Review of Quantitative Finance and Accounting, 9 (1997): 251–267
© 1997 Kluwer Academic Publishers, Boston. Manufactured in The Netherlands.

Nonparametric Smoothing of Yield Curves

CARSTEN TANGGAARD
Department of Finance, The Aarhus School of Business

Abstract. This paper proposes a new nonparametric approach to the problem of inferring term structure estimates using coupon bond prices. The nonparametric estimator is defined on the basis of a penalized least squares criterion. The solution is a natural cubic spline, and the paper presents an iterative procedure for solving the non-linear first-order conditions. Besides smoothness, there are no a priori restrictions on the yield curve, and the position of the knots and the optimal smoothness can be determined from data. For these reasons the smoothing procedure is said to be completely data driven. The paper also demonstrates that smoothing a simple transformation of the yield curve greatly improves the stability of longer-term yield curve estimates.

Key words: Yield curves, term structure of interest rates, nonparametric regression, cubic splines, generalized cross validation, optimal smoothness

1. Introduction

The use of observed yields on bonds represents problems in financial applications and in research on the term structure of interest rates. First, small pricing errors resulting from market imperfections necessitate smoothing. Second, observed yields from coupon bonds are biased estimators of the underlying zero-coupon bond yields. Finally, even when zero-coupon bond prices are available, there is a need for interpolation of yields between adjacent maturities.

All these problems can be solved by assuming a smooth functional specification of the underlying zero-coupon yield curve. Non-linear least squares software can then be used for inference on the term structure. The main problem with this approach, however, is the selection of a proper functional form. For example, in the polynomial model of the yield curve suggested by Chambers, Carleton and Waldman (1984), the resulting yield curve is highly unstable in the long term, and while other parametric specifications e.g. (Nelson and Siegel, 1987) have remedied this problem, the functional form is often too rigid to capture the shape of actual yield curves.

Cubic regression splines were first used for term structure estimation by McCulloch (1975), who also suggested a rule of thumb for the position of the knots. McCulloch (1971) was the first article to use splines of any type on the term structure problem, and both articles were among the first to use splines on economic problems at all. Other uses of splines for term structure estimation are found in Vasicek and Fong (1982), and Shea (1984).

In recent years, nonparametric regression has gained popularity in economic and financial applications (Robinson, 1988; Delgado and Robinson, 1992). The standard nonparametric regression model (Eubank, 1988; Härdle, 1990) is not directly useful for term structure estimation, however, except in the special case where data consists of zero-coupon bonds. The aim of this article is thus to develop a procedure for nonparametric estimation in the more general case of coupon bond prices.

In section 2 of the article, a natural cubic spline smoothing estimator is derived from a generalization of the traditional penalized least squares approach to nonparametric regression. Section 3 presents some details of the solution procedure in a number of special cases. The cubic spline smoothing procedure is highly dependent on a smoothness parameter, and section 4 proposes generalized cross validation for estimating the correct degree of smoothness. An application for Danish government bonds is presented in section 5. Finally, some concluding remarks are given in section 6.

2. Spline smoothing of the term structure

The following notation will be used. Let $p_1,\ldots,p_n$ denote the observed prices of n bonds from which the term structure is to be inferred. Bonds can either be pure discount bonds or bonds with intermediate payments. Bond i has *fixed* payments, $c_i(t_j)$: $j = 1,\ldots,m_i$, due on dates t_j, $j = 1,\ldots,m_i$. The payment, $c_i(t_j)$, consists of coupon and repayment of principal. Furthermore, let $m = \max(m_1,\ldots,m_n)$ be the number of dates on which at least one payment is due, and let $t_1,\ldots,t_m$ denote the corresponding set of dates. m is bounded below by $\max(m_1,\ldots,m_n)$, but could be as high as $\Sigma\, m_i$, since m_i is the number of payments of bond i.

The full set of payments is arranged in an $n \times m$ matrix, $\mathbf{C}$, with $c_i(t_j)$ in cell i, j. It should be clear from the definition that $c_i(t_j) = 0$, if bond i has no payment on date t_j.

The term structure can be represented by the yield curve, $y(t)$, of pure discount bonds, or directly by the discount function, $d(t) = \exp(-ty(t))$. These alternative ways of looking at the term structure are equivalent from an economic point of view. From a statistical point of view, however, the curvature properties depend very much on the functional form. In consequence, the procedure presented here is so general as to allow a variety of different statistical models of the present value function.

In future, let φ denote whatever function ($d(t)$, $y(t)$ …) is to be estimated and let the notation $pv = pv(\varphi)$ be used to emphasize the dependency on φ. When φ is the discount function, $d(t)$, the present value of bond i is

$$pv_i(\varphi) = \sum_{j=1}^{m_i} c_i(t_j)d(t_j) \equiv \sum_{j=1}^{m_i} c_i(t_j)\varphi(t_j) \tag{1}$$

It has been argued that, in the context of regression splines, direct estimation of the discount function (as proposed by McCulloch (1975)) gives severe problems with the implied estimates of the yield curve (Vasicek and Fong, 1982; Shea, 1984). All objections to these models also apply to the model suggested in this article, and the model given by (1) will only be used for illustrative purposes.

If, instead, φ is the yield curve, $y(t)$, the present value is

$$pv_i(\varphi) = \sum_{j=1}^{m_i} c_i(t_j) \exp[-t_j y(t_j)] \equiv \sum_{j=1}^{m_i} c_i(t_j) \exp[-t_j \varphi(t_j)] \tag{2}$$

This specification solves most of the problems implied by specification (1). However, many theories of term structure assume some kind of mean reverting behavior, which results in yield curves being asymptotically constant as $t \rightarrow \infty$ (Cox, Ingersoll and Ross (1985) is an important example). Unfortunately, the cubic spline smoothing estimators considered here are asymptotically linear, though with no guarantee of a zero slope[1]. Consider, therefore, estimation of the curve, $u(t)$, related to the yield curve by $y(t) = u(t)/(1 + t)$. Using the smoothing procedure, $\lim u'(t) = y^*$, say. This gives the desired asymptotic behavior, $\lim y(t) = y^*$. In addition the exponential curvature of the discount function is automatically imposed[2].

In this model, the present value of bond i is

$$pv_i(\varphi) = \sum_{j=1}^{m_i} c_i(t_j) \exp\left[-\frac{t_j}{1+t_j} u(t_j)\right] \equiv \sum_{j=1}^{m_i} c_i(t_j) \exp\left[-\frac{t_j}{1+t_j} \varphi(t_j)\right] \tag{3}$$

Any of the three models (1), (2) and (3) can be estimated by using the techniques presented in this article.

Because of non synchronous trading, credit risks, tax effects, and market imperfections in general, the no-arbitrage condition, or law of one price, is not fulfilled in a strict sense, so an error term, ϵ_i, is added to catch the pricing errors, i.e. $p_i = pv_i + \epsilon_i$. The error terms will be assumed to be independent across the sample. However, the variance, σ_i^2, of ϵ_i will be allowed to vary from bond to bond.

The nonparametric estimator, $\hat{\varphi}$, of φ is now defined as the solution of a penalized least squares problem:

$$\hat{\varphi} = \arg\min_{\varphi} \left\{ \frac{1}{n} \sum_{=1}^{n} w_i [p_i - pv_i(\varphi)]^2 + \lambda \int [\varphi''(\tau)]^2 d\tau \right\} \tag{4}$$

where λ is a smoothness parameter which fulfills $0 < \lambda < \infty$. Apart from a constant, the weights, w_i, are assumed to be known. A natural choice of the weight system is to set w_i to be proportional to the inverse of the variance on p_i, i.e. $V(p_i) = \sigma^2 w_i^{-1}$. This will be assumed throughout[3].

Criterion (4) is a natural generalization of the criterion for nonparametric estimation in a simple regression model with one explanatory variable (Eubank, 1988). Accordingly, the interpretation is similar. The first term on the right hand side of (4) is a goodness-of-fit criterion, while the second term is a function of the speed of change of the first-order derivative of φ. For a linear function, φ, $\int \varphi''(\tau)^2 d\tau = 0$ while $\int \varphi''(\tau)^2 d\tau > 0$ for any nonlinear φ. Generally speaking, the more nonlinear φ is, the larger is the penalty term. In this sense, the smoothness parameter, λ, determines the balance between the two conflicting goals, goodness of fit and smoothness of the estimated curve. How the choice of λ affects the estimator of φ will be discussed in section 4 below.

3. The solution procedure

Let $C^2[a, b]$ denote the vector space of functions with 2 continuous derivatives in $[a, b]$, and assume that $a < t_1 < \ldots < t_m < b$. Assuming that φ belongs to $C^2[a, b]$, it follows that the solution to (4) is a *natural cubic spline* with knots at $t_1,\ldots,t_m$. This is a consequence of the nature of the penalty term[4].

A cubic spline function, φ, is any function on $[a, b]$ of the following type:

$$\varphi(t) = \sum_{k=0}^{3} \theta_k t^k + \sum_{j=1}^{m} \delta_j (t - t_j)_+^3 \tag{5}$$

where $t_+ = \max(0,t)$. φ coincides with a 3-degree polynomial on $[t_j, t_{j+1})$ and has at least 2 continuous derivatives in $[a, b]$ (see Eubank (1988), chapter 5). The third order derivative is a step function that is constant for t in $[t_j, t_{j+1})$.

The set of spline functions, $S[t_1,\ldots,t_m]$, with representation (5) is a vector space of dimension $m + 4$. A *natural cubic spline* is linear outside $[t_1, t_m]$. It can be shown that this imposes 4 independent restrictions across the parameters, $\theta_0,\ldots,\theta_3$, $\delta_1,\ldots,\delta_m$. Thus, the vector space, $NS[t_1,\ldots,t_m]$, of natural cubic splines has a dimension m. Two of these so-called natural restrictions are $\theta_2 = \theta_3 = 0$, while the remaining 2 are more complicated[5]. Because the representation (5) is not used directly for estimation purposes, this topic will not be discussed further here.

When $x_1(t),\ldots,x_m(t)$ is a set of independent functions spanning the solution space, $NS[t_1,\ldots,t_m]$, φ can be expressed as

$$\varphi(t) = \sum_{k=1}^{m} \beta_k x_k(t) \tag{6}$$

where the parameters, β_k, $i = k,\ldots,m$, must be determined from the data.

Before proceeding with the solution, a brief explanation is in order of the difference between using regression splines and smoothing splines. The same mathematical tool, cubic splines, is used in both regression and smoothing splines. However, from an application point of view, there are important differences. One main difference is that regression splines require determination of the knot positions, while smoothing splines require an estimate of the smoothness parameter, λ (Eubank, 1988, chapter 7). A further difference is seen in the number of parameters to be determined from data. This number, m, may very well be much bigger than the number, n, of observations. In one sense, therefore, the problem is over-parameterized, and cannot be solved without imposing some further restrictions. The purpose of the estimation procedure developed below can therefore also be seen as a way of determing which restrictions are necessary. In addition, in the regression spline approach, the cubic spline solution is imposed, while in the nonparametric approach, the functional form is the outcome of a proper estimation criterion.

In deriving the solution, some matrix notation will become useful. Let ϕ, β denote $m \times 1$ vectors: ϕ is the vector of evaluations, $\varphi(t_i)$, in the knots, and β is the vector of β_i parameters. The $n \times 1$ vectors of prices, p_i, and present values, pv_i, will be denoted **p** and **pv** respectively. In addition, let $\mathbf{X} = \{x_{ij}\}$ denote a matrix with elements $x_{ij} = x_j(t_i)$. **X** is non-singular, and maps β into φ, i.e. $\phi = X\beta$. Finally, let Δ be the $m \times m$ matrix with elements, $\{\delta_{ij}\}$, defined in Lemma 5.1, p. 199 in Eubank (1988). Because $x_i(t)$ is a spline function, it has a representation (5), and Δ is the $m \times m$ matrix with the i'th row consisting of the δ's of the representation (5) for $x_i(t)$.

In a second application of Lemma 5.1, Eubank (1988), the penalized least squares criterion (4) becomes equivalent to

$$\hat{\beta} = \arg\min_{\beta} \left\{ \frac{1}{n} (\mathbf{p} - \mathbf{pv}(\mathbf{X}\beta))' \mathbf{W} (\mathbf{p} - \mathbf{pv}(\mathbf{X}\beta)) + \lambda 6 \beta' \mathbf{X} \Delta \beta \right\} \tag{7}$$

where parentheses around β emphasize the dependency on β. **W** is the $n \times n$ matrix with w_i along the diagonal and zeros outside. **W** is related to the variance by $\mathbf{V}(\mathbf{p}) = \sigma^2 \mathbf{W}^{-1}$. By both differentiation w.r. to β, and rearranging, the first-order condition for determination of β becomes

$$\mathbf{X}'\mathbf{B}'(\phi)\mathbf{W}(\mathbf{p} - \mathbf{pv}(\phi)) - n\lambda \mathbf{X}'\mathbf{G}\beta = 0 \tag{8}$$

where $\mathbf{G} = 3!\Delta$, $\phi = \mathbf{X}\beta$, and

$$\mathbf{B}' = \frac{\partial \mathbf{pv}}{\partial \beta} \tag{9}$$

is an $m \times n$ matrix of partial derivatives of pv_j w.r. to β_i.[6]

There are several ways to establish a cubic spline basis (see Eubank (1988), section 5.3.3 for a detailed discussion). For example, there is a basis that parameterizes in terms of the values in the knots, and in which $\mathbf{X} = \mathbf{I}$ and $\beta = \phi$. This keeps notation more simple, and this basis will be assumed throughout the remaining part of the article[7].

The elements, b_{ij}, of **B** depend on the specification of $pv(\varphi)$. In the case of (1)

$$b_{ij} = c_i(t_j) \tag{10}$$

Thus, $\mathbf{B} = \mathbf{C}$ does not depend on ϕ. In models (2) and (3), **B** is defined by

$$b_{ij} = -c_i(t_j) t_j \exp[-t_j \phi_j] \tag{11}$$

and

$$b_{ij} = -c_i(t_j)\frac{t_j}{1+t_j}\exp\left[-\frac{t_j}{1+t_j}\phi_j\right] \tag{12}$$

It is now obvious that **B** depends on ϕ.

In the first case, where **B** does not depend on ϕ (equation (10)), the first-order condition becomes

$$\mathbf{B}'\mathbf{W}(\mathbf{p} - \mathbf{B}\phi) - n\lambda\mathbf{G}\phi = 0 \tag{13}$$

with the solution

$$\hat{\phi} = (\lambda n\mathbf{G} + \mathbf{B}'\mathbf{W}\mathbf{B})^{-1}\mathbf{B}'\mathbf{W}\mathbf{p} \tag{14}$$

Where **B** does depend on ϕ (equations (11) and (12)), the first-order condition

$$\mathbf{B}(\phi)'\mathbf{W}(\mathbf{p} - \mathbf{pv}(\phi)) - n\lambda\mathbf{G}\phi = 0 \tag{15}$$

cannot be solved analytically. A numerical solution can, however, be obtained by the iterative scheme

$$\hat{\phi}^{k+1} = \hat{\phi}^k + [\lambda n\mathbf{G} + \mathbf{B}'(\hat{\phi}^k)\mathbf{W}\mathbf{B}(\hat{\phi}^k)]^{-1}[\mathbf{B}'(\hat{\phi}^k)\mathbf{W}(\mathbf{p} - \mathbf{pv}(\hat{\phi}^k)) - \lambda n\mathbf{G}\hat{\phi}^k] \tag{16}$$

where $\hat{\phi}^k$ denotes the estimate at the k'th stage of the iterative procedure. This is essentially the Newton-Raphson algorithm, with the derivatives replaced by their expectation. This is, of course, closely related to the Gauss-Newton algorithm for nonlinear least squares estimation.

The dimension, m, of ϕ can be quite large, and it may seem difficult to solve the equations. However, several of the matrices in (16) have a band-diagonal structure, which is exploited in the computer code, and my experiments with the solution procedure show that, in most cases, the problem is well-behaved. In other words, the algorithm typically converges within a short time and with few iterations[8].

4. Optimal smoothness

An optimal value of λ must take both bias and variance into account. Consider the example of risk, $R(\lambda) = 1 / n\, \mathrm{E}[(\mathbf{pv} - \widehat{\mathbf{pv}})'\mathbf{W}(\mathbf{pv} - \widehat{\mathbf{pv}})]$, and *predictive risk*, $P(\lambda) = 1 / n\, \mathrm{E}[(\mathbf{p}^* - \widehat{\mathbf{pv}})'\mathbf{W}(\mathbf{p}^* - \widehat{\mathbf{pv}})]$. The risk is the weighted expected squared distance between the true vector of present values, **pv**, and the estimate, $\widehat{\mathbf{pv}}$, while the predictive risk is the weighted expected squared distance between the estimate, $\widehat{\mathbf{pv}}$, and a hypothetical new sample vector, $\mathbf{p}^*$. Risk is related to predictive risk by

$$P(\lambda) = \sigma^2 + R(\lambda) \tag{17}$$

Predictive risk is a natural criterion for estimation of λ, and, when σ^2 is known, minimizing risk is equivalent to minimizing predictive risk. Unfortunately, σ^2 is not known, and minimizing risk requires at least a good estimator of σ^2. Consequently, alternatives are needed.

The $n \times n$ matrix, $\bar{\mathbf{S}}_\lambda$, which maps the price vector p into smoothed (or predicted) present values, $\widehat{\mathbf{pv}}$, will be called the *smoother matrix*, i.e.

$$\widehat{\mathbf{pv}} = \bar{\mathbf{S}}_\lambda \mathbf{p} \tag{18}$$

For ϕ, defined by (14), $\bar{S}_\lambda$ is

$$\bar{\mathbf{S}}_\lambda = \mathbf{B}(\lambda n\mathbf{G} + \mathbf{B}'\mathbf{W}\mathbf{B})^{-1}\mathbf{B}'\mathbf{W} \tag{19}$$

which means that the smoothing procedure is linear in data. However, this result does not apply to the nonlinear smoothers defined by (16), where $\bar{\mathbf{S}}_\lambda$ depends on ϕ. Nonetheless, in these cases, $\bar{\mathbf{S}}_\lambda$ will be defined by (19), and an estimate can be obtained from the last iteration of the estimation procedure. The equality of $\widehat{\mathbf{pv}}$ and $\bar{\mathbf{S}}_\lambda\mathbf{p}$ is then only valid as an approximation[9].

The problems in determination of λ can be further illustrated by means of the following well-known expression for risk:

$$R(\lambda) = \frac{1}{n}\mathbf{pv}'(\mathbf{I} - \bar{\mathbf{S}}_\lambda)'\mathbf{W}(\mathbf{I} - \bar{\mathbf{S}}_\lambda)\mathbf{pv} + \sigma^2\frac{1}{n}\,\mathrm{tr}\,[\bar{\mathbf{S}}_\lambda\bar{\mathbf{S}}_\lambda] \tag{20}$$

or

$$R(\lambda) = \frac{1}{n}b'_{\widehat{\mathbf{pv}}}(\lambda)\mathbf{W}b_{\widehat{\mathbf{pv}}}(\lambda) + \frac{1}{n}tr\,[\mathbf{W}\mathrm{V}(\widehat{\mathbf{pv}})] \tag{21}$$

where $b_{\widehat{\mathbf{pv}}} = (\bar{\mathbf{S}}_\lambda - \mathbf{I})\mathbf{pv}$ is the bias and $\mathrm{V}(\widehat{\mathbf{pv}}) = \sigma^2\bar{\mathbf{S}}'_\lambda\mathbf{W}^{-1}\bar{\mathbf{S}}_\lambda$ is the variance of $\widehat{\mathbf{pv}}$. Both the bias and variance components of the risk react in the same direction when λ is varied.

In all but trivial and useless cases ($\mathbf{pv} = \mathbf{0}$ or $\lambda = 0$), $\widehat{\mathbf{pv}}$ is a biased estimator of $\mathbf{pv}$. However, bias can be asymptotically reduced by letting λ tend to zero at a rate faster than n. At the same time, however, the variance will increase, which will leave us with the classical problem of balancing bias against variance.

A criterion for the estimation of λ which does not require knowledge of σ^2 is the GCV criterion. First, consider the mean squared error, $\mathrm{MSE}(\lambda) = \frac{1}{n}(\mathbf{p} - \widehat{\mathbf{pv}})'\mathbf{W}(\mathbf{p} - \widehat{\mathbf{pv}})$, with expected value

$$E(\mathrm{MSE}) = \frac{1}{n}\mathbf{pv}'(\mathbf{I} - \bar{\mathbf{S}}_\lambda)'\mathbf{W}(\mathbf{I} - \bar{\mathbf{S}}_\lambda)\mathbf{pv} + \sigma^2 \frac{1}{n}\,\mathrm{tr}\,[(\mathbf{I} - \bar{\mathbf{S}}_\lambda)(\mathbf{I} - \bar{\mathbf{S}}_\lambda)] \tag{22}$$

or

$$E(\mathrm{MSE}) = \frac{1}{n}\mathbf{pv}'(\mathbf{I} - \bar{\mathbf{S}}_\lambda)'\mathbf{W}(\mathbf{I} - \bar{\mathbf{S}}_\lambda)\mathbf{pv} + \sigma^2 + \sigma^2 \frac{1}{n}\,\mathrm{tr}\,[\bar{\mathbf{S}}\mathrm{f}_\lambda \bar{\mathbf{S}}_\lambda] - 2\sigma^2 \frac{1}{n}\,\mathrm{tr}\,[\bar{\mathbf{S}}_\lambda] \tag{23}$$

In equations (22) and (17), the first three terms equal the risk. Thus, MSE(λ) is a (downward) biased estimator of the predictive risk, $P(\lambda)$, with bias, $-2\sigma^2 \frac{1}{n} tr\,[\bar{\mathbf{S}}_\lambda]$, and the GCV statistic, as defined by

$$\mathrm{GCV}(\lambda) = \frac{\mathrm{MSE}(\lambda)}{(\mathrm{tr}[\mathbf{I} - \bar{\mathbf{S}}_\lambda]/n)^2} \tag{24}$$

can be seen as a bias-corrected mean squared error. Under the assumptions of the GCV theorem (Eubank, 1988, theorem 2.1) originally by Craven and Wahba (1979), the GCV statistic is an approximately unbiased estimator of predictive risk. The GCV theorem, as stated in Eubank (1988), does not strictly apply here. However, such details will not be discussed further.

A detailed discussion of GCV and related criteria can be found in textbooks on nonparametric regression (Härdle, 1990; Hastie and Tibshirani, 1990; Eubank, 1988).

GCV is a convex function of λ, and is easily calculated as a by-product of the smoothing procedure. λ can therefore be estimated by a *golden section search* in one dimension (Press, Teukolsky, Vetterling and Flannery, 1992, chapter 10) for the minimum of the GCV function. Implicit in this procedure is thus the hope that the minimizer of GCV is also the minimizer of predictive risk.

The problem of the determination of the smoothness parameter, λ, can also be seen as equivalent to the determination of the number of parameters in a parametric curve-fitting model. If $\widehat{\mathbf{pv}}$ was in fact estimated from a parametric model with k parameters, then the second term in (20) and (21) would be $\sigma^2 k/n$. This leads to the definition of $\kappa = \mathrm{tr}\,(\bar{\mathbf{S}}_\lambda \bar{\mathbf{S}}_\lambda)$ as an equivalent number of parameters. In this sense, a nonparametric estimator of $\widehat{\mathbf{pv}}$ with, say, an equivalent number of parameters of $\kappa = 4$, uses the same degrees of freedom as a cubic polynomial—or any other parametric curve—with 4 parameters. Thus, a simple—and in many ways satisfactory—way of determining λ is by setting some fixed value for κ and then carrying out a simple search for the corresponding value of λ. In all but very ill-behaved cases, there is a non-increasing relationship between λ and κ, which facilitates the search for λ.

To sum up, there are (at least) two methods for determination of λ. One is a direct estimation method based on the GCV criterion. The other is based on setting an equivalent number of parameters, and searching for the corresponding value of λ. Both methods are used in the empirical application to the Danish term structure discussed in the next section.

5. An empirical application

A set of monthly sampled Danish government bond prices will be used for the empirical example. The period covers 120 days from 1985 through 1994. Danish government bonds are non-callable, and by excluding low coupon bonds, the effects of taxation can also be ignored[10]. Historically, officially listed prices have been very noisy, and a good estimation procedure must also be able to deal with that problem. However, in later years, and especially since the introduction of electronic trading on the Copenhagen Stock Exchange (CSE), the problem has diminished[11].

Figure 1 shows two different nonparametric estimates of the yield curve. One (solid line) is derived by smoothing the function $u(t)$ (equation (3)), while the other (broken line) is found by direct smoothing of the yield curve, $y(t)$ (equation 2). For both curves, the smoothness parameter, λ, was determined by minimizing the GCV criterion. The difference between the two smoothed curves is striking, however. Direct smoothing of the yield

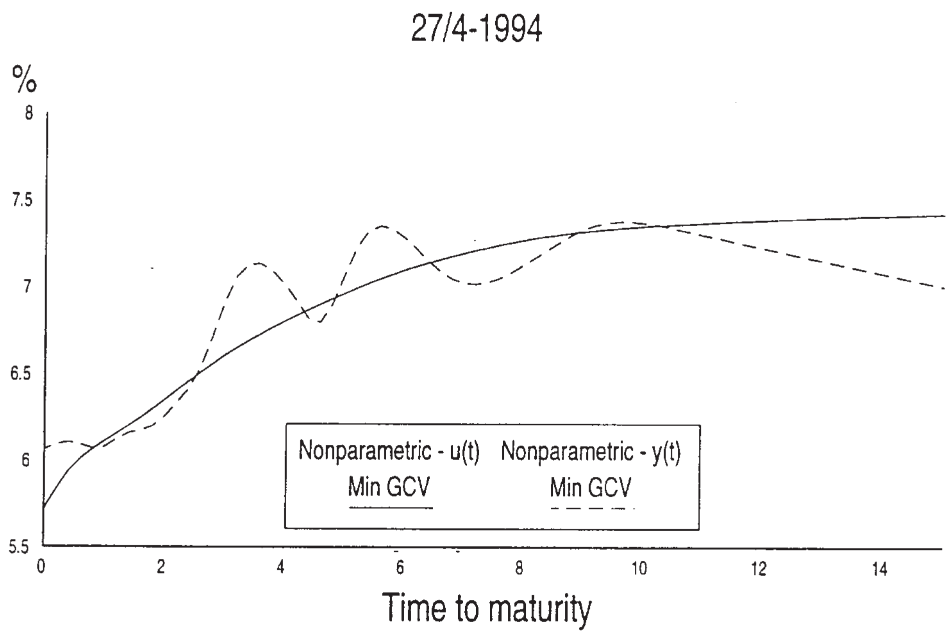

Figure 1. Two different nonparametric yield curve estimates

curve $y(t)$ greatly under-smoothes the data (the equivalent number of parameters is $\kappa = 8.27$). In contrast, the yield curve derived from $u(t)$ is much smoother with an equivalent number of parameters of $\kappa = 4.06$.

It is important to remember that under smoothing is a common problem in any application of nonparametric regression to small data sets with smoothness determined by the GCV criterion (see Hastie and Tibshirani (1990), section 3.4), and for this particular trading day (April 27, 1994) the data set is very small indeed ($n = 16$). A closer inspection reveals that the data are also noisy, and at least one observation can be considered an outlier. Figure 2 shows the smoothed yield curves after the outlier has been removed. Now both curves are similar, except for maturities longer than 10 years. However, since the time to maturity of the longest bond in the sample is 10.62 years, the yield curve beyond 10.62 years is pure extrapolation. Figure 2 thus supports one of the conjectures of section 3, namely that smoothing $u(t)$ has a stabilizing effect on the long end of the yield curve. Nevertheless, there are still signs of under smoothness when $y(t)$ is smoothed. In this case, however, removing the outlier has in fact decreased κ from 8.27 to 6.26.

For the full period, 1985–94, as well as the subperiod, 1990–94, smoothing $u(t)$ generally produces a smoother curve. This is further confirmed by the summary of the distribution of κ in tables 1 and 2. In these tables, the smoothness parameter, λ, is determined by GCV. The difference in smoothness between the two curves is even more manifest in the subperiod 1990–94.

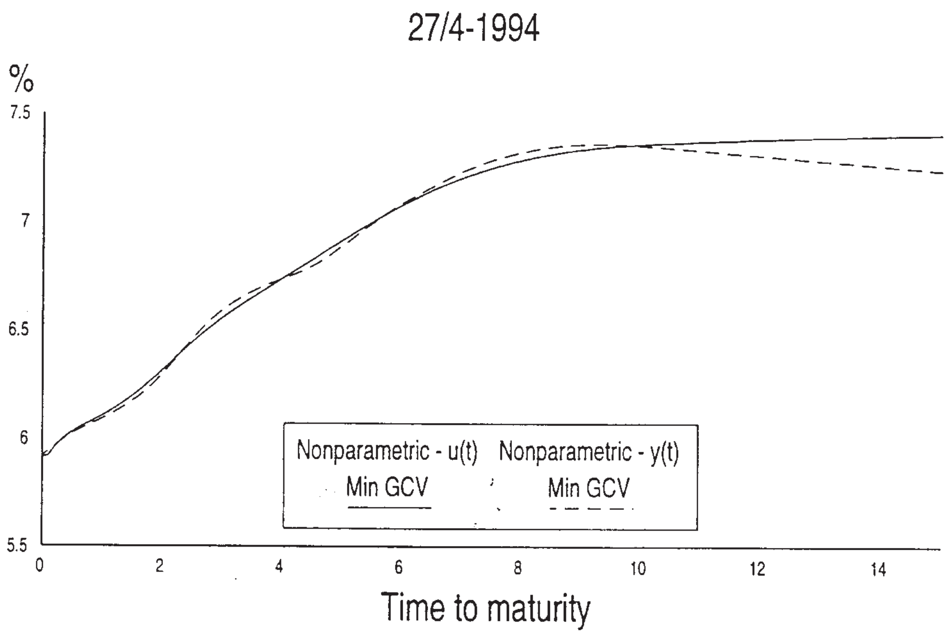

Figure 2. Yield curves when outliers are removed

Table 1. Distribution of equivalent number of parameters for two different nonparametric models in two periods.

Period 1985–1994	$\kappa \leq 3.5$	$3.5 < \kappa \leq 5.5$	$5.5 < \kappa \leq 7.5$	$7.5 < \kappa$
Smoothing $u(t)$	17	74	10	19
Smoothing $y(t)$	4	44	41	31
Period 1990–1994	$\kappa \leq 3.5$	$3.5 < \kappa \leq 5.5$	$5.5 < \kappa \leq 7.5$	$7.5 < \kappa$
Smoothing $u(t)$	15	44	1	0
Smoothing $y(t)$	1	23	20	16

Smoothness parameter, λ; estimated by minimum GCV.

Table 2. Summary statistics of κ in the full period and in the subperiod, 1990–94.

Period 1985–1994	Average κ	Median κ
Smoothing $u(t)$	4.95	4.28
Smoothing $y(t)$	6.13	5.82
Period 1990–1994	Average κ	Median κ
Smoothing $u(t)$	3.98	3.85
Smoothing $y(t)$	6.18	5.89

Smoothness parameter, λ, estimated by minimum GCV.

Table 3. Ranking of nonparametric yield curve models according to GCV and MSE.

1985–94	Smoothness	Ranked 1 (GCV)	Av. GCV	Med. GCV
Smoothing $u(t)$	min GCV	88	5.23	3.55
Smoothing $y(t)$	min GCV	32	5.72	3.84
1985–94	Smoothness	Ranked 1 (MSE)	Av. MSE	Med. MSE
Smoothing $u(t)$	$\kappa = 4$	103	3.56	2.43
Smoothing $y(t)$	$\kappa = 4$	17	6.59	3.82
1990–94	Smoothness	Ranked 1 (GCV)	Av. GCV	Med. GCV
Smoothing $u(t)$	min GCV	49	7.07	5.12
Smoothing $y(t)$	min GCV	11	7.87	5.85
1990–94	Smoothness	Ranked 1 (MSE)	Av. MSE	Med. MSE
Smoothing $u(t)$	$\kappa = 4$	58	4.55	3.41
Smoothing $y(t)$	$\kappa = 4$	2	10.16	6.43

The Rank 1 column shows how many times the corresponding model is ranked as number 1 (according to the criterion in brackets). The other two columns show the average and median values of MSE and GCV (multiplied by 100).

Despite the difference in model type and choice of smoothing parameter the two models can be compared by the GCV criterion[12]. Table 3 indicates that—judged by GCV—smoothing $u(t)$ is superior to direct smoothing of $y(t)$.

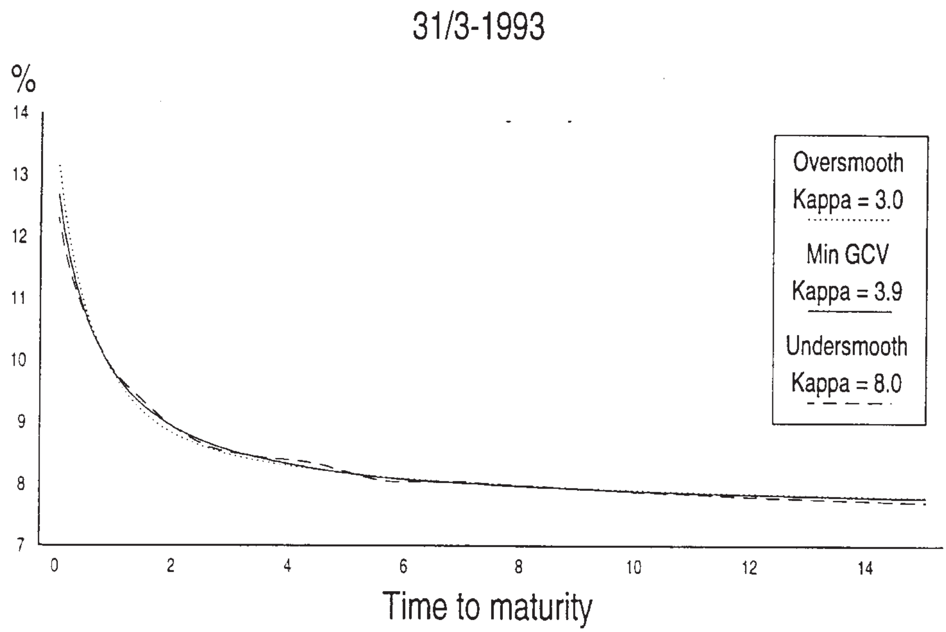

Figure 3. Three yield curves with different smoothness.

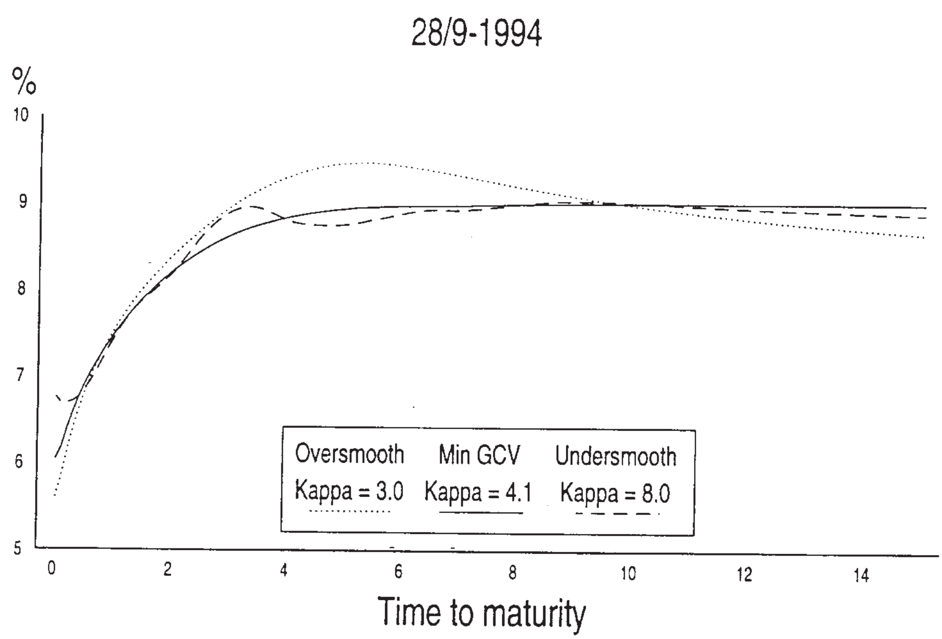

Figure 4. Three yield curves with different smoothness.

Smoothness can also be controlled by finding a λ corresponding to some prefixed value of κ (e.g. 4). In this case, MSE is used to compare the two models. As table 3 shows, the superiority of smoothing $u(t)$ is even more striking in this case.

The next example illustrates the use of GCV for estimating λ. The minimum GCV value of κ on the trading day, March 31, 1993 is 3.9, and, as figure 3 shows, there is no indication of excess or lack of smoothing of the data. Changing the equivalent number of parameters to 8 and 3 gives only a slight change in the estimated yield curves. On the trading day, September 28, 1994, however, the situation is different. The minimum GCV value of κ is now 4.1, which seems to give a reasonable degree of smoothness (figure 4). Changing κ to 3 or 8 has a dramatic impact on the smoothness of the yield curve, however. A further interpretation of the example is shown by figure 5. For the trading day, September 28, 1994, the GCV curve has a well-defined minimum at $\kappa = 4.1$. In contrast, the GCV curve for the trading day March 31, 1993, is almost flat. A flat GCV curve cannot be used to discriminate between alternative values of λ.

The final example consists of a comparison of the nonparametric procedure with three different parametric models:

- The Cox et al. (1985) model (CIR)
- The Nelson and Siegel (1987) model (NS)
- A regression spline model of the yield curve

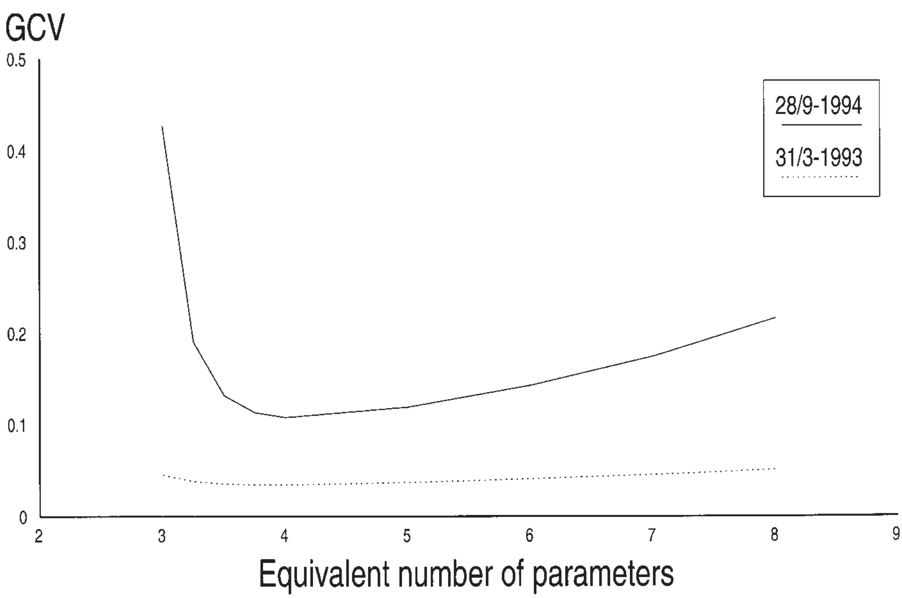

Figure 5. GCV as a function of equivalent number of parameters, $\kappa(\lambda)$

The NS and CIR models have 4 parameters each, while the regression spline model can have any number of parameters—one for each knot. Regression spline models of the *yield curve* with 4 knots were selected. No restrictions were imposed on the long end of the term structure. This is because the parametric models only serve as benchmarks against which the nonparametric model is to be tested, and an unrestricted regression spline model—even with a comparable number of parameters—presumably performs better than a restricted model. The placement of the knots were chosen by rule of thumb, as suggested in McCulloch (1975).

Two cases are considered for the full period and the last subperiod. First, the nonparametric model is estimated with λ determined by the minimum GCV criterion, and the models are compared using the GCV criterion[13]. In the second case, the smoothness parameter is determined by setting the value of κ to 4. In this case, the competing models can be compared directly by the MSE. The results are summarized in table 4. The nonparametric model (NPM) performs very well in comparison with the parametric models. In the full period, NPM is ranked as number 1 on more than half of the days considered, while in the last subperiod the NPM performs slightly worse. However, it is still the model

Table 4. Comparison of parametric and nonparametric yield curve models.

1985–94	Smoothness	Rank 1 (GCV)	Av. GCV	Med. GCV
Smoothing $u(t)$	min GCV	64	5.23	3.55
NS	4 parameters	16	5.69	4.07
CIR	4 parameters	3	6.80	4.47
Regression spline	4 parameters	37	5.72	3.96
1985–94	**Smoothness**	**Rank 1 (MSE)**	**Av. MSE**	**Med. MSE**
Smoothing $u(t)$	$\kappa = 4$	77	3.56	2.43
NS	4 parameters	14	3.98	2.83
CIR	4 parameters	2	4.71	3.25
Regression spline	4 parameters	27	4.02	2.77
1990–94	**Smoothness**	**Rank 1 (GCV)**	**Av. GCV**	**Med. GCV**
Smoothing $u(t)$	min GCV	25	7.06	5.12
NS	4 parameters	12	7.08	4.71
CIR	4 parameters	3	8.67	4.98
Regression spline	4 parameters	20	7.60	4.69
1990–94	**Smoothness**	**Rank 1 (MSE)**	**Av. MSE**	**Med. MSE**
Smoothing $u(t)$	$\kappa = 4$	35	4.55	3.41
NS	4 parameters	10	4.81	3.37
CIR	4 parameters	2	5.81	3.47
Regression spline	4 parameters	13	5.23	3.43

The Rank 1 column shows how many times the corresponding model is ranked as number 1 (according to the criterion in brackets). The other two columns show the average and median values of MSE and GCV (multiplied by 100).

with the highest percentage of number one rankings. On the whole, table 4 confirms the general assumption of this article: nonparametric cubic spline smoothing out-performs the parametric models.

Further evidence, of this superiority is presented in figure 6. In contrast with the NPM, the Nelson and Siegel model is unable to capture the shape of the term structure on the trading day, September 30, 1992. This becomes especially clear by comparison with the actual bond yields in the short end[14].

6. Conclusion

This article has presented evidence that nonparametric smoothing is a feasible solution to the problem of inferring the zero-coupon bond yield curve from noisy data on coupon bonds. The computational problems are surmountable, despite the absence of an analytical solution. The article has also presented evidence that nonparametric smoothing is preferable to parametric models. Finally, it has documented that smoothing a simple transformation of the yield curve is less affected by outliers than direct smoothing of the yield curve.

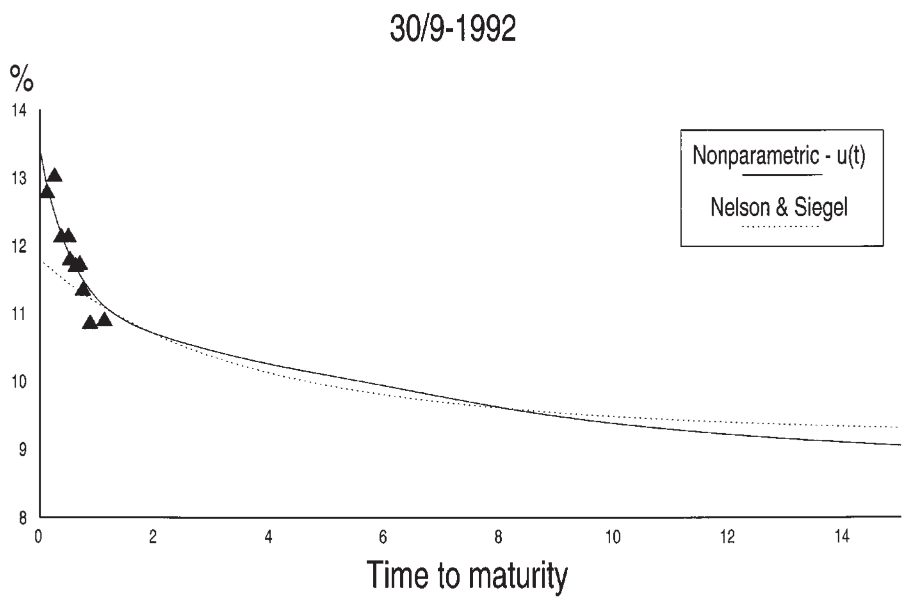

Figure 6. Comparison of nonparametric and parametric (NS) model.

Acknowledgment

A preliminary version was presented at the *IFAC Workshop on Computational Methods in Economics and Finance*, Amsterdam, June 8–10, 1994. Helpful comments and suggestions from Leif Andersen, Tom Engsted, Svend Jakobsen, Jesper Lund, John Huston McCulloch, and a referee are gratefully acknowledged, as is the financial support from the *Danish Social Science Research Council*.

Notes

1. The precise meaning of this statement will become clear from the discussion in section 3.
2. John Huston McCulloch—in private communication—has suggested estimation of the curve, $q(t) = ty(t)$. The first-order derivative of $q(t)$ is the instantaneous forward curve. Moreover, q has the same extrapolative behavior as the curve, $u(t)$, suggested here. However, using q necessitates a further restriction, $q(0) = 0$, in order for the yield curve to be defined in terms of q. Furthermore, there are no a priori statistical arguments in favor of $q(t)$ rather than $u(t)$. Finally, standard computer packages for calculation of cubic spline smoothers do not allow restrictions of the type, $q(0) = 0$. This model will not, therefore, be discussed in this article.
3. In the empirical application to the Danish term structure in section 5 below $w_i = t_{m_i}^{-1}$. This specification is consistent with the one suggested by Chambers et al. (1984). A similar specification is suggested by Vasicek and Fong (1982). My experience with the model also shows that the resulting yield curve is fairly unaffected by minor changes in the variance specification.
4. An appendix with further details is available on request. The proof of this result requires some extra restrictions on φ, all of which are fulfilled by the models considered here (equations (1), (2) and (3)).
5. In principle, the interval $[a, b]$ can be chosen freely as long as it contains the knots. In the term structure fitting problem, a natural choice of a is $a = 0$. Because φ is linear outside $[t_1, t_m]$, it can be extrapolated to infinity by a linear function with slope equal to $\varphi'(t_m)$. It is in this sense that the solution to the smoothing problem is said to be asymptotically linear as $t \to \infty$.
6. Readers not familiar with the notation of vector-derivatives can consult Magnus and Neudecker (1988).
7. An appendix with further details is available on request.
8. In the application on the Danish term structure discussed in section 5 below, m varies between 47 and 91, with an average of 61.2. The sample size, n, varies between 14 and 36, with an average of 26.2. The average computing time was 2.73 minutes on a 66 MHz PC (i486). This time estimate includes the time used for the search of the optimal smoothness parameter, λ.
9. Similar approximations will be used below without further comment.
10. The lower limit for the coupon varies with the general level of interest rates. Details, and a copy of the data set, is available on request.
11. Electronic trading on the CSE started in the late 1980s. This, and changes in the definition of officially listed prices, is the reason for the special consideration of the subperiod, 1990–94, below.
12. There is no solid theoretical basis for this statement. However, GCV is very much related to information criteria (AIC, BIC, etc.), which are often used in model selection (see Eubank (1988), section 2.4).
13. GCV can also be calculated for parametric models.
14. The coupon effect is the reason for not plotting the longer-term bond yields in figure 6. For bonds with time to maturity of less than 1 year, the coupon effect is either absent or can be ignored.

References

Chambers, D. R., W. T. Carleton, and D. W. Waldman. A new approach to estimation of the term structure of interest rates, *Journal of Financial and Quantitative Analysis*, 19, 233–269 (1984).

Cox, J., J. Ingersoll, and S. Ross. A theory of the term structure of interest rates, *Econometrica*, 53, 385–407 (1985).

Craven, P. and G. Wahba. Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross validation, *Numerische Mathematik*, 31, 377–403 (1979).

Delgado, M. A. and P. M. Robinson. Nonparametric and semiparametric methods for economic research, *Journal of Economic Surveys*, 6, 201–249 (1992).

Eubank, R. L. *Spline Smoothing and Nonparametric Regression*, Vol. 90 of *Statistics, textbooks and monographs*, Marcel Dekker (1988).

Härdle, W. *Applied Nonparametric Regression*, Cambridge University Press (1990).

Hastie, T. J. and R. J. Tibshirani. *Generalized Additive Models*, Vol. 43 of *Monographs on Statistics and Probability*, Chapman and Hall (1990).

Magnus, J. R. and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Series in Probability and Mathematical Statistics, John Wiley & Sons (1988).

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

McCulloch, J. H. The tax-adjusted yield curve, *The Journal of Finance*, 30, 811–830 (1975).

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

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*, Cambridge University Press (1992).

Robinson, P. M. Semiparametric econometrics: A survey, *Journal of Applied Econometrics*, 3, 35–51 (1988).

Shea, G. S. Pitfalls in smoothing interest rate term structure data: Equilibrium models and spline approximations, *Journal of Financial and Quantitative Analysis*, 19, 253–269 (1984).

Vasicek, O. A. and H. G. Fong. Term structure modeling using exponential splines, *The Journal of Finance*, 37, 339–356 (1982).