Applied Mathematical Finance 7, 257–270 (2000)

A generalized bootstrap method to determine the yield curve

RICHARD DEAVES[1] and MAHMUT PARLAR[2*]

[1]*Crummer Graduate School of Business, Rollins College, 1000 Holt Avenue, 2722 Winter Park, FL 32789-4499, USA*

[2]*DeGroote School of Business, McMaster University, Hamilton, Ontario, L8S 4M4, Canada*

A new technique is described for operationalizing the bootstrap methodology to estimate the yield curve given any available data set of bond yields. The problem of missing data points is dealt with using symbolic cubic spline interpolation. To make such an approach tractable the computer algebra system Maple is employed to symbolically generate the interpolation equations for the missing data points and to solve the nonlinear equation system in order to obtain the points on the yield curve. Several examples with real data demonstrate the usefulness of the methodology.

Keywords: bootstrap methodology, yield curve, symbolic cubic spline interpolation

1. Introduction

The term structure of interest rates is one of the most fundamental relationships in finance. This is because the yield curve can be used for such diverse purposes as predicting future real economic activity (Harvey, 1991), fixed income return enhancement (Deaves, 1997) and pricing interest rate derivatives (Jarrow, 1996). Unfortunately the relationship between yield and maturity is not directly observable in the market,[1] which means it must be estimated based on the data that analysts have at their disposal.

This is a task that has occupied researchers for some time. Early approaches (Bradley and Crane, 1973; Echols and Elliott, 1976), sought to account for nonlinearities and coupon effects. More sophisticated mathematical modelling began at about this time with McCulloch's (1971) polynomial spline method and later with the exponential spline method of Vasicek and Fong (1982). More recently, Nelson and Siegel (1987) and Svensson (1994) have proposed parametric models specifying functional forms for the forward rate from which yield curves can be straightforwardly derived.

If an analyst were able to possess a sufficient quantity of reliable friction-free government bond

Research supported by the Social Sciences and Humanities Research Council of Canada and the Natural Sciences and Engineering Research Council of Canada

*Corresponding author. e-mail: parlar@mcmaster.ca. Telephone: 1 (905) 525-9140, Ext. 22858.

[1] With the advent of the strips market, this statement is not true as it once was. Nevertheless, this market is characterized by certain frictions that render implied yields less than ideal; see Sundaresan (1997, pp. 201–202).

Applied Mathematical Finance ISSN 1350-486X print/ISSN 1466-4313 online
http://www.tandf.co.uk/journals
DOI: 10.1080/13504860010021162

data, a technique known as the bootstrap would straightforwardly derive the yield curve.[2] This technique is based on the notion that individual coupon-paying bonds can be viewed as 'packages' of pure discount bonds. For example, a three-year bond is comprised of six pure discount bonds, namely the five coupon payments to be paid every six months, and the final payment which is the sum of the final coupon and the return of principal. This suggests that a bond's value can be viewed as either the present value of future cash flows discounted at the yield to maturity, or as the sum of the values of individual pure discount bonds, each of which is a present value of a cash flow discounted at its own time-specific yield.

The classic 'textbook' explication of the bootstrap usually begins by assuming the existence of a set of perfectly spaced bonds: for example, a 6-month, a 12-month, an 18-month and so on.[3] If these bonds have frictionless market prices, the bootstrap renders the correct yield curve in a straightforward fashion, as will be illustrated below.

The problem with the bootstrap is that it relies heavily on the existence of a suitable body of data. In particular there are two problems: illiquidity and missing data points. Quotes coming from a thin market may be at divergence from true market prices due to spreads and asynchronous trading. This is why applying the bootstrap to a raw sample of bonds is likely to lead to an unreasonably 'choppy' yield curve. One approach is to perform some form of curve fitting to arrive at a reasonably smooth representation of the yield curve. This can be done either before or after utilizing the bootstrap. Nevertheless there is a problem with any smoothing procedure which is unavoidable: the final curve will always be a function of an assumed functional form. For example, the Nelson and Siegel (1987) procedure allows for a single hump while the Svensson (1994) approach allows for two humps.

There is another approach for eliminating inappropriate shapes, namely to make use of averaged yield data. For example, one could take all bonds in the neighbourhood of the five-year maturity, average their yields and then use the resultant average as the yield on a hypothetical five-year par bond in the belief that some of the bonds comprising the average will be discount bonds and others will be premium bonds. This, in fact, will be the approach that we will employ in some examples discussed later.

As for the second problem with the bootstrap — the lack of a full set of data — this must be solved by imposing conditions on the intermediate points. Of course, one can never avoid the arbitrariness of the choice of various interpolation procedures. We argue below that a cubic spline is an appropriate choice.

The purpose of this paper is to illustrate a new technique for estimating the yield curve which is useful in the presence of real-world data limitations. Our approach can be viewed as a kind of generalized bootstrap procedure. Using symbolic manipulation of algebraic expressions, it straightforwardly deals with any data set regardless of time spacing. Different interpolation assumptions can be accommodated. The illiquidity issue is dealt with by using highly liquid T-bills at the short end of the maturity and average yields beyond one year.

More specifically, beginning with a series of bond value expressions for a given set of fixed income securities, one cannot — except under ideal conditions — solve for the yields corresponding to all payment dates, since there are more unknowns than equations. Each maturity date is accounted for by a single bond value expression. Generating equations for the points corresponding to the

[2] When we say yield curve, this will refer to the spot or zero-coupon yield curve. An alternative useful for some purposes is the yield curve for bonds whose coupon rates are equivalent to their yields, that is the par bond yield curve.

[3] For such textbook explications, see, among others, Fabozzi and Fabozzi (1989) and Sundaresan (1997).

coupon dates, however, requires the use of an interpolation approach that involves the manipulation of symbolic quantities. By using a computer algebra system, such as Maple or Mathematica, it becomes possible to perform the symbolic interpolation and generate the required algebraic equations. Once a sufficient number of algebraic equations are obtained, the resultant system is solved numerically to obtain the points on the yield curve.

A key advantage of this approach is its simple 'one-shot' nature. Unlike the methods used by McCulloch (1971), Vasicek and Fong (1982) and Svensson (1994) who make *a priori* assumptions about the form of the yield curve, we let the available data determine the exact form of the yield curve by solving a system of nonlinear equations. It is worth stressing that this simpler — yet, general — method requires the processing of certain symbolic quantities to interpolate the points on the yield curve corresponding to coupon dates. Such symbolic processing is only possible by using sophisticated computer algebra systems such as the sort mentioned above.

In Section 2 the bootstrap method is described as presented in standard texts and simple example given. A mathematical description of the generalized bootstrap methodology is then provided and a a more complicated example involving symbolic interpolation and the solution of a nonlinear system of equations that could not be solved using the simple bootstrap method is discussed. In Section 3 three additional examples are provided using some recent Canadian bond data. Two appendices describe Maple's capabilities that are essential in symbolic interpolation and the numerical solution of the system of equations to obtain the yield curve. The paper concludes in Section 4 with a brief summary.

2. The bootstrap method and its generalization

2.1 *The textbook bootstrap*

The bootstrap method — as discussed in standard texts — is used to solve sequentially a system of *nonlinear* equations which has at least one equation whose solution for a single unknown can be obtained in a straightforward manner. In this section a brief review of the bootstrap method is provided by utilizing a numerical example.[4] The data on the maturity, coupons and prices of four bonds are given below:[5]

Bond number	Time to maturity (years)	Annual coupon (dollars)	Bond price (dollars)
1	0.50	0	94.9
2	1.00	0	90.0
3	1.50	8	96.0
4	2.00	12	101.6

[4] This example is similar to that of Hull (1997, p. 82). Also see the short expository article by Kitter (1998) that appeared in *Markets*, a trade magazine for financial professionals.

[5] Note that it is assumed that the bond principal is \$100 and half of the stated coupon is paid every six months.

Denoting the discount rate for a maturity of t years by r_t, it is easy to see that for $t = 0.50$ and $t = 1.00$ we must solve $94.9 = 100e^{-r_{0.50}\times 0.5}$ and $90 = 100e^{-r_{1.00}\times 1.0}$ to obtain $r_{0.50} = 0.1047$ and $r_{1.00} = 0.1054$, respectively.[6]

For the third bond which matures in 1.5 years there are three payments of \$4, \$4 and \$104 at $t = 0.50$, 1.00 and 1.50, respectively. Since the discount rates at $t = 0.50$ and 1.00 are already available from the previous calculations, the rate $r_{1.50}$ for $t = 1.50$ can be computed by solving $96 = 4e^{-0.1047\times 0.5} + 4e^{-0.1054\times 1.0} + 104e^{-r_{1.50}\times 1.5}$ for $r_{1.50}$ which gives $r_{1.50} = 0.1086$. In a similar fashion, it is straightforward to calculate $r_{2.00} = 0.1081$.

Note that the bootstrap succeeded because there were four equations and four unknown yields. What if in our example we had a 2.75-year bond (with coupon payments at 0.25, 0.75, 1.25, 1.75, 2.25); or what if the 0.50-year T-bill did not exist? Although the textbook bootstrap can no longer be applied in such a case, the generalized bootstrap method, which will be characterized next, can easily deal with such problems. The proposed method can also automate the task of symbolically generating the interpolation equations that are crucial in determining the yield curve.[7] Naturally, the generality of the method implies that simpler problems — easily amenable to the textbook bootstrap method — can also be solved in essentially one step instead of sequentially.

In the general model, the necessity to use natural spline interpolation requires the symbolic manipulation of certain quantities. The program used for this purpose is the computer algebra system Maple (Heal *et al.*, 1998), which automatically performs the symbolic computations necessary in the development of our model. In particular, we make use of Maple's `spline()` command to automate the task of generating cubic spline equations symbolically and `fsolve()` command to solve the general model involving nonlinear system of equations. More details are provided in Appendix A.

2.2 *Mathematical representation of the general method*

A particular bond can be characterized by its maturity, annual coupons, the time sequence of coupon payments and its price. Thus, if we have data on K bonds, the ith bond $\mathbf{B}_i$ can be represented using vector notation as

$$\mathbf{B}_i = [n_i, c_i, \mathbf{t}_i, p_i], \quad i = 1, \ldots, K$$

Here, n_i is the number of coupon payments, c_i is the annual coupon (in dollars), the time vector $\mathbf{t}_i = [t_{i,1}, \ldots, t_{i,n_i}]$ is the list of occurrence times of coupon payments (in years) with $t_{i,j}$ being the time of the jth coupon payment, and p_i is the bond price (in dollars). Note that t_{i,n_i} is the maturity of the ith bond and if $n_i = 1$, then the ith bond is zero-coupon implying $c_i = 0$. Conversely, if $c_i = 0$ for bond i, then $n_i = 1$ and the bond is zero-coupon. We assume throughout our discussion that the bonds are listed in increasing order of their maturity dates, i.e. $t_{1,n_1} < t_{2,n_2} < \cdots < t_{K,n_K}$.

Given the available information on the K bonds, we can write K nonlinear equations that relate

[6] Note that we will use continuously compounded yields in all cases.

[7] This methodology can also straightforwardly deal with extrapolation. This would, for example, be necessary if we had a 2.75-year bond, since the first cash flow would come before any of the known yields.

the price of a bond to other data. Defining r_j as the discount rate (to be determined) for the payment at the end of t_j time units (years), we have

$$p_i = \sum_{j=1}^{n_i-1} \tfrac{1}{2}c_i \exp(-r_j \cdot t_{i,j}) + (100 + \tfrac{1}{2}c_i)\exp(-r_{n_i} \cdot t_{i,n_i}), \quad i = 1, \ldots, K \tag{1}$$

In the unlikely case that the number of r_js is equal to the number of equations, this system can be solved to obtain the discount rates.[8] This, however, is an exception rather than the rule, and the number of unknowns generally exceeds the number of equations. Thus, we now examine the more general case where additional equations must be introduced (via symbolic cubic spline interpolation) in order to solve the nonlinear system of equations.

For each bond i, a set $\mathcal{T}_i = \{t_{i,1}, \ldots, t_{i,n_i}\}$ is defined that is obtained from the bond's time vector $\mathbf{t}_i = [t_{i,1}, \ldots, t_{i,n_i}]$. Thus, the union $\mathcal{T} = \bigcup_{i=1}^{K} \mathcal{T}_i = \{t_1, t_2, \ldots, t_N\}$ is the set of all *distinct* time points at which a coupon and/or face value payment is made for *some* bond. In the example discussed in Section 2.1, we have $\mathcal{T}_1 = \{0.50\}$, $\mathcal{T}_2 = \{1.00\}$, $\mathcal{T}_3 = \{0.50, 1.00, 1.50\}$, and $\mathcal{T}_4 = \{0.50, 1.00, 1.50, 2.00\}$. Thus $\mathcal{T} = \{0.50, 1.00, 1.50, 2.00\}$. The cardinality of $\mathcal{T}$ is $|\mathcal{T}| = N$ which corresponds to the total number of distinct times at which coupons and/or face value are paid. In the above example, $N = 4$.

The smallest value of all the elements in $\mathcal{T}$ is denoted by $t_1 = \min\{t_{1,1}, t_{2,1}, \ldots, t_{K,1}\}$ and the largest value in $\mathcal{T}$ by $t_N = t_{K,n_K}$ so that the elements of the set $\mathcal{T}$ can be listed in an increasing order as $[t_1, t_2, \ldots, t_N]$. Clearly, any element $t_{i,j}$ of the vector $\mathbf{t}_i$ for bond i defined above can be found in the set $\mathcal{T} = \{t_1, t_2, \ldots, t_N\}$. In the example discussed in Section 2.1, we have $t_1 = 0.50$ and $t_4 = 2.00$.

When $K < N$ the system of K nonlinear equations given in Equation 1 is 'underdetermined' since there are fewer equations than there are unknowns. In such a case, additional equations can be introduced by interpolating some of the intermediate maturities.

Based on the discussion above, it can be seen that many of the elements in the set $\mathcal{T}$ correspond to the maturities of the K different bonds, namely, $t_{1,n_1}, t_{2,n_2}, \ldots, t_{K,n_K}$. The set of maturities of the K bonds is denoted by $\mathcal{K} = \{t_{1,n_1}, t_{2,n_2}, \ldots, t_{K,n_K}\}$ with cardinality $|\mathcal{K}| = K$. Naturally, for each of these time points we have exactly one nonlinear equation as given in Equation 1. For the remaining $L = N - K$ maturities we can develop $L = N - K$ additional equations using some form of natural spline interpolation as will be seen below. This would bring the total number of equations to $K + (N - K) = N$ which matches the total number N of unknowns. The set of those points for which interpolation is necessary is denoted by $\mathcal{L} = \{t_j : t_j \in \mathcal{T} \text{ and } t_j \notin \mathcal{K}\}$. In the example of Section 2.1, we would have $\mathcal{K} = \{0.50, 1.00, 1.50, 2.00\}$ and $\mathcal{L} = \varnothing$, the empty set. Solving the resulting N equations in N unknowns would produce the points r_j on the yield curve for each maturity t_j.

[8] In fact, for the simple example discussed in Section 2.1, using Equation 1 we can compute the yield r_{n_i} for the ith bond ($i = 1, \ldots, K$) as

$$r_{n_i} = \frac{1}{t_{i,n_i}} \ln\left(\frac{100 + \frac{1}{2}c_i}{p_i - \sum_{j=1}^{n_i-1} \frac{1}{2}c_i \exp(-r_j \times t_{i,j})}\right)$$

Of course, this formula implicitly assumes that all previous yields have been obtained in a sequential manner.

2.3. *An example of the solution of the general model*

In this section an extension of the previous example where the textbook bootstrap method fails to solve the problem is described. In this case $K = 5$ and the bonds have the following characteristics:

Bond $\mathbf{B}_i$	*Number of coupons* n_i	*Annual coupon* c_i	*Coupon payment times* $\mathbf{t}_i$	*Price* p_i
1	1	0	[0.25]	97.5
2	2	2	[0.50, 1.00]	90.0
3	3	8	[0.50, 1.00, 1.50]	96.0
4	4	12	[0.50, 1.00, 1.50, 2.00]	101.6
5	6	10	[0.25, 0.75, 1.25, 1.75, 2.25, 2.75]	99.8

The point on the zero-coupon curve for $t_1 = 0.25$ can be easily computed as $r_{0.25} = 0.1013$ using the information on the first bond. This information, however, would not be useful in determining another point on the curve since substituting this value into the equation for bond number 5 (which has a coupon payment at $t_1 = 0.25$) could not produce a solution for any other point. In this case, the textbook bootstrap fails, so we use our general method to determine all yields.

The general model is now developed in terms of nonlinear equations and the interpolated equations (which prove to be linear). Solving the resultant system of equations using Maple enables one to find all the points on the yield curve for the time points given in the table above.

In this example we have $\mathscr{K} = \{0.25, 1.00, 1.50, 2.00, 2.75\}$ with $K = 5$ elements as the set of maturities. Using Maple's spline() command, we can easily fit a cubic spline to these points for the unknown values on the yield curve. This gives

$$CS(t) = \begin{cases} -0.05r_{2.00} - 0.61r_{1.00} + 1.43r_{0.25} + 0.01r_{2.75} + 0.23r_{1.50} \\ +(0.15r_{2.00} + 2.16r_{1.00} - 1.62r_{0.25} - 0.01r_{2.75} - 0.68r_{1.50})t \\ +(0.31r_{2.00} + 1.66r_{1.00} - 0.57r_{0.25} - 0.03r_{2.75} - 1.37r_{1.50})t^2 \\ +(-0.41r_{2.00} - 2.21r_{1.00} + 0.76r_{0.25} + 0.04r_{2.75} + 1.82r_{1.50})t^3, & \text{if } t \leqslant 1 \\ -4.02r_{2.00} - 9.14r_{1.00} + 3.64r_{0.25} + 0.39r_{2.75} + 10.1r_{1.50} \\ +(12.1r_{2.00} + 27.8r_{1.00} - 8.25r_{0.25} - 1.18r_{2.75} - 30.4r_{1.50})t \\ +(-11.6r_{2.00} - 23.9r_{1.00} + 6.07r_{0.25} + 1.13r_{2.75} + 28.3r_{1.50})t^2 \\ +(3.56r_{2.00} + 6.32r_{1.00} - 1.45r_{0.25} - 0.34r_{2.75} - 8.08r_{1.50})t^3, & \text{if } 1 \leqslant t \leqslant 1.5 \\ 29r_{2.00} + 24.3r_{1.00} - 2.52r_{0.25} - 5.58r_{2.75} - 44.2r_{1.50} \\ +(-53.8r_{2.00} - 39.1r_{1.00} + 4.05r_{0.25} + 10.7r_{2.75} + 78.1r_{1.50})t \\ +(32.3r_{2.00} + 20.6r_{1.00} - 2.14r_{0.25} - 6.82r_{2.75} - 44r_{1.50})t^2 \\ +(-6.20r_{2.00} - 3.58r_{1.00} + 0.371r_{0.25} + 1.42r_{2.75} + 8r_{1.50})t^3, & \text{if } 1.5 \leqslant t \leqslant 2 \\ -38r_{2.00} - 7.44r_{1.00} + 0.77r_{0.25} + 11.8r_{2.75} + 33.9r_{1.50} \\ +(46.7r_{2.00} + 8.57r_{1.00} - 0.88r_{0.25} - 15.3r_{2.75} - 39r_{1.50})t \\ +(-17.9r_{2.00} - 3.2r_{1.00} + 0.33r_{0.25} + 6.23r_{2.75} + 14.6r_{1.50})t^2 \\ +(2.17r_{2.00} + 0.39r_{1.00} - 0.04r_{0.25} - 0.75r_{2.75} - 1.77r_{1.50})t^3, & \text{if } 2 \leqslant t \leqslant 2.75 \end{cases}$$

where the $CS(t)$ function provides the cubic spline interpolated value at any time t. Cubic spline interpolation is chosen for the unknown points on the yield curve since it provides sufficient flexibility to ensure that the interpolant is continuously differentiable and it has continuous second derivatives. For details of this method, see Burden and Faires (1985, Section 3.6). Appendix A presents the computations of this interpolation using Maple.

Now armed with this cubic spline approximation of the yield curve, linear equations can easily be developed for all points $t_j \in \mathcal{X} = \{0.50, 0.75, 1.25, 1.75, 2.25\}$ that do not correspond to the maturities of the bonds. This gives (to two significant digits):

$$r_{0.50} = CS(0.50) = 0.05r_{2.00} + 0.61r_{1.00} + 0.58r_{0.25} - 0.005r_{2.75} - 0.22r_{1.50} \quad (2)$$

$$r_{0.75} = CS(0.75) = 0.06r_{2.00} + 1.01r_{1.00} + 0.21r_{0.25} - 0.006r_{2.75} - 0.28r_{1.50} \quad (3)$$

$$r_{1.25} = CS(1.25) = -0.11r_{2.00} + 0.51r_{1.00} - 0.04r_{0.25} - 0.01r_{2.75} + 0.6r_{1.50} \quad (4)$$

$$r_{1.75} = CS(1.75) = 0.51r_{2.00} - 0.11r_{1.00} + 0.01r_{0.25} - 0.04r_{2.75} + 0.62r_{1.50} \quad (5)$$

$$r_{2.25} = CS(2.25) = 1.01r_{2.00} + 0.06r_{1.00} - 0.01r_{0.25} + 0.21r_{2.75} - 0.28r_{1.50} \quad (6)$$

Note that by using cubic spline interpolation any point on the yield curve can be expressed as a *linear combination* of all other points corresponding to the maturities. For example, $r_{2.25}$ is obtained as a linear combination of $r_{0.25}$, $r_{1.00}$, $r_{1.50}$, $r_{2.00}$ and $r_{2.75}$. If simpler linear interpolation had been used, $r_{2.25}$ would be obtained as a convex combination of the nearest two neighbouring points $r_{2.00}$ and $r_{2.75}$, in particular $r_{2.25} = 0.67r_{2.00} + 0.33r_{2.75}$.

Next, the $K = 5$ nonlinear equations arising from Equation 1 are written:

$$97.5 = 100e^{-r_{0.25}\times 0.25} \quad (7)$$

$$90.0 = 1 \cdot e^{-r_{0.50}\times 0.50} + 101e^{-r_{1.00}\times 1.0} \quad (8)$$

$$96 = 4e^{-r_{0.50}\times 0.50} + 4e^{-r_{1.00}\times 1.0} + 104e^{-r_{1.50}\times 1.50} \quad (9)$$

$$101.6 = 6e^{-r_{0.50}\times 0.50} + 6e^{-r_{1.00}\times 1.0} + 6e^{-r_{1.50}\times 1.5} + 106e^{-r_{2.00}\times 2.0} \quad (10)$$

$$99.8 = 5e^{-r_{0.25}\times 0.25} + 5e^{-r_{0.75}\times 0.75} + 5e^{-r_{1.25}\times 1.25} + 5e^{-r_{1.75}\times 1.75} + 5e^{-r_{2.25}\times 2.25} + 105e^{-r_{2.75}\times 2.75} \quad (11)$$

Thus, $L + K = 10$ equations in 10 unknowns are obtained. Note that the problem can be further simplified by substituting the linear expressions for $r_{0.50}$, $r_{0.75}$, etc., into the nonlinear equations (7)–(11) and solving for only the five unknowns $r_{0.25}$, $r_{1.00}$, $r_{1.50}$, $r_{2.00}$ and $r_{2.75}$. Once these quantities are obtained, they are re-substituted in $r_{0.50}$, $r_{0.75}$, etc., to obtain the complete solution. Solving the system (2)–(11) using Maple's `fsolve()` function yields the following results, where the numbers in **bold** correspond to the final maturities of the five bonds: (For details of Maple's `fsolve()` function that can solve systems of nonlinear equations, see Appendix B.)

t_j	**0.25**	0.50	0.75	**1.00**	1.25	**1.50**	1.75	**2.00**	2.25	**2.75**
r_j (in %)	10.13	11.53	12.50	12.59	11.61	10.61	10.49	10.73	10.92	10.83

3. Examples using Canadian bond data

The examples described above involved hypothetical bonds. We now present three other examples employing Canadian fixed income market data. In particular, we have data on four Canadian T-bills and four synthetic par bonds, whose yields are calculated as averages over all bonds in the neighbourhood of the relevant maturity.[9] We obtain yield curves for three months over the last decade: February 1990, December 1993 and April 1998. These different dates were chosen for variety of yield curve slope.

The data provided in Table 1 are used to estimate the yield curve for February 1990. The approach discussed in Section 2.3 is implemented to generate the system of linear and nonlinear equations. The solution of this system in terms of the unknowns r_j gives the points on the yield curve for this month.

In this problem we have $\mathscr{K} = \{0.08, 0.25, 0.50, 1.00, 2.00, 4.00, 7.50, 18.00\}$ as the maturities of four T-bills and four par bonds. That is, the cardinality of this set is $|\mathscr{K}| = K = 8$. The set of all distinct time points at which a coupon and/or face value payment is made for some bond is found as $\mathscr{T} = \{0.08, 0.25, 0.50, 1.00, 1.50, \ldots, 17.00, 17.50, 18.00\}$. This set has cardinality $|\mathscr{T}| = N = 38$. The set of time points for which we perform cubic spline interpolation is $\mathscr{S} = \{1.50, 2.50, 3.00, \ldots, 17.00, 17.50\}$ with a cardinality of $|\mathscr{S}| = L = 30$. For example, at $t = 8.50$, we obtain

$$r_{8.50} = -0.07r_{0.25} + 0.02r_{0.08} - 0.18r_{1.00} + 0.13r_{0.50} - 0.41r_{4.00} + 0.26r_{2.00} + 1.22r_{7.50} + 0.03r_{18.00}$$

Table 1. T-bill and par bond data for February 1990

T-bill/Par bond $\mathbf{B}_i$	*Number of coupons* n_i	*Annual coupon* c_i	*Coupon times* $\mathbf{t}_i$	*Yields* r_{n_i}	*Price* p_i
1 (1 m)	1	0	[0.08]	12.93	98.93
2 (3 m)	1	0	[0.25]	13.15	96.81
3 (6 m)	1	0	[0.50]	13.04	93.87
4 (1 yr)	1	0	[1.00]	12.78	88.66
5 (2 yrs)	4	12.05	[0.50, 1.00, ..., 2.00]		100
6 (4 yrs)	8	11.45	[0.50, 1.00, ..., 4.00]		100
7 ($7\frac{1}{2}$ yrs)	15	10.74	[0.50, 1.00, ..., 7.50]		100
8 (18 yrs)	36	10.64	[0.50, 1.00, ..., 18.00]		100

[9] These data were obtained from various issues of the *Bank of Canada Review*, Table F1.

as the linear equation corresponding to $r_{8.50}$. Thus, we have a system of $N = 38$ equations (8 nonlinear and 30 linear) and 38 unknowns r_j with $j \in \mathscr{T}$. After substituting the linear equations into the nonlinear equations, we effectively reduce the problem to the solution of $K = 8$ nonlinear equations in 8 unknowns. Note that these substitutions and all the necessary simplifications are performed automatically by Maple.

The T-bill yields r_{n_i} are already available from published data, and these are listed in the fifth column of the table. To establish equations for these zero-coupon instruments, we compute the implied prices p_i (shown in the last column of the table).[10] For the par bonds, the yields at maturity are the same as the annual coupons.

Solving this system using Maple, we obtain the points on the yield curve as a function of time t for February 1990, which is shown as the solid curve in Fig. 1.[11] Note that in this example the yield curve is inverted since the yields are higher for shorter maturities. We also plot the forward rate curve (implied by the yield curve) as the dashed line in the same figure.[12]

In the second example we consider the December 1993 bond data as presented in Table 2. The resultant yield curve shown in Fig. 2 is now upward sloping. Finally, Table 3 and Fig. 3 correspond to April 1998. Once again, the yield curve is an increasing function of time, though, with an economic slowdown anticipated by the market, the yield curve was close to flat.

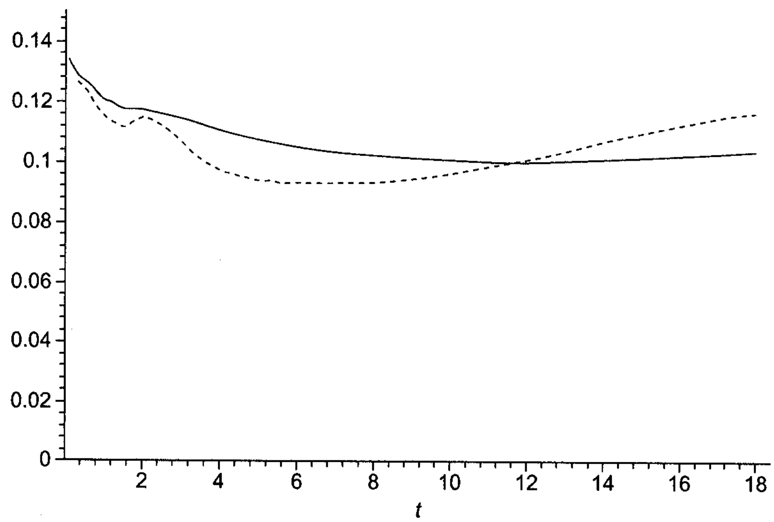

Fig. 1. Yield curve (solid) and forward rate curve (dashed) as a function of time t for February 1990 bond data.

[10] This is done using the formula $p_i = 100/(1 + r_{n_i}/t_{i,n_i})$.

[11] All Maple programs used in this paper are available from the authors, if requested.

[12] The forward rates f_{t_1,t_2} for any two years t_1 and t_2 with $t_1 < t_2$ can be obtained in a straightforward manner using the well-known formula

$$f_{t_1,t_2} = \frac{r_{t_2} \times t_2 - r_{t_1} \times t_1}{t_2 - t_1}$$

Table 2. T-bill and par bond data for December 1993

T-bill/Par bond $\mathbf{B}_i$	*Number of coupons* n_i	*Annual coupon* c_i	*Coupon times* $\mathbf{t}_i$	*Yields* r_{n_i}	*Price* p_i
1 (1 m)	1	0	[0.08]	3.60	99.70
2 (3 m)	1	0	[0.25]	3.87	99.04
3 (6 m)	1	0	[0.50]	4.04	98.01
4 (1 yr)	1	0	[1.00]	4.23	95.94
5 (2 yrs)	4	4.57	[0.50, 1.00, ..., 2.00]		100
6 (4 yrs)	8	5.47	[0.50, 1.00, ..., 4.00]		100
7 ($7\frac{1}{2}$ yrs)	15	6.33	[0.50, 1.00, ..., 7.50]		100
8 (18 yrs)	36	7.12	[0.50, 1.00, ..., 18.00]		100

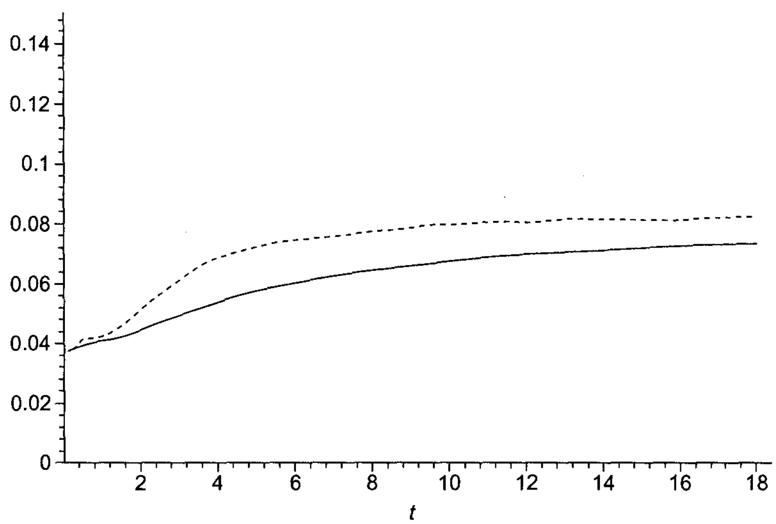

Fig. 2. Yield curve (solid) and forward rate curve (dashed) as a function of time t for December 1993 bond data.

As is necessary for any reasonable yield curve estimation methodology, all forward rate curves exhibit a fair degree of smoothness. This is in contrast to some other methodologies, e.g. McCulloch (1971), where forward rate curves exhibit 'knuckles' and undesirable asymptotic properties.

4. Summary and conclusions

This paper has described a new technique for operationalizing the bootstrap methodology for estimating the yield curve given any available data set of bond yields. The problems of illiquidity and missing data points must always be confronted. Illiquidity can be tackled by using some sort

Table 3. T-bill and par bond data for April 1998

T-bill/Par bond $\mathbf{B}_i$	*Number of coupons* n_i	*Annual coupon* c_i	*Coupon times* $\mathbf{t}_i$	*Yields* r_{n_i}	*Price* p_i
1 (1 m)	1	0	[0.08]	4.37	99.63
2 (3 m)	1	0	[0.25]	4.54	98.87
3 (6 m)	1	0	[0.50]	4.67	97.71
4 (1 yr)	1	0	[1.00]	4.89	95.33
5 (2 yrs)	4	4.99	[0.50, 1.00, ..., 2.00]		100
6 (4 yrs)	8	5.11	[0.50, 1.00, ..., 4.00]		100
7 ($7\frac{1}{2}$ yrs)	15	5.24	[0.50, 1.00, ..., 7.50]		100
8 (18 yrs)	36	5.50	[0.50, 1.00, ..., 18.00]		100

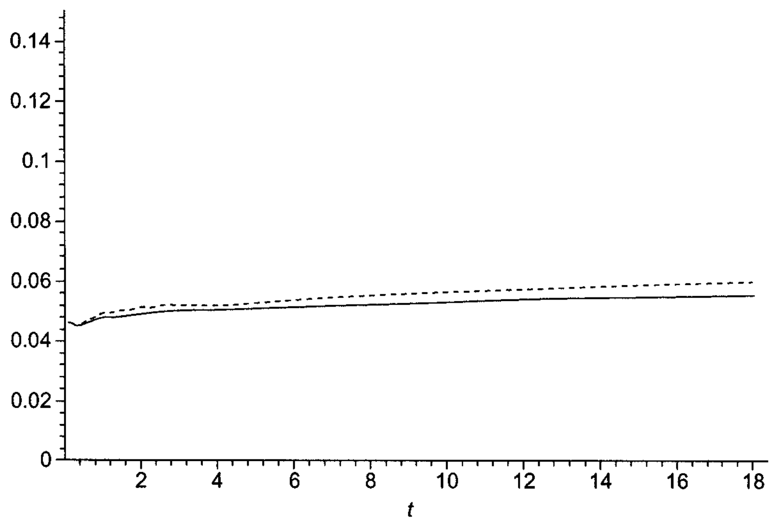

Fig. 3. Yield curve (solid) and forward rate curve (dashed) as a function of time t for April 1998 bond data.

of smoothing technique (either before or after the bootstrap) or averaged yield data. In the illustration of the proposed methodology using Canadian bond data averaged yields were employed. Missing data points were dealt with by the use of spline interpolation. More specifically, the cubic spline was selected as it provides sufficient flexibility to ensure that the interpolant is continuously differentiable and has continuous second derivatives. The computer algebra system Maple was used to symbolically generate the interpolation equations for the missing data points and to solve the nonlinear equation system in order to obtain the points on the yield curve.

The three examples presented here demonstrate the usefulness of the methodology. The yield curves resulting from the estimation appear to have reasonable shapes, and in addition the forward rate curves are clearly well-behaved as well. Finally, another strength of the methodology is its ease of implementation. Of course, given the importance of ascertaining the correct term structure for a variety of purposes, further research on this issue is desirable.

References

Bradley, S.P. and Crane, D.B (1973) Management of commercial bank government security portfolio: An optimization approach under uncertainty, *Journal of Bank Research*, **4**, 18–30.

Burden, R.L. and Faires, J.D. (1985) *Numerical Analysis*, 3rd edn. Prindle, Weber and Schmidt, Boston.

Deaves, R. (1997) Predictable excess fixed-income returns: The Canadian case, *Journal of Fixed Income*, **7**, September, 67–76.

Echols, M.E. and Elliott, J.W. (1976) A quantitative yield curve model for estimating the term structure of interest rates. pp. 87–114, March.

Fabozzi, F.J. and Fabozzi, T.D. (1989) *Bond Markets, Analysis and Strategies*. Prentice Hall, Englewood Cliffs, NJ.

Harvey, C.R. (1991) The term structure and world economic growth, *Journal of Fixed Income*, **1**, June, 7–20.

Heal, K.M., Hansen, M.L. and Rickard, K.M. (1998) *Maple V Learning Guide*. Springer, New York.

Hull, J.C. (1997) *Options, Futures and Other Derivatives*. Prentice Hall, Upper Saddle River, NJ.

Jarrow, R.A. (1996) *Modeling Fixed Income Securities and Interest Rate Options*. McGraw-Hill, New York.

Kitter, G. (1971) Riding the zero yield curve, *Markets*, **3**, March/April, 40–41.

McCulloch, J.H. (1971) Measuring the term structure of interest rates, *Journal of Business*, **44**, 19–31.

Nelson, C.R. and Siegel, A.F. (1987) Parsimonious modeling of yield curves, *Journal of Business*, **60**, 473–489.

Sundaresan, S. (1997) *Fixed Income Markets and Their Derivatives*. South-Western College Publishing, Cincinnati.

Svensson, L.E.O. (1994) Estimating and interpreting forward interest rates: Sweden 1992–1994. Technical Report Discussion Paper Number 1051, Centre for Economic Policy Research, October.

Vasicek, O.A. and Fong, H.G. (1982) Term structure modeling using exponential splines, *Journal of Finance*, **37**(2), 329–348.

Appendix A Maple's treatment of interpolation using the spline( ) function

Additional equations for the set $\mathscr{L}$ can be generated using natural spline interpolation. Assuming that $\mathbf{x} = [x_1, \ldots, x_K]$ is the independent variable vector with $x_1 < x_2 < \cdots < x_K$, and $\mathbf{y} = [y_1, \ldots, y_K]$ is the dependent variable vector, natural spline interpolation computes a piecewise polynomial approximation of degree d to the $(\mathbf{x}, \mathbf{y})$ data. The use of the computer algebra system Maple to automate the computations involved in natural spline interpolation are now demonstrated.

For $d = 1$, natural spline interpolation reduces to the well-known linear interpolation. For example, when $\mathbf{x} = [1, 2, 3]$ and $\mathbf{y} = [3, 1, 2]$ we enter the following Maple commands to load the spline( ) library and compute the linear interpolation function $LI(x)$.

```
> readlib(spline);
```

$$\mathbf{proc}(X, Y, z, d) \ldots \mathbf{end}$$

> LI:=spline([1,2,3],[3,1,2],x,linear);

$$LI := \begin{cases} 5 - 2x & x < 2 \\ -1 + x & otherwise \end{cases}$$

Since the independent variable x varies between 1 and 3, Maple's result implies that $LI(x)$ is

$$LI(x) = \begin{cases} 5 - 2x, & \text{for } 1 \leqslant x \leqslant 2 \\ -1 + x, & \text{for } 2 \leqslant x \leqslant 3 \end{cases}$$

To compute the interpolated values of y for any value of x, we need to convert Maple's result for LI to a function of x using unapply().

> LI:=unapply(spline([1,2,3],[3,1,2],x,linear),x);

$$LI := x \rightarrow \text{piecewise}(x < 2,\ 5 - 2x,\ -1 + x)$$

Thus, the linear interpolation of the data at, say, $x = 1.6$, gives $LI(1.6) = 1.8$.

> LI(1.6);

$$1.8$$

When the dependent variable vector $\mathbf{y} = [y_1, \ldots, y_K]$ is not given numerically (as is the case for the generalized algebraic approach we are using in this paper), then Maple becomes indispensable for the required symbolic computations of the interpolation. For $d = 1$ and $\mathbf{x} = [1, 2, 3]$ as above, we assume that $\mathbf{y} = [y_1, y_2, y_3]$ is the symbolic vector for the dependent variable. Applying spline( ) to these $(\mathbf{x}, \mathbf{y})$ data gives

> LI:=spline([1,2,3],[y[1],y[2],y[3]],x,linear);

$$LI := \begin{cases} 2y_1 - y_2 + (-y_1 + y_2)x & x < 2 \\ 3y_2 - 2y_3 + (-y_2 + y_3)x & otherwise \end{cases}$$

> LI:=unapply(spline([1,2,3],[y[1],y[2],y[3]],x,linear),x);

$$LI := x \rightarrow \text{piecewise}(x < 2,\ 2y_1 - y_2 + (-y_1 + y_2)x,\ 3y_2 - 2y_3 + (-y_2 + y_3)x)$$

> LI(1.6);

$$0.4y_1 + 0.6y_2$$

These results imply that

$$LI(x) = \begin{cases} 2y_1 - y_2 + (-y_1 + y_2)x, & \text{for } 1 \leqslant x \leqslant 2 \\ 3y_2 - 2y_3 + (-y_2 + y_3)x, & \text{for } 2 \leqslant x \leqslant 3 \end{cases}$$

Thus, at $x = 1.6$, we would obtain $y_{1.6} = LI(1.6) = 0.4y_1 + 0.6y_2$ since $1 < x = 1.6 < 2$. Note that the interpolated value $y_{1.6}$ at $x = 1.6$ is a *convex combination* of the nearest neighbouring points.

A much better interpolation approach that provides a smooth (i.e. differentiable) fit to the available data points is the cubic spline interpolation when the degree of the piecewise polynomial approximation is $d = 3$. For details of this approach, see Burden and Faires (1985, Section 3.6). When the data points are given numerically as $\mathbf{x} = [1, 2, 3]$ and $\mathbf{y} = [3, 1, 2]$, the Maple command

> CSI:=spline([1,2,3],[3,1,2],x,cubic);

produces the following cubic spline interpolation function $CSI(x)$:

$$CSI(x) = \begin{cases} 5 - \frac{1}{2}x - \frac{9}{4}x^2 + \frac{3}{4}x^3, & \text{for } 1 \leqslant x \leqslant 2 \\ 17 - \frac{37}{2}x + \frac{27}{4}x^2 - \frac{3}{4}x^3, & \text{for } 2 \leqslant x \leqslant 3 \end{cases}$$

Thus, if we wish to interpolate the data using cubic splines at $x = 1.6$, we would obtain $CSI(1.6) = 1.512$.

When the dependent variable vector $\mathbf{y} = [y_1, \ldots, y_K]$ is not given numerically, i.e. when $\mathbf{x} = [1, 2, 3]$ as above but $\mathbf{y} = [y_1, y_2, y_3]$, the symbolic cubic spline interpolation

```
> CSI:=spline([1,2,3],[y[1],y[2],y[3]],x,cubic);
```

is obtained as

$$CSI(x) = \begin{cases} -y_2 + 2y_1 + (-\frac{1}{2}y_1 + \frac{1}{2}y_3)x \\ +(\frac{3}{2}y_2 - \frac{3}{4}y_1 - \frac{3}{4}y_3)x^2 + (-\frac{1}{2}y_2 + \frac{1}{4}y_1 + \frac{1}{4}y_3)x^3, & \text{for } 1 \leqslant x \leqslant 2 \\ -9y_2 + 6y_1 + 4y_3 + (12y_2 - \frac{13}{2}y_1 - \frac{11}{2}y_3)x \\ +(-\frac{9}{2}y_2 + \frac{9}{4}y_1 + \frac{9}{4}y_3)x^2 + (\frac{1}{2}y_2 - \frac{1}{4}y_1 - \frac{1}{4}y_3)x^3, & \text{for } 2 \leqslant x \leqslant 3 \end{cases}$$

Thus, at $x = 1.6$, we find

```
> CSI(1.6);
```

$$0.3040000000y_1 + 0.792000000y_2 - 0.096000000y_3$$

i.e., $y_{1.6} = CSI(1.6) = 0.304y_1 + 0.792y_2 - 0.096y_3$. Note that the interpolated value $y_{1.6}$ at $x = 1.6$ is no longer a convex combination of its nearest neighbouring points y_1 and y_2, but it is a *linear combination* of all three points y_1, y_2 and y_3. This result has the important implication that even with cubic spline interpolation, for a given value of the independent variable x, the point on the interpolated curve is obtained as a linear function of the symbolic data points.

Appendix B Maple's solution of a system of nonlinear equations

Maple's fsolve( ) function is a very powerful solver that can be used to find the roots of a system of nonlinear equations. As an example, consider the problem where we wish to find the positive solution of the nonlinear system of equations $x^2 + y^2 = 1$ and $y = e^{-x}$. The following Maple statements define the two equations and fsolve() is used to compute the required root.

```
> eq1:=x^2+y^2=1;
```

$$eq1 := x^2 + y^2 = 1$$

```
> eq2:=y=exp(-x);
```

$$eq2 := y = e^{(-x)}$$

```
> fsolve({eq1,eq2},{x,y},x=0.001 .. 1);
```

$$\{x = 0.9165625831,\ y = 0.3998912743\}$$

It is worth noting that the user can instruct Maple to limit the search to an interval of x (and y) values such as x=0.001 ... 1.