LTA 2/98 • P. 159–177

JARI KÄPPI[1]

Estimating A Smooth Term Structure of Interest Rates

ABSTRACT

This paper extends the literature of the term structure estimation with splines. We fit the term structure of interest rates with a smoothing spline method that uses a different smoothing norm and locates the knot points by the size of the fitting errors. The method is applied to the Finnish fixed income market and compared to the usual smoothing spline methods and to the equally spaced knot locations. The results show that the new method where the spline is placed on the log of the discount function and the knots are located freely outperforms the other methods.

Keywords: *Term Structure of Interest Rates, Yield Curve, Smoothing Splines, Generalized Cross Validation.*

I. INTRODUCTION

The term structure of interest rates represents the yields to maturity of zero-coupon bonds as a function of time to maturity. It can be presented by any of the following ways; using the discount function, the zero-coupon interest rates, or the forward rates. The yield curve is the ba-

1 I am grateful to Vesa Puttonen, Peter Honoré, Stefan Jaschke, Miikka Tauren and two anonymous referees for helpful comments and suggestions. I am also especially indebted to Darrell Duffie for numerous comments and suggestions.

JARI KÄPPI, Ph.D, Associate Professor

Helsinki School of Economics and Business Administration • e-mail: kappi@hkkk.fi

sic tool in fixed income markets. It provides a framework for active bond portfolio management (Ilmanen (1995)), where forward rates are used as break-even rates for expected bond return analysis. The Value-at-Risk applications need good estimates of the yield curve to map the cash flows from fixed income instruments in order to estimate the risks in the portfolios. The fast growing market for fixed income derivatives instruments needs a term structure model for pricing purposes. For example, the implementation of Hull and White (1996) interest rate trees requires an estimate of a smooth term structure function in order to calibrate the tree to fit the initial time bond prices. These are only a few examples where a smooth term structure of interest rates plays a crucial role.

The estimation of the term structure of interest rates is usually done by parsimonious parameterization of the yield curve, e.g. Nelson and Siegel (1987), or by spline-based methods. The spline-based methods, pioneered by McCulloch (1971, 1975) and extended by Vasicek and Fong (1982), Coleman, Fisher and Ibbotson (1992), Adams and Van Deventer (1994), and Fisher, Nychka and Zervos (1995), among others, have received a lot of attention lately, and for example Fisher *et al.* (1995) claim that their spline method produces smaller pricing errors than the Nelson and Siegel model. We also estimated the yield curves using the Nelson-Siegel method and found that the pricing errors were several times larger than the pricing errors from a simple spline estimation method. Partly, this can be due to the small number of instruments available.

The spline-based methods can be divided into interpolating, least-squares and smoothing methods. When the interpolating spline is used, the pricing errors are the smallest possible but the term structure of the interest rates will not necessarily be the smoothest, because the interpolating spline function picks all the noise from the data (i.e. it overfits the data). Most of the current literature has used the least-squares approach to the fitting of the term structure where they use only a subset of data points, the maturities of the instruments, as knots. For example, McCulloch (1975) has presented that the number of knot points can be selected as the square root of the number of instruments and they should be located such that there is an equal number of instruments between the knot points. However, the selection of the knot points is less trivial and most of the time it is more or less a trial and error process. The empirical results show, see e.g. Fisher *et al.* (1995), that when the number of knots is increased the pricing errors of the different spline functions also change and another spline function may have smaller pricing errors with different knot sets. The shape of the splined function affects not only the number of knots but also the smoothness of the curve.

The smoothing splines try to combine the two different approximation objectives, to fit a smooth curve to the data and to simultaneously keep the pricing errors as small as possible. However, these two properties are often contradictory, and a compromise between the two properties needs to be found. Fisher *et al.* (1995) have proposed a method where the residual errors are minimized with a penalizing roughness function. Even though their simulation results show that the smoothing spline method gives better estimation results, as should be expected, among the compared spline fitting methods their model uses a subset of the maturities of the instruments as knot points.[2]

The objective of this paper is to extend the literature of the spline-based estimation methods by applying a smoothing spline method presented by Dierckx (1975, 1981, 1982) to the fitting of the term structure of interest rates.[3] The model deviates from other smoothing spline methods by using a different smoothing norm, the square of the discontinuity jump in the third derivatives at the interior knot points, and by locating the internal knot points by the size of the fitting errors. The proposed method uses the generalized cross validation to detect the smoothing parameter.

We apply the estimation method to the Finnish fixed income markets, where the number of instruments is much less than the number of cash flow dates. On this kind of market the estimation of the term structure of interest rates is very difficult without a spline-based method. However, the proposed estimation method will work on all kinds of markets. The proposed method is compared to other smoothing spline methods and also to the equally spaced knot positions. Our results show that the new method where the spline is placed on the log of the discount function and the knots are located by the size of errors outperforms the other fitting methods.

The rest of the paper is organized as follows. Section 2 reviews the theory of curve fitting with splines. Section 3 presents the term structure concepts and the estimation model, section 4 illustrates the empirical results on the Finnish market, and section 5 summarizes the paper.

II. SPLINE FUNCTIONS

Before applying spline functions to the estimation of the term structure we first discuss briefly the basic concepts, definitions and properties of the spline functions. We present the numerically stable *B*-spline basis, which is the most often used spline representation form, and discuss the different smoothing norms. These basic properties are important to understand when using splines.

2 Although Fisher *et al.* (1995) maintain that their model lets the data determine the "effective number of parameters", they fix the knot positions beforehand and use the GCV method to smooth the spline. As they use one third of the total number of the data points as knots, we do not know whether another combination of knot points will give a better smoothing spline or not. At least the value of the penalty function will be different with a different combination of the knots as it is completely determined by the knot positions.

3 Dierckx (1995) gives a good description of smoothing spline functions and presents the algorithms and the proofs.

Basic properties

Splines are used in different kinds of data fitting problems to approximate an unknown function. The increased interest in splines can be explained by their flexibility and easy usage as opposed to polynomials. In addition, they can also be used both for interpolation and smoothing. The following definition illustrates the basic properties of the spline functions.

Definition 1. [Dierckx (1995)] *A function s(x), defined on a finite interval [a, b], is called a spline function of degree k > 0, having as knots the strictly increasing sequence* λ_i, $i = 0,\dots,g+1$ $(\lambda_0 = a, (\lambda_{g+1} = b)$, *if the following two conditions are satisfied:*

1. *On each knot interval* $[\lambda_i, \lambda_{i+1}]$, *s(x) is given by a polynomial of degree k at most.*
2. *The function s(x) and its derivatives up to order k-1 are all continuous on [a,b].*

The definition shows that splines are more flexible than polynomials and they guarantee the continuity of the function and its derivatives up to the *k-1* degree. If the definition is relaxed and coincident knots, say *r*, are allowed, the continuity conditions in derivatives decrease to *k-1-r* at the given point. Even though the degree of the splines is not restricted, the most often used degree of the spline function is cubic because it is easy to implement and it oscillates less than higher degree splines.[4]

The splines can be represented in several different forms, but most of the time they are represented on a *B*-spline basis, which is a numerically stable form. The following recurrence relation defines the cubic *B*-spline basis on [a, b]

$$(1) \qquad N_{i,r+1}(x) = \frac{x-\lambda_i}{\lambda_{i+1}-\lambda_i} N_{i,r}(x) + \frac{\lambda_{i+r+1}-x}{\lambda_{1+r+1}-\lambda_{i+1}} N_{i+1,r}(x),$$

$$N_{i,1}(x) = \begin{cases} 1, & if\ x \in [\lambda_i, \lambda_{i+1}) \\ 0, & if\ x \notin [\lambda_i, \lambda_{i+1}) \end{cases},$$

$$r = 1, \dots, 3, \; i = -3, \dots, g.$$

The minimum number of augmented knot points is $8 + g$, where g is the number of internal knot points, and $\lambda_{-3} = \lambda_{-2} = \lambda_{-1} = \lambda_0 = a$ and $\lambda_{g+1} = \lambda_{g+2} = \lambda_{-g+3} = \lambda_{g+4} = b$. By using the *B*-spline basis every cubic spline has a unique representation on [a, b]

$$(2) \qquad s(x) = \sum_{i=-3}^{g} c_i N_{i,3+1}(x)$$

where the *B*-spline coefficients, c_i, have to be estimated for each spline function. It is also important to understand that the *B*-spline basis is completely defined by the knot points, and

4 See Schwarz (1989).

the splined function does not affect it. By construction the *B*-spline basis is normalized and it keeps the estimation of the coefficients numerically very stable.

(3) $$\sum_{i=-k}^{g} N_{i,k+1}(x) = 1, \quad \forall x \in [a, b]$$

On the other hand, a natural cubic interpolating spline function can also be defined as a solution to the following variational problem on an interval [a, b]

(4) $$\text{Min} \int_a^b \left(s^{(2)}(x)^2\right) dx \quad s.t. \quad y_j = s(x_j), \; j = 1, \ldots, n,$$

where $s^{(2)}(x)$ is the second derivative of the spline function, y_j is the value to be approximated, and $a \le x_1 < \ldots < x_n \le b$. The smoothing norm, the integral of the squared second derivative of the spline function, is referred as the "standard" smoothing norm hereafter.

By using the basic *B*-spline properties from de Boor (1978) and Gaffney (1976) the value of equation (4) can be calculated. Adams and Van Deventer (1994) propose that it can be used to measure the smoothness of the curve and rank the fitting methods. However, by definition the natural cubic spline minimizes the value but the value is hardly a good measure of smoothness as a ranking measure. For example, when we have the same knot positions but different kinds of curves to be fitted, we have an unchanged *B*-spline basis but the coefficients are different. The integration affects only the *B*-spline basis and not the coefficients. The result is that even with a perfect fit with the data (no noise in the data) the value of equation (4) depends on the shape of the curve. A smaller value does not necessarily mean a smoother curve. The natural interpretation of equation (4) is that each cubic spline guarantees the value to be a minimum amongst the fitting methods for the estimated curve.

Smoothing splines

The purpose of the smoothing spline is to remove the extra noise from the data. The properties we have discussed can also be used with smoothing splines. The most often used smoothing spline method is based on the following constrained variational problem on an interval [a, b].

(5) Minimize $$\eta := \int_a^b (s^{(2)}(x))^2 dx,$$

s.t. $$\delta := \sum_{j=1}^{n} (w_j(y_j - s(x_j)))^2 \le S,$$

where w_j is the weight value assigned to the data point. This can be interpreted as minimizing the roughness of the spline while simultaneously keeping the closeness of the fit at the desired level S, which is called the smoothing factor. The problem can be rewritten as finding a smoothing cubic spline $s_p(x)$ which minimizes

(6) $\varepsilon = \eta + p\delta$,

where the smoothing parameter p is determined so that the value $F(p) = S$, where

(7) $F(p) := \sum_{j=1}^{n} (w_j (y_j - s_p(x_j)))^2$,

and where $s_p(x_j)$ is a smoothing cubic spline adjusted by the smoothing parameter p. The smoothing spline $s_p(x)$ has the following properties:

1. If p tends to infinity, the smoothing spline becomes an interpolating spline, if all data points are used as knots as presented by Reinsch (1967). If only a subset of data points are used as knots, the spline becomes the least-squares spline corresponding to the given set of knots.
2. If p tends to zero, the smoothing spline becomes a straight line.

Dierckx (1975, 1981, 1982) has proposed another smoothing spline that uses a different smoothing norm. It is based on the following constrained minimization problem.

(8) Minimize $\tilde{\eta} := \sum_{i=1}^{g} (s^{(3)}(\lambda_i+) - s^{(3)}(\lambda_i-))^2$,

s.t. $\delta := \sum_{j=1}^{n} (w_j (y_j - s(x_j)))^2 \leq S$,

where $s^{(3)}(\lambda_i)$ is the third derivative of the spline function at the internal knot point. The method can be interpreted as minimizing the discontinuity jump at each internal knot point, λ_i, while simultaneously keeping the closeness of the fit at the desired level S. Although the interpretations are similar in these two estimation methods, the smoothing splines behaves a little bit differently. The solution to this problem is also a smoothing cubic spline and it can also be written using equations (6) and (7), where the objective is to minimize equation (6) and to determine the smoothing parameter p such that $F(p) = S$. The smoothing spline $s_p(x)$ has the following properties (see Dierckx (1995)):

1. If p tends to infinity, the smoothing spline becomes the least-squares spline corresponding to the given set of knots.

2. If p tends to zero, the smoothing spline becomes the weighted least-squares polynomial of degree three.

The Choice of the Smoothing Factor

Both smoothing spline methods require the user to provide the smoothing factor S to control the tradeoff between the roughness of the fit and the closeness of the fit. The recommended values for the smoothing factor depend on the relative weights w_j. Reinsch (1967) suggests choosing S in the range $n \pm \sqrt{2n}$ *if the weights are taken as the inverse of the standard deviation of* y_j. If the volatility is not known, S is determined by trial and error.

Wahba (1990) have presented a statistical technique called generalized cross validation, GCV, that detects the smoothing parameter p automatically by minimizing the following function

$$(9) \quad G(p) = \frac{nF(p)}{(n - Tr(A(p)))^2},$$

where n is the number of data points, $F(p)$ is the sum of squared residuals, $Tr(\mathbf{A}(p))$ is the trace of the matrix $\mathbf{A}(p)$, the so-called influence matrix (see Appendix B). The denominator of $G(p)$ can be interpreted as the square of the number of degrees of freedom for the smoothing spline. Wahba (1990) also maintains that GCV is a predictive mean-square error criterion. This criterion can be used to rank the models and we use it to rank the models in the empirical section.

There are a couple of important remarks that should be remembered when the GCV is used. First, the value of $G(p)$ depends on the knot positions. If only a subset of data points is used as knots, we should search the minimum over all possible subsets of knots. However, this is not usually done, but it is important to understand that knot locations can seriously affect the shape of the smoothing spline. Second, a small number of data points do not guarantee optimal smoothness when equation (9) is minimized if the number of knots is big compared to the total number of instruments. Indeed the value of equation (9) tends to zero when the number of data points is small. We will discuss more about this in the next sections.

III. ESTIMATING A SMOOTH TERM STRUCTURE

Term Structure Concepts

The term structure of interest rates is a curve that represents the relationship, at a given point of time, between time to maturity and yield to maturity of a pure discount bond. There are three different ways to characterize the term structure: 1) the discount factors, 2) the zero-coupon interest rates, and 3) the forward rates. In general, none of these representation forms are directly observable, since there are very few, if any, liquid zero-coupon bonds beyond maturities of one year. The yield curve has to be estimated from coupon bonds or from swap data.

The discount factor or the price of a pure discount bond specifies the present value of a unit payment in the future, i.e. the pure discount bond pays 1 unit at maturity. A set of discount factors specifies the discount function, represented by equation (10) in continuous time.

$$(10) \quad P(t, T) = e^{-r(t, T)(T-t)}$$

where $P(t,T)$ is the price of a T maturity pure discount bond at time t that pays 1 unit at maturity (or equivalently the discount factor for that bond), $r(t,T)$ is the yield of the pure discount bond. Even though the discount function may be difficult to interpret as a description of the term structure, it is the easiest function to estimate from coupon bonds. A coupon bond is a portfolio of zero-coupon bonds where every coupon payment is a separate discount bond. A second way to represent the term structure is zero-coupon interest rates. They are the yields of the pure discount bonds of the specific maturity, defined by equation (11).

$$(11) \quad r(t, T) = -\frac{\log(P(t, T))}{T-t}$$

This is also the most well known representation of the term structure of interest rates. A third way to represent the term structure is forward rates, defined by equation (12). They are the future interest rates implied by the current term structure.

$$(12) \quad f(t, T) = \frac{-\partial \log(P(t, T))}{\partial T},$$

where $f(t,T)$ is the forward rate for period T at time t.

The mathematical link between the equations (10) – (12) guarantees that if we know any single function of them we can calculate all other functions from the known one. Even though the term structure of interest rates can be upward sloping, flat, downward sloping, humped, or inverted humped, negative interest rates are not allowed. This can be achieved by restricting the discount factors to be monotonically decreasing. The proof follows from equation (12).

Estimation Method

The presented method estimates the term structure from coupon bond data. The price of a coupon bond is the present value of its coupon payments, and each payment has its own discount factor. Usually the estimation is done using the following least-squares method

(13) $\underset{c}{\text{Minimize}}\ [V(c) - P]^T[V(c) - P],$

where P is an $n \times 1$ vector of bond prices, $V(c)$ is the vector of the present values of the bonds, and c is a vector of the B-spline coefficients to be estimated. The spline can be placed directly on i) the discount function, ii) the log of the discount function, or iii) the forward rate function as presented by Fisher *et al.* (1995). The present value function, $V(c)$, can be presented as follows in these different approaches

(14i) $V(c) = \mathbf{BE}c,$

(14ii) $\mathbf{B}\exp(-\mathbf{E}c),$

(14iii) $V(c) = \mathbf{B}\exp(-\int_0^T \mathbf{E}dsc),$

where $\boldsymbol{B}$ is an $n \times m$ matrix of cash flows b_{ij} from bond i in period j for $i = 1,...,n,$ and $j = 1,...,m,$ $\boldsymbol{E}$ is an m *3* k evaluation matrix composed of the given knot set by equation (1), and c is a $k \times 1$ vector of B-spline coefficients.[5]

If the spline is placed on the discount function, equation (14i), we get the basic spline method presented by McCulloch (1971). When the forward rate function is splined, equation (14iii), we have to integrate the evaluation matrix (see Gaffney (1976)), $\boldsymbol{E}$, and we lose the normalized B-spline basis property (equation (3)) and the estimation method becomes numerically less stable. In general, it is tempting to spline the discount function, because it is a linear estimation problem, whereas the others are non-linear problems. However, there is an advantage in splining the log of the discount function, equation (14ii), as we can linearly extrapolate zero-coupon rates beyond the longest available maturity, which is not possible if we spline the discount function.[6]

In our smoothing spline method we combine the smoothing spline method from Dierckx (1975, 1981, 1982), equation (8), and the standard least-squares method applied to coupon bonds, equation (13). We propose the following estimation problem:

(15) $\underset{c(p)}{\text{Minimize}}\ c(p)^T\mathbf{H}c(p) + p([V(c(p)) - \mathrm{P}]^T\mathbf{W}[V(c(p)) - P]),$

where $\boldsymbol{H}$ is a matrix of the squared discontinuity jumps in the third derivatives of the spline at the interior knot points, $\boldsymbol{W}$ is a diagonal weighting matrix, $c(p)$ is a column vector of B-spline

5 The evaluation matrix should be presented as $\boldsymbol{E(t)}$ as it is time dependent, but we have suppressed it for notational simplicity. For example, in equation (14iii) we integrate only the matrix and not the c vector.
6 If we spline the log of the discount function, we can make an assumption that the yield curve is linear beyond the longest available maturity date.

coefficients which depends on the smoothing parameter *p*, *P* is a column vector of bond prices, and *V (c (p))* is the vector of the present values of the bonds.

When we spline the discount function, the optimization problem reduces to the closed form, but when we spline the log of the discount function the optimization problem has to be solved numerically (see Appendix C). The smoothing parameter *p* can be detected by the GCV method, equation (9), where *n* is the number of instruments. When we have a non-linear fitting problem we have to estimate the coefficients for each *p* estimate, so the GCV estimation slows down quite a lot from the linear case. Furthermore, the GCV method is derived for the linear case (see Wahba (1990)).

There is still one important aspect in the estimation problem (the number and locations of the knot points. This is due to the fact that knot points also affect the GCV value, especially when the number of instruments, *n*, is small. McCulloch (1975) proposes that square root of *n* is a good number of knots. He also recommends that the knots should be located so that there is an equal number of instruments between the knot points. His argument is that it allows the spline to fit equally complex shapes for all values of the abscissae (see McCulloch (1975)). Fisher *et al.* (1995), in contrast, uses the maturity of every third instrument as a knot point. Other papers of the term structure estimation do not specify how they position the knots.

Dierckx (1995) presents that we add knots only to the points where the fit is poor. We start with a cubic polynomial, i.e. only end points are used as knots, and add knots to the areas where the squared residuals are largest. The knot is added approximately in the middle of the interval that has the largest squared residuals between the current knot points. The previous knots are not relocated, the new knots are only added to the set. The addition of knots guarantees that the sum of the squared residuals decreases after each knot addition (see Dierckx (1995)). The number of knots can be detected by the smoothing factor *S*, i.e. we add knots until the sum of squared residuals is smaller than *S*. If we use the GCV method to find the optimal smoothness, such as in this paper, we should search over all possible set of knots to find the minimum of the GCV function.[7]

In this paper we limit the number of internal knots points to two. This limitation is due to the small number of instruments. We compare this approach to the method presented by McCulloch (1975), as we use the same number of knots but an equal number of instruments falls between the knots. The method that has the smallest GCV value is ranked a better fitting method.

7 None of the papers that we are aware of have done this.

IV. EMPIRICAL RESULTS

The Data

In Finland, government benchmark bonds have existed only since the introduction of the primary dealer system in August 1992. The benchmark bonds are government bonds, for which primary dealers have to give two-way quotations. During the period June 3, 1993 through February 6, 1996, there have been traded five benchmark bonds, the longest bond maturing in 2004. All the benchmark bonds are bullet bonds with annual coupons. They are quoted on an annual yield basis with a 30/360 year basis. The bid-ask spread has been no more than five basis points of yield to maturity in all maturities. The size of the market has doubled during the period and in February 1996 was about 103 billion Finnish markkas (FIM) (USD 23 billion).[8]

The data from the money market is from the bank CDs as they are the only instruments that have been traded during the whole period with sufficient liquidity.[9] The money market instruments include 1, 2, 3, 6, 9, and 12-month CDs. The CDs are zero-coupon instruments and they are quoted on a money market yield basis with an Actual/365 year basis. The size of the market of the bank CDs has been quite steady during the period, about FIM 80 billion (USD 18) and the liquidity has been good.

Our data are limited to these six money market instruments and the five benchmark bonds. The Repo market is very illiquid and the data from it cannot be used in the term structure estimation. When the maturity of a bond approaches one year, the bond trades at considerably lower yields than the one-year bank CDs.[10] This feature distorts the estimation, so that bonds that have less than one year and three months time to maturity were omitted from the sample. As the bonds and the CDs have a different day count basis, the CDs are converted to a 30/360 year basis after their yields are annualized.[11] The data cover the period June 3, 1993 through February 6, 1996 and the total number of days is 679.

8 A good description of the Finnish Bond and Money Market is given by Valtonen *et al.*(1996).

9 The T-bill market is younger and the liquidity of the market is thin. During the last year T-bills have traded 1–3 bp lower than bank CDs.

10 This phenomenon seems to be related to the discontinuation of the benchmark status of a bond. The Bank of Finland announces the discontinuation date and the conversion period during which the bond can be converted to other instruments. During the estimation period the benchmark status of two bonds was discontinued and new benchmark bonds were introduced.

11 The continuously compounded annualized yield of the money market instrument can be calculated by using equation (2), where the price of the instrument is

$$P(t,T) = \frac{100\ \%}{1 + rd/365}$$, where d is the number of days between t and T, d ≤ 365, and r is the quoted yield.

The conversion is done by multiplying the CD yields by 360/365.

The Results

We fit the term structure of interest rates by splining the discount function and the log of discount function, equations (14i) and (14ii), respectively.[12] First we discuss the number of knot points. Second we compare the results to the results fitted by the "standard" smoothing norm. Third we compare the method of the freely located knot points to the method of equally spaced knot points. The comparison is done in all cases by the GCV value — the smallest GCV value is ranked as the best method.

Figure 1 illustrates an example of what happens when we have three internal knot points instead of two internal points and use the GCV function to find the optimal smoothness. The knot points are located by the size of the fitting errors. The spline is placed on the log of the

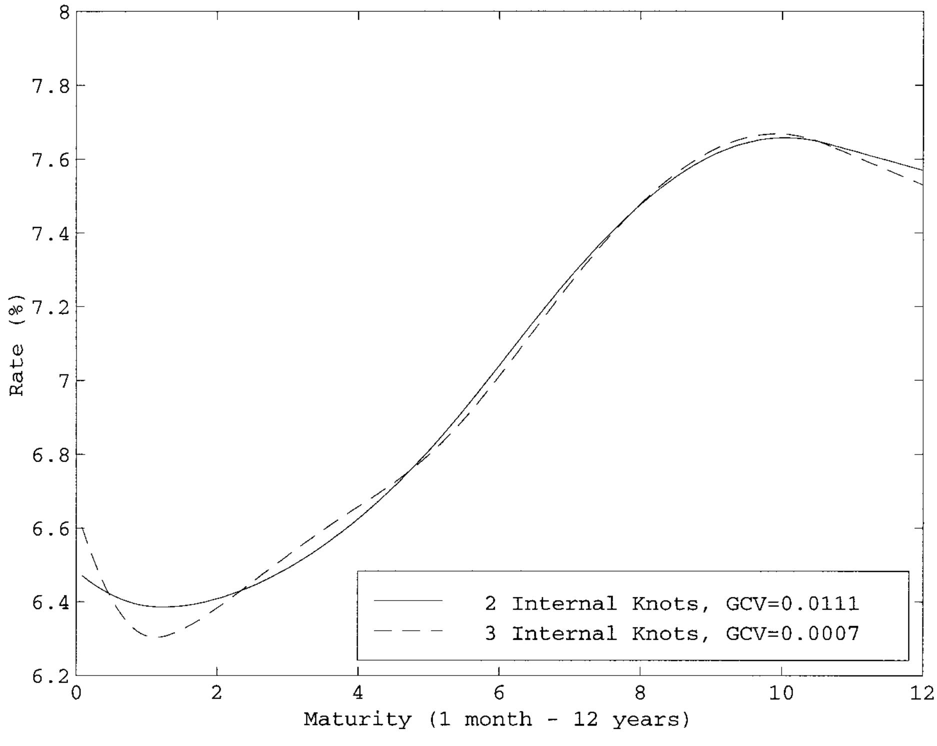

FIGURE 1. Fitting with 2 and 3 Internal Knots

12 We also tested the forward rate fitting, but the results were much worse. Our main findings were that the forward rate curve required more knot points than other methods and pricing errors were at least 10 times larger. One issue that can partly explain the results is the small number of instruments, but we also found that the integrated *B*-spline basis was numerically less stable. The smoothing has nothing to do with the results, because the results were similar without smoothing.

discount function and the values beyond 10.6 years are linearly extrapolated. The solid line represents the yield curve with two internal knot points and the dashed line is the fitted yield curve with three internal knot points. The dashed line is clearly overfitted. If the GCV value is used to rank the models, we would have selected the model with three internal knot points. This kind of problem occurs because the number of instruments is too small. For example, Wahba (1990) states that the GCV method requires over 25 data points to be effective. In order to get acceptable results with a small number of instruments we have to limit the number of knot points. Intuitively the limitation forces the pricing errors to be big enough that we can smooth the curve (see also figure 2). The maximum number of knots can be found by trial and error. For this sample it has been found to be two internal knot points.

Table 1 panel A shows the ranking statistics of the fitting methods when the knots are located freely by the size of the fitting error. The method that splines the log of the discount function with the smoothing norm of the discontinuity jumps at the third derivatives is ranked clearly the best model with 422 number one rankings of the 679 possibilities. Surprisingly, the fitting with the discount function and using the "standard" smoothing norm gets the second highest number of number one rankings.

Panel B shows the results when the models are compared with the equally spaced knot points. The previous winner is also ranked the best with 261 number one rankings. Interestingly, the same model but where the equal number of instruments fall between the knot points is ranked the second best model. The third best model is the discount function method with the jump smoothing norm and freely located knot points.

Panel B also shows the average and median of the effective number of parameters. The effective number of parameters is the trace of the influence matrix $\boldsymbol{A}(p)$. The mean and median values are between 5 and 6 in all estimated methods. The values are a little bit smaller when we have equally spaced knot points. However, the equally spaced methods get less number one rankings in general. The reason is that the equally spaced methods have higher pricing errors.

Panel C shows the average absolute pricing errors of the fitting methods. The smallest errors are when knot points are located by the size of the pricing errors and the highest errors are around the one-year maturity. When the knots are located so that there is an equal number of instruments between the knot points, the average absolute pricing errors are much higher but similar in all four methods. Furthermore, the highest pricing errors are now at the range of 2 to 6 years' maturity.

Figure 2 illustrates an example of what might happen when the knots are freely located or equally spaced. The fitting method is the log of the discount function with a jump smoothing norm. The solid line represents the method with freely located knots and the dashed line the

Table 1. Descriptive statistics of the ranking of the fitting methods, the mean and median of the number of effective parameters (EP), and the mean absolute pricing errors in basis points. Instruments are 1– to 12-month bank CDs and five benchmark bonds during the period June 3, 1993 through February 6, 1996 and the total number of days is 679. The acronyms of the models are the following: DFJ is the discount function and LDFJ is the log of the discount function with a jump smoothing norm, DFS and LDFS are smoothed with the "standard" smoothing norm. The knot points are located freely by the size of the fitting errors in these models. The acronyms that have E as the last letter are the models where an equal number of instruments falls between the knot points.

PANEL A: RANKING OF THE MODELS (FREELY LOCATED KNOT POINTS)

	DFJ	LDFJ	DFS	LDFS
RANK 1	50	422	121	86

PANEL B: RANKING OF THE MODELS (FREELY LOCATED AND EQUALLY SPACED KNOT POINTS) AND THE EPS

	DFJ	LDFJ	DFS	LDFS	DFJE	LDFJE	DFSE	LDFSE
RANK 1	36	261	103	52	73	134	11	9
MEAN EP	5.859	5.511	5.956	5.699	5.539	5.190	5.682	5.192
MEDIAN EP	5.949	5.448	5.952	5.901	5.681	5.285	5.705	5.323

PANEL C: PRICING ERRORS IN BASIS POINTS

TIME TO MATURITY	DFJ	LDFJ	DFS	LDFS	DFJE	LDFJE	DFSE	LDFSE
1 M	0.30	0.59	0.29	0.65	0.29	0.60	0.37	1.13
2 M	0.73	1.19	0.69	1.37	0.96	1.09	0.95	1.86
3 M	1.40	1.63	1.35	1.96	2.00	1.68	1.74	2.23
6 M	3.80	3.46	3.75	4.28	5.19	3.64	5.23	4.14
9 M	4.74	4.31	4.85	5.45	4.29	3.81	4.39	4.48
12 M	10.38	8.75	10.59	9.44	7.02	8.98	6.39	8.70
1–2 Y	8.70	11.32	8.51	10.09	8.49	9.26	8.78	8.93
2–4 Y	6.86	7.32	6.83	6.42	12.64	10.69	12.53	10.64
4–6 Y	4.56	5.97	4.37	4.34	13.06	14.82	13.01	15.35
6–8 Y	1.21	2.18	1.29	2.58	5.53	9.86	4.82	11.14
8–11 Y	0.33	0.74	0.46	1.38	1.89	4.20	1.44	5.09

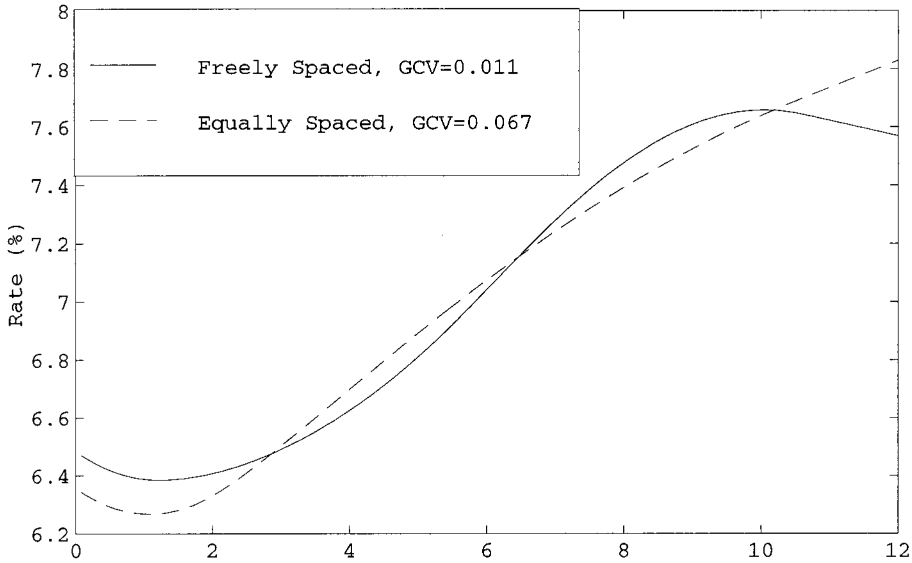

FIGURE 2A. Fitting with Freely and Equally Spaced Knots

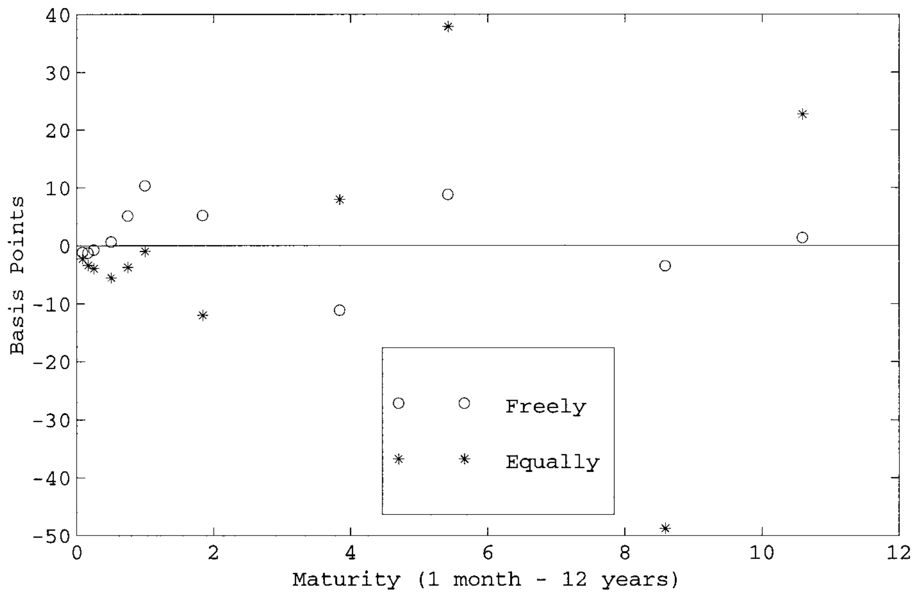

Figure 2B. Pricing Errors with Freely and Equally Spaced Knots

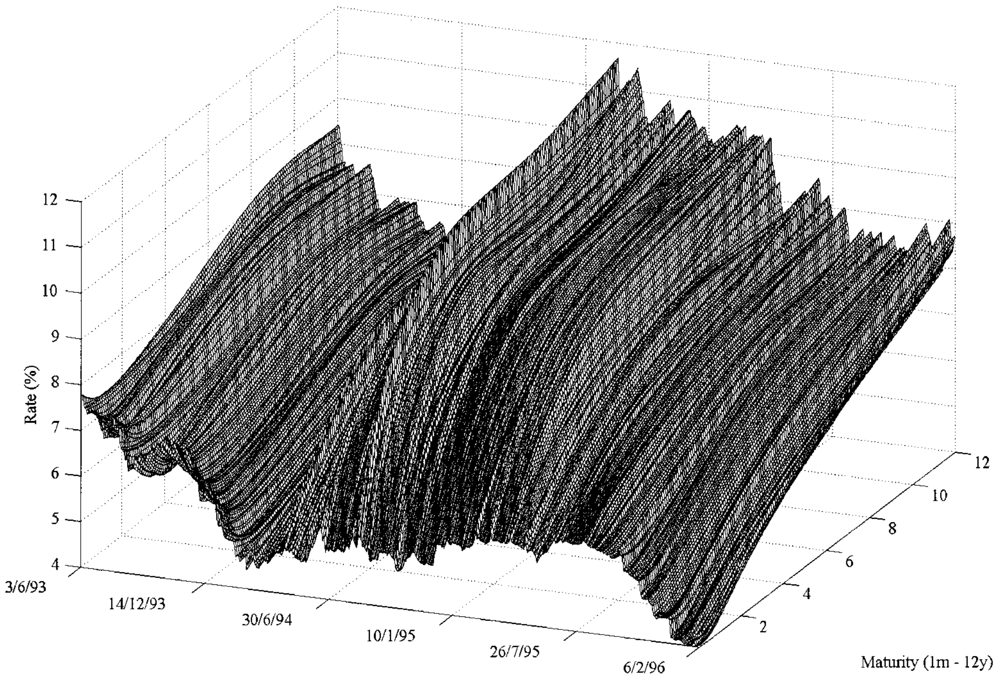

Figure 3. Yield Curves When Splining the Log of the Discount Function

method with equally spaced knots. The curves look very different and it is tempting to say that the dashed line represents the better model. However, when we look at the lower panel which shows the pricing errors, we see that they are much higher when knots are equally spaced. Furthermore, the GCV value is also higher in the equally spaced method.

Finally figure 3 illustrates the fitted yield curves for the whole sample period. The method is the log of the discount function with a jump smoothing norm and freely located knots. During the sample period the longest maturity has been a little less than 11 years in the beginning of the period and at the end of the period it has been a little over 8 years. The values beyond the longest maturity have been linearly interpolated.

V. CONCLUSIONS

This paper has presented a new method to estimate a smooth term structure of interest rates. The method uses a different smoothing norm, the square of the discontinuity jumps in the third derivatives at the internal knot points. The method is compared to the ”standard” smoothing norm fitting methods. In addition, we also discuss the GCV method and the problems that might occur with it. Moreover, we also located the knot points by the size of the fitting errors and compare the results to the method proposed by McCulloch (1975).

The results show that the best model is the method that places the spline on the log of the discount function, uses the jump smoothing norm and locates the knots by the size of the fitting errors. As our sample is limited to a small number of instruments, it would be interesting to study a larger market before any further conclusions are made.

REFERENCES

ADAMS, K. J., and D. R. VAN DEVENTER (1994), Fitting Yield Curves and Forward Rate Curves with Maximum Smoothness, *Journal of Fixed Income*, 4, 52–62.

DE BOOR, C. (1978), *A Practical Guide to Splines*, Applied Mathematical Sciences, Vol. 27, New York.

COLEMAN, T. S., L. FISHER and R. IBBOTSON (1992), Estimating the Term Structure of Interest Rates from Data that Include the Prices of Coupon Bonds, *Journal of Fixed Income*, 2, 85–116.

DIERCKX, P. (1975), An Algorithm for Smoothing, Differentiation, and Integration of Experimental Data Using Spline Functions, *Journal of Computational and Applied Mathematics*, 1, 165–181.

DIERCKX, P. (1981), An Improved Algorithm for Curve Fitting with Spline Functions, *TW Report 54*, Department of Computer Science, Katholieke Universiteit Leuven.

DIERCKX, P. (1982), A Fast Algorithm for Smoothing Data on a Rectangular Grid While Using Spline Functions, *SIAM Journal on Numerical Analysis*, 19, 1286–1304.

DIERCKX, P. (1995), *Curve and Surface Fitting with Splines*, Clarendon Press, Oxford.

FISHER, M., D. NYCHKA, and D. ZERVOS (1995), Fitting the Term Structure of Interest Rates with Smoothing Splines, Federal Reserve System Working Paper 95–1.

GAFFNEY, P. W. (1976), The Calculation of the Indefinite Integrals of B-splines, *Journal of the Institute of Mathematics and its Applications*, 12, 329–336.

HULL, J., AND A. White (1996), Using Hull-White Interest Rate Trees, *Journal of Derivatives*, 4, 26–36.

ILMANEN, A. (1995), Overview of Forward Rate Analysis (Part 1), Salomon Brothers Research Paper.

MCCULLOCH, J. H. (1971), Measuring the Term Structure of Interest Rates, *Journal of Business*, 44, 19–31.

MCCULLOCH, J. H. (1975), The Tax-Adjusted Yield Curve, *Journal of Finance*, 30, 811–830.

NELSON, C., and **A. SIEGEL** (1987), Parsimonious Modeling of Yield Curves, *Journal of Business*, 60, 473–489.

REINSCH, C. (1967), Smoothing by Spline Functions, *Numerische Mathematik*, 10, 177–183.

SCHWARZ, H. R. (1989), *Numerical Analysis: A Comprehensive Introduction*, John Wiley & Sons, New York.

VALTONEN, E., P. HEINARO, and **A. MÄKINEN** (1996), A Guide to the Finnish Bond and Money Market, Handelsbanken Markets.

VASICEK, O. A., and **H. C. FONG** (1982), Term Structure Modeling Using Exponential Splines, *Journal of Finance*, 37, 339–348.

WAHBA, G. (1990), *Spline Models for Observational Data*, SIAM, Philadelphia.

Appendix A: Calculation of the Discontinuity Jump Matrix

The elements of the matrix **H**, the squared discontinuity jump in the third derivative of the spline function, *s(x)*, at the internal knot point can be calculated as follows

$$\text{(A1)}\qquad a_{i,q} = \begin{cases} 0 & , \text{ if } i < q-4 \text{ or } i < q, \\ \dfrac{3!\lambda_{i+4} - \lambda_i}{\prod\limits_{\substack{j=i\\ j\neq q}}^{i+4} (\lambda_q - \lambda_j)} & , \text{ if } q-4 \le i \le q, \end{cases}$$

where $q = 1,\ldots,g$, g is the number of internal knot points, and $i = -3,\ldots,g$. The matrix **H** is as follows

$$\text{(A2)}\qquad H = \begin{bmatrix} a_{-3,1} & \cdots & a_{g,1} \\ \vdots & \ddots & \vdots \\ a_{-3,g} & \cdots & a_{g,g} \end{bmatrix}^T \begin{bmatrix} a_{-3,1} & \cdots & a_{g,1} \\ \vdots & \ddots & \vdots \\ a_{-3,g} & \cdots & a_{g,g} \end{bmatrix}.$$

Appendix B: The Influence Matrix A(p)

The influence matrix in the GCV criteria is defined in the linear case, i.e. the spline is placed on the discount function, as follows

$$\text{(B1)}\qquad \mathbf{A}(p) = \mathbf{BE}[\mathbf{E}^T\mathbf{B}^T\mathbf{WBE} + \mathbf{H}p^{-1}]^{-1}\mathbf{E}^T\mathbf{B}^T\mathbf{W}.$$

When the spline is placed on the log of the discount function, i.e. the nonlinear case, the equation (B1) is as follows

$$\text{(B2)}\qquad \mathbf{A}(p) = \mathbf{Z}(c(p))[\mathbf{Z}(c(p))^T\mathbf{WZ}(c(p)) + \mathbf{H}p^{-1}]\,\mathbf{Z}(c(p))^T\mathbf{W},$$

where $Z(c(p))^T = \dfrac{\partial V(c(p))}{\partial c(p)}$.

Appendix C: Simplification of the Estimation Problem

The optimization problem can be simplified in the linear case as follows

$$\text{(C1)}\qquad c = [\mathbf{E}^T\mathbf{B}^T\mathbf{WBE} + \mathbf{H}p^{-1}]^{-1}\mathbf{E}^T\mathbf{B}^T\mathbf{W}P.$$

The nonlinear case can be linearized around an initial guess c^i using the standard Taylor series approach as follows

(C2) $$V(c(p)) = V(c(p)^i) + [c(p) - c(p)^i]\left.\frac{\partial V(c(p))}{\partial c(p)^T}\right|_{c(p) = c(p)^i},$$

and define

(C3) $$\mathbf{Z}(c(p)^i) = \left.\frac{\partial V(c(p))}{\partial c(p)^T}\right|_{c(p) = c(p)^i},$$

(C4) $$Y(c(p)^i) = P - V(c(p)^i) + c(p)^i\mathbf{Z}(c(p)^i).$$

The minimizer for the optimization problem can be found by

(C5) $$c(p)^{i+1} = [\mathbf{Z}(c(p)^i)\mathbf{W}\mathbf{Z}(c(p)^i)^T + \mathbf{H}p^{-1}]^{-1}\mathbf{Z}(c(p)^i\mathbf{W}Y(c(p)^i),$$

where $c(p)^{i+1}$ is an updated $c(p)^i$. We iterate until convergence.