapolate the right-hand end of the forward-rate curve with constant value f_n where

$$f_n = a_{n-1} + b_{n-1}x_n + c_{n-1}x_n^2 + d_{n-1}x_n^3 + e_{n-1}x_n^4$$

Thus, the integral of the forward rate is

$$\begin{aligned} \int_0^x f(u)du &= \int_0^{x_n} f(u)du + \int_{x_n}^{x} f(u)du \\ &= x_n y_n + \int_{x_n}^{x} f_n du \\ &= x_n y_n + (x - x_n) f_n \end{aligned} \qquad (22)$$

From Equation 22, the zero rate is

$$y = \frac{x_n y_n + (x - x_n) f_n}{x}$$

for $x_n \le x$.

Comparing interpolation methods

Most practitioners judge the quality of a zero curve not by the quality of the underlying mathematics, but by the quality of the curve itself. In the absence of an objective measure of quality, they rely on subjective observation of the candidate curves. To demonstrate the nature of the curves produced by the methods discussed, zero curves constructed from the same initial data using different interpolation methods are presented and compared. The data is presented in Table 1.

Maturity (years)	Zero rate
0.5	0.0552
1	0.0600
2	0.0682
4	0.0801
5	0.0843
10	0.0931
15	0.0912
20	0.0857

Table 1: Zero-rate data

The following three graphs show the zero- and forward-rate curves interpolated from this data using the smoothest forward rate, natural cubic-spline and financial cubic-spline interpolation methods. Note that though the maximum maturity in the data is 20 years, the graphs have been extended to 30 years to show the differences between the extrapolated values.

Figure 1 shows the curves constructed from the smoothest forward-rate interpolation method. Note that the gradient of the forward-rate curve is zero at the maximum maturity of 20 years and the gradient is zero when extrapolated. Beyond the maximum maturity, the zero curve tends towards the forward-rate curve with increasing maturity. Both the zero and the forward-rate curves are smooth.

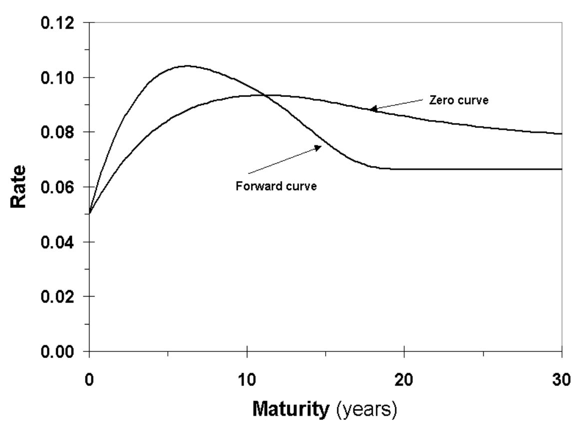

Figure 1: Smoothest forward-rate interpolation

Figure 2 shows the curves constructed from the natural cubic-spline interpolation method. Note that the extrapolated zero curve is downward sloping, whereas most zero curves tend to zero gradient at the right-hand end. Also, note that the extrapolated zero curve is linear and leads inevitably to negative zero and forward rates. Both the zero and the forward-rate curves are smooth.

Figure 3 shows the curves constructed from the financial cubic-spline interpolation method. This method corrects the problem with the extrapolated rates, but distorts the forward-rate curve. Though the zero curve is smooth, there is an abrupt transition at the maximum maturity (20 years) where the forward-rate curve ceases to be smooth.

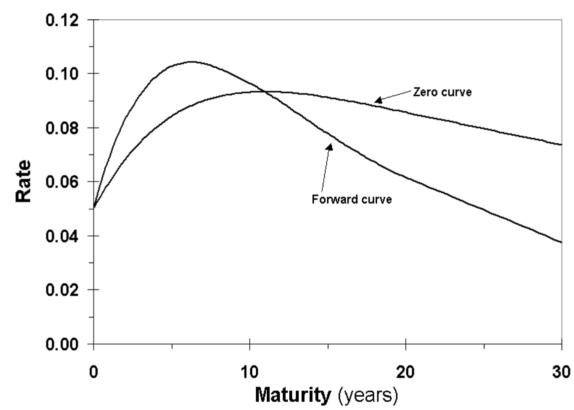

Figure 2: Natural cubic-spline interpolation

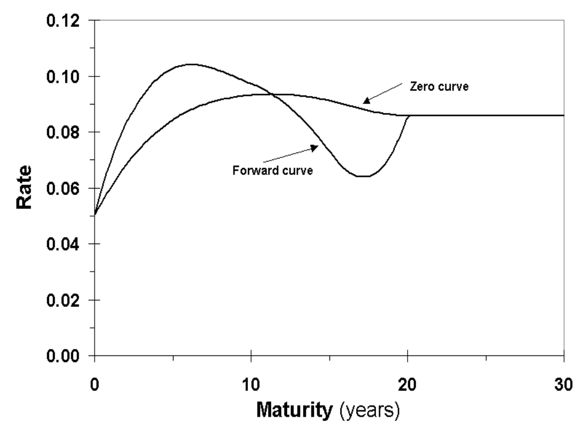

Figure 3: Financial cubic-spline interpolation

The zero curves are very similar in the maturity range of the given data, but differ substantially in the extrapolated sections. As illustrated in Figure 4, the zero curves are similar up to about 11 years when the zero rates attain their maximum value. Beyond that point, they show considerable differences and only the smoothest forward-rate curve has a shape consistent with that of the majority of zero curves.

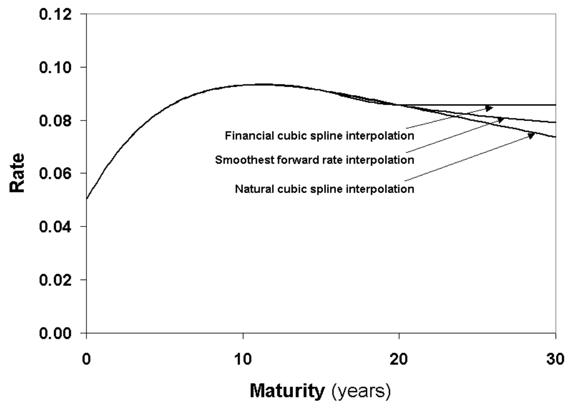

Figure 4: Comparing zero curves

The differences in the resulting forward-rate curves are more notable, as illustrated in Figure 5. Though the data was chosen to illustrate the differences among the interpolation methods, financial cubic-spline interpolation always causes similar distortions of the forward-rate curve. Though natural cubic-spline interpolation does not distort the forward-rate curve, the right-hand end of the curve does not flatten, and the method is less suitable for calculating rates for maturities close to the maximum maturity.

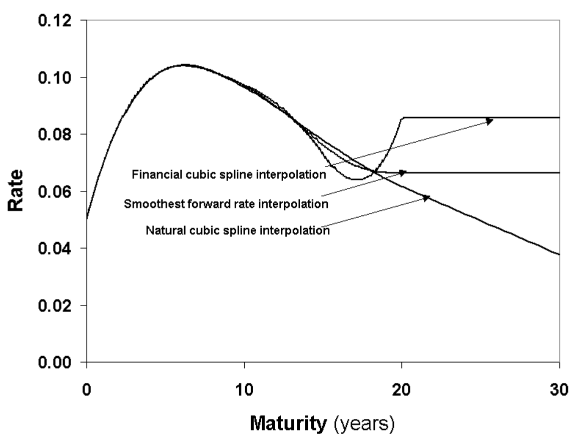

Figure 5: Comparing forward curves

Discussion

Given that there are three ways to interpolate the zero curve, it is natural to ask what differences arise from the use of these methods. First, it is important to note that all three methods interpolate the same set of points, and all produce smooth zero curves. Only smoothest forward-rate and natural cubic-spline interpolation methods produce smooth (instantaneous) forward-rate curves. In addition, all three methods guarantee the smoothest interpolation of a curve, though the smoothest curve depends on the chosen method. The natural cubic-spline interpolation method guarantees the smoothest zero curve and a smooth, but not necessarily the smoothest, (instantaneous) forward-rate curve. Similarly, the financial cubic-spline interpolation method guarantees the smoothest zero curve whose right-hand end is constrained to have a zero gradient; it does not guarantee a smooth forward-rate curve. Finally, the smoothest forward-rate interpolation guarantees, as its name implies, the smoothest (instantaneous) forward curve and a smooth, but not necessarily the smoothest, zero curve.

There is some evidence that using smooth zero curves results in more accurate pricing, though there is insufficient evidence to show that one interpolation method is always the best. Hence, practitioners have to make a choice, and that will depend on the nature of the market data from which the zero curve is constructed. Swap traders, in particular, desire smooth forward rates and one would expect them to prefer smoothest forward-rate interpolation to cubic-spline interpolation. Other practitioners may regard the degree of smoothness of the zero curve to be paramount. In that case, they will have to choose between the two cubic-spline interpolation methods, choosing the one that best suits their needs. Both cubic-spline interpolation methods give better results for short- and medium-term maturities, whereas, the smoothest forward-rate interpolated curve can be used over the whole range of maturities. As smoothest interpolation methods are more widely adopted, the benefits and trade-offs of the various methods will be better defined and better understood.

Conclusions

This discussion has extended the mathematical concept that a smooth function is one that has a continuous first derivative. To do this, a measure of smoothness used by engineers when fitting smooth curves to a finite set of points was adopted. This measure of smoothness implies that there in no smoother interpolating function than the set of cubic splines interpolating the given points.

In addition, we develop a theory of the mathematics of zero curves that enables the definition of a real zero curve in terms of a set of market data points and an interpolation method. Without this dual identity (data points and interpolation method) the assumed bond market is incomplete and the theory does not apply. Many practitioners have been in the habit of extracting a curve by bootstrapping, which typically implies the use of a two-point interpolation formula, and then using a different interpolation formula when using the curve to value cash flows. The dual nature of interpolated zero curves implies that more care should be taken in defining the way in which the rates are interpolated.

Finally, by combining the mathematical theories of zero curves and smoothness, we show that the smoothest interpolation method depends on whether the smoothest zero rates or the smoothest, continuously compounded forward rates are desired. The well-known cubic-spline interpolation method ensures that the smoothest zero rates approach is well established as an interpolation method in financial software systems. The smoothest forward-rate interpolation method, which ensures the smoothest continuously compounded forward rates, is as well known. This method uses quartic splines and it was possible to give a full account of how to set up the system of linear equations to solve for the coefficients of these quartic splines. In addition, it is shown how to interpolate values and how to deal with rates beyond the ends of the zero curve.

There is anecdotal evidence that finance practitioners consider maximum smoothness to be intuitively important. The little research that has been done indicates there is some basis for this intuitive judgement. A better understanding of the interpolation of zero curves depends on the dual nature of the definition of a zero curve in terms of both the market data points and the interpolation method. This understanding—together with the knowledge of which method guarantees which smoothest curve, and experience in the use of those methods—will allow risk practitioners to make informed choices about the appropriate interpolation method to use in different situations.

References

Adams, K. and D. van Deventer, 1994, “Fitting yield curves and forward rate curves with maximum smoothness,” *Journal of Fixed Income*, June, 4(1):52–62.

Burden, R. and J. Douglas Faires, 1997, *Numerical Analysis,* New York, NY: Brooks/Cole Publishing Co.

Schwarz, H., 1989, *Numerical Analysis: A Comprehensive Introduction*, Stanford, CT: Wiley & Sons.