Estimating Term Structure Using Nonlinear Splines: A Penalized Likelihood Approach

[1]Yoshinori Kawasaki and [2]Tomohiro Ando

[1]The Institute of Statistical Mathematics, Research Organization of Information and Systems
4–6–7 Minami-Azabu, Minato-ku, Tokyo 106–8569, Japan. E-Mail: kawasaki@ism.ac.jp
[2]Graduate School of Business Administration, Keio University,
2-1-1 Hiyoshi Honcho, Kouhoku-ku, Yokohama, Kanagawa 223-8523, Japan.

Keywords: *yield curve, penalized likelihood, nonlinear functional, smoothing parameter, information criteria*

EXTENDED ABSTRACT

The spline-based models are widely used in practice to estimate the term structure of interest rates from a set of observed coupon-bond prices. The most popular method can be traced back to McCulloch (1971). Assuming that the price of a bond is equal to the present value of its future coupon payments and redemption, cash flows are regressed on a set of basis functions to estimate discount functions. Once the discount function is estimated, the zero-coupon yield and the forward rate can be obtained by transformations of the discount function. Though this method was followed by a lot of researchers, some serious drawbacks have been reported.

The most important problem is the instability of estimated yield curves. As is widely known, the discount function $\delta(t)$, the zero coupon yield $\eta(t)$ and the instantaneous forward rate $f(t)$ are closely tied with one another by explicit relationships. McCulloch's method gives approximated discount function first, so the zero coupon yield curve and the forward rate curve can be derived once $\delta(t)$ is estimated. The problem is, however, seemingly reasonable estimate of the discount function does not always lead to acceptable shapes of yield curves, especially for the forward rate curve. Some researchers are concerned with the choice of basis functions when defining a spline function, while others question how to place knots efficiently.

The choice of basis functions and/or knot locations undoubtedly affects the estimation results. However, the present article focuses on a different point. It is considered here that instability of the estimated yield curves is caused by the ill-posed nature of the regression spline, rather than by the inappropriate choice of the basis function. By ill-posed it is meant that a model may be over-parameterized compared to the amount of sample information. Without a addressing this ill-posed nature specifically, any modification of the choice of basis functions, approximating functional forms, or knots placement may provide only minor improvements.

Throughout this article, a penalty term is added to the original log-likelihood of a yield curve model, that is, a penalized likelihood approach is adopted for this treatment. In this sense, the work of Fisher, Nychka and Zervos (1995) is the most closely related and influential to this study. Those authors fitted smoothing splines (with B-splines bases) instead of regression splines. Moreover, Fisher, Nychka and Zervos (1995) fitted smoothing spline to the zero coupon yield and the forward rate as well as to the discount function. Their simulation results suggest that the best way to estimate yield curves is to place spline bases on the forward rate curve. However, splining the forward rate or the zero coupon yield is not linear operation, hence the use of GCV or the effective number of parameters in model selection lacks its theoretical foundation.

This paper proposes a penalized likelihood approach accompanied by generalized information criteria (GIC) that determine the desired degree of smoothness of yield curves in a data-dependent way. GIC, proposed by Konishi and Kitagawa (1996), is an extension of AIC, Akaike Information Criterion. Originally AIC is proposed on the assumption that the models to be compared are estimated by the method of maximum likelihood. GIC is extended to the cases where the models are not necessarily estimated by ML. Model selection among penalized (nonlinear) regressions comes within the range of GIC. Our approach is theoretically valid even if the regression functional is nonlinear with respect to the unknown coefficients of basis functions, of which typical case is 'splining the forward rate' or 'splining the zero coupon yield' case. As will be shown in Section 2, these cases are reduced to the problems of nonlinear regression spline. The derived GICs enable us to compare the models with various choices of basis functions under different regression functional forms in a unified manner. In addition, the number of basis function can be chosen based in minimum GIC method. Monte Carlo simulations reveal that choosing the appropriate number of bases by GIC reduces MSE rather than controlling a plenty of bases by a single smoothing parameter.

1 INTRODUCTION

There have been a number of studies attempting to establish an excellent technique for estimating the term structure of interest rates from a cross-section of coupon bond prices. Under the assumption that the price of a bond is equal to the present value of its future coupon payments and redemption, McCulloch (1971) regressed cash flows on a set of basis functions to estimate discount functions. Once the discount function is estimated, the zero-coupon yield and the forward rate can be obtained by transformations of the discount function.

Although the approach adopted by McCulloch (1971, 1975) was followed by several related studies, the approach has been criticized on a number of points. The central issue has been the instability of regression spline. Hence, throughout this article, a penalty term is added to the original log-likelihood of a yield curve model, that is, a penalized likelihood approach is adopted for this treatment. In this sense, the work of Fisher, Nychka and Zervos (1995) is the most closely related and influential to this study.

One important assertion made by Fisher, Nychka and Zervos (1995) based on their simulation studies is that smoothing splines could be used to spline an arbitrary transformation of the discount function. Their simulation results suggest that the best way to estimate yield curves is to place spline bases on the forward rate curve. As will be described soon below, however, splining the forward rate or the zero coupon yield is doubly nonlinear; the regression functional is nonlinear with respect to the coefficients on the spline bases, and the basis function is also nonlinear with respect to maturity, t. For such a case, the use of GCV or the effective number of parameters lacks its theoretical foundation.

It is widely known that the discount function $\delta(t)$ and the instantaneous forward rate $f(t)$ are related by

$$f(t) = -\delta'(t)/\delta(t), \tag{1}$$

where $\delta'(t)$ is the derivative of the discount function $\delta(\cdot)$ evaluated at the point t. The zero coupon yield $\eta(t)$ is tied to the discount function $\delta(t)$ by

$$\eta(t) = -\ln(\delta(t))/t. \tag{2}$$

See for example Anderson, Breedon, Deacon and Derry (1996) for the derivations of these relationships. Hence, we do not have to start by approximating the discount function $\delta(t)$. From equations (1) and (2), it is recognized that if splines are placed on $\eta(t)$ or $f(t)$, then $\delta(t)$ will be expressed as an exponential function with an approximating function for $\eta(t)$ or $f(t)$ as its argument. That is, splining the zero coupon yield or the forward rate is equivalent to exponential splining of the discount function.

By fitting a smoothing spline with cubic B-spline bases, Fisher et al. compared all three options; splining $\delta(t)$, $\eta(t)$ and $f(t)$, and determined the roughness penalty using GCV in all three cases. From a theoretical viewpoint, however, the application of GCV is questionable except the case of splining $\delta(t)$. As point out themselves (see footnote 10 and appendix B in Fisher et al. 1995), GCV cannot be applied unless the regressor is expressed as a linear combination of basis functions. In other words, a basis function can be nonlinear in t as is usual with many nonparametric regression schemes, but the regression functional should be linear with respect to the unknown parameters for the use of GCV. Clearly this does not hold in splining $\eta(t)$ or $f(t)$. Supposing that the splined zero coupon yield $\eta_s(t)$ is expressed as $\eta_s(t) = \sum w_k \phi_k(t)$, where $\{\phi_k(t); k = 1, 2, ...\}$ is a set of spline bases with coefficients w_k, then (2) implies that the splined discount function $\delta_s(t)$ is expressed as $\delta_s(t) = \exp(-t \sum w_k \phi_k(t))$. Here, δ_s is clearly not linear in $\{w_k\}$.

Sharing the motivation of Fisher et al., the aim of the present study is to propose a theoretically valid criterion that enables us to determine the desired level of smoothness of yield curves in a data-dependent way even when the regression functional is not always linear with respect to the unknown parameters. Therefore the models considered in this article are all penalized or regularized in principle. In this treatment, the generalized information criteria (GIC) introduced by Konishi and Kitagawa (1996) are tailored to various cases. Use of the GIC also makes it possible to choose the optimal number of basis functions. This is an important feature, as allowing excess knots can lead to an undesirable shape of the forward rate function. Selection of the appropriate number of basis by an objective criterion is therefore desirable.

2 PENALIZED LIKELIHOOD APPROACH

2.1 Bond equation

Consider a set of n bonds traded on one day. Let p_α be the price of bond α, c_α be its coupon payment, which is paid at time t_1^α,...,$t_{L_\alpha}^\alpha$, let R_α be the redemption payment, and let L_α be the number of remaining payments. Following the theory of bond pricing (McCulloch, 1971), we assume that the price of a bond (plus accrued interest a_α) is equal to the present value of its future coupon payments and the redemption, i.e., for $\alpha = 1, ..., n$,

$$p_\alpha + a_\alpha = c_\alpha \sum_{k=1}^{L_\alpha} \delta(t_k^\alpha) + R_\alpha \delta(t_{L_\alpha}^\alpha) + \varepsilon_\alpha, \tag{3}$$

where $\delta(\cdot)$ is the discount function, ε_α are independent and normally distributed with mean of zero and variance σ^2. The discount function $\delta(t)$ gives the present value of a monetary unit, e.g., \$1.00 after t years. Most researchers follow McCulloch (1971) in explicitly constraining cash flows from different bonds due at the same time to be discounted at the same rate, and estimate the discount function $\delta(\cdot)$ from which the other yield curves can be derived.

If splines are placed on the discount function, $\delta(\cdot)$ is expressed as a linear combination of a set of m underlying basis functions, as follows.

$$\delta(t; \boldsymbol{w}) = 1 + \sum_{k=1}^{m} w_k \phi_k(t) = 1 + \boldsymbol{w}' \boldsymbol{\phi}(t), \qquad (4)$$

where $\boldsymbol{\phi}(t) = (\phi_1(t), ..., \phi_m(t))'$ is an m-dimensional vector constructed from a set of basis functions $\{\phi_j(t); j = 1, \cdots, m\}$, and $\boldsymbol{w} = (w_1, ..., w_m)'$ is an unknown parameter vector to be estimated. It follows from equations (3) and (4) that the bond price model based on a linear combination of basis functions is as follows.

$$f_B(y_\alpha | \boldsymbol{t}_\alpha; \boldsymbol{w}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(y_\alpha - \boldsymbol{c}_\alpha' \Phi_\alpha \boldsymbol{w})^2}{2\sigma^2} \right\}, \quad (5)$$

where $\boldsymbol{t}_\alpha = (t_1^\alpha, ..., t_{L_\alpha}^\alpha)'$ is the vector of the points of time at which payments occur, $y_\alpha = p_\alpha + a_\alpha - L_\alpha c_\alpha - R_\alpha$, $\Phi_\alpha = (\boldsymbol{\phi}(t_1^\alpha), ..., \boldsymbol{\phi}(t_{L_\alpha - 1}^\alpha), \boldsymbol{\phi}(t_{L_\alpha}^\alpha))'$ and $\boldsymbol{c}_\alpha = (c_\alpha, ..., c_\alpha, c_\alpha + R_\alpha)'$, respectively. This specification is very convenient for parameter estimation because the functional form of (4) is linear with respect to the unknown parameters.

A number of functional forms have been proposed for the basis functions $\phi_j(t)$. See McCulloch (1971, 1975), Schaefer (1981), Mastronikola (1991), Steeley (1991). Throughout this paper, we consider B-spline only. Also note that the argument so far is still on the framework of regression spline. We will soon move on to smoothing spline in 2.3 after introducing nonlinear spline models in the next section.

2.2 Nonlinear spline

It is also possible to place spline on zero coupon yield or even on forward rate rather than on discount function. Langetieg and Smoot (1989) fitted a cubic B-spline to the zero coupon yield, a technique they refer to as the exponential yields model. In this model, $\{\phi_k(t)\}_{k=1}^m$ are cubic B-spline bases. From equation (2), $\eta(t) = \sum w_k \phi_k(t)$ implies $\delta(t) = \exp(-t \sum w_k \phi_k(t))$. Hence, splining the zero coupon yield is equivalent to fitting an exponential spline model to the discount function:

$$\delta(t; \boldsymbol{w}) = \exp\left(-t \sum_{k=1}^{m} w_k \phi_k(t) \right). \qquad (6)$$

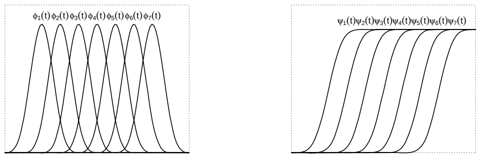

Figure 1. Basis functions for the B-spline and its integral

They found that their model gave better results than the exponential spline specification of Vasicek and Fong (1982), and argued that it is not surprising since the exponential transformation model can be viewed as an approximation of the exponential yields model.

Fisher, Nychka and Zervos (1995) suggest the placement of a B-spline on the forward rate curve:

$$f(t) = \sum_{k=1}^{m} w_k \phi_k(t). \qquad (7)$$

From the equation (1), (7) can be rewritten as

$$\delta(t; \boldsymbol{w}) = \exp\left\{ -\sum_{k=1}^{m} w_k \psi_k(t) \right\} \qquad (8)$$

where $\psi_k(t) = \int_0^t \phi_k(s) ds$. The functional form in (8) resembles the exponential spline specification (6), but the choice of basis function is different. Figure 1 shows the definition of a cubic B-spline basis over six equally spaced knots and the corresponding integral. As is clearly seen from Figure 1, the difference between (6) and (8) is that the basis functions are monotone in the model (8). This monotonicity is seemingly advantageous in approximating $\delta(t)$ which is expected to be monotonically decreasing in t.

Combining the equation (3) with either (6) or (8) as a discount function, we obtain a slightly different form of the bond pricing model, as follows.

$$f_E(y_\alpha | \boldsymbol{t}_\alpha; \boldsymbol{w}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(y_\alpha - \boldsymbol{c}_\alpha' \boldsymbol{\delta}(\boldsymbol{t}_\alpha; \boldsymbol{w}))^2}{2\sigma^2} \right\}, (9)$$

where $y_\alpha = p_\alpha + a_\alpha$, $\boldsymbol{\delta}(\boldsymbol{t}_\alpha; \boldsymbol{w}) = (\delta(t_1^\alpha; \boldsymbol{w}), ..., \delta(t_{L\alpha}^\alpha; \boldsymbol{w}))'$ is the discount vector, and $\boldsymbol{w} = (w_1, ..., w_m)'$ is an unknown parameter vector to be estimated from the data. An important point of (9), in comparison with (5), is that the discount function is not a linear combination of basis functions in fitting the B-spline for either $\eta(t)$ or $f(t)$. We will return to this point when we construct the model selection criterion.

2.3 Penalized likelihood

Here we present the maximum penalized likelihood method for estimating the unknown coefficients $\boldsymbol{w}$ and σ^2 in the bond price model (5) and (9). For parameter estimation in the bond price model (5), the maximum likelihood estimate of the weights is given explicitly by $\hat{\boldsymbol{w}} = (B'B)^{-1} B'\boldsymbol{y}$, where $B = (\Phi_1'\boldsymbol{c}_1, ..., \Phi_n'\boldsymbol{c}_n)'$, $\boldsymbol{y} = (y_1, ..., y_n)'$. In practice, however, the maximum likelihood method does not yield satisfactory results because the parameter estimates tend to be unstable and lead to overfitting. For example, suppose there is a hump in the estimated discount function, perhaps due to overfitting. No matter how small the hump, the derived forward rate may be negative at some maturity unless the discount function is non-increasing everywhere. The same problem can also arise in the nonlinear model (9) that model only guarantees the positiveness of the discount function.

These instabilities of the estimated yield curves all originate from the ill-posed nature of the regression spline, rather than from any inappropriate choice of the basis function. To avoid overfitting, a penalty term on the smoothness of the unknown coefficients is introduced into the log-likelihood. Specifically, we maximize

$$
\begin{aligned}
l_\lambda(\boldsymbol{w}, \sigma^2) &= \sum_{\alpha=1}^{n} \log f_{(\cdot)}(y_\alpha | \boldsymbol{t}_\alpha; \boldsymbol{w}, \sigma^2) \\
&- \frac{n\lambda}{2} \sum_{j=2}^{m} (\Delta^2 w_j)^2, \qquad (10)
\end{aligned}
$$

where λ is the smoothing parameter controlling the smoothness of the discount function, and $\Delta w_k = w_k - w_{k-1}$ is the difference operator.

Given λ and m, the unknown parameters $\boldsymbol{w}$ and σ^2 can be obtained as the solution of $\partial l_\lambda(\boldsymbol{w}, \sigma^2)/\partial \boldsymbol{w} = 0$ and $\partial l_\lambda(\boldsymbol{w}, \sigma^2)/\partial \sigma^2 = 0$. If these estimation equations can be solved explicitly in terms of $\boldsymbol{w}$ and σ^2, then by replacing the unknown parameters with the parameter estimates $\hat{\boldsymbol{w}}$ and $\hat{\sigma}^2$, we have the bond price model constructed by the penalized likelihood: $f_{(\cdot)}(y_\alpha | \boldsymbol{x}_\alpha; \hat{\boldsymbol{w}}, \hat{\sigma}^2)$.

If the bond model belongs to the f_B-class, it is always possible to give the maximum penalized estimates of $\boldsymbol{w}$ and σ^2 explicitly. Using an $(m-2) \times m$ difference matrix D_2 where $D_2'D_2$ makes a certain quadratic penalty matrix, the penalty term in (10) can be represented by $\sum_{j=2}^{m} (\Delta^2 w_j)^2 = \boldsymbol{w}'D_2'D_2\boldsymbol{w}$. Hence, for fixed λ and m, the maximum penalized likelihood estimates of $\boldsymbol{w}$ and σ^2 in the bond price model (5) are explicitly provided by

$$
\begin{aligned}
\hat{\boldsymbol{w}} &= (B'B + n\beta D_2'D_2)^{-1} B'\boldsymbol{y}, \\
\hat{\sigma}^2 &= \frac{1}{n} \sum_{\alpha=1}^{n} \{y_\alpha - \boldsymbol{c}_\alpha' \Phi_\alpha \hat{\boldsymbol{w}}\}^2, \qquad (11)
\end{aligned}
$$

where $\beta = \lambda\sigma^2$. Clearly an f_B-class model requires quite simple linear operations, hence much less computation is required. On the other hand, if the bond models belong to the f_E-class, explicit estimators are no longer available. In such a case, a numerical maximization procedure must be invoked. In this article, the Newton-Raphson method based on the first and second derivatives of the penalized likelihood function is adopted for estimation.

An important remaining problem is the criterion by which we should choose the smoothing parameter λ and the number of basis functions m. Here, we derive a criterion for evaluating the bond price model from an information-theoretic point of view. Once the criterion is established, the optimum roughness penalty λ and number of bases m are determined by searching the grid of ($\log\lambda$, m).

3 INFORMATION CRITERIA FOR MODEL EVALUATION

Akaike Information Criterion (AIC, Akaike (1974)) is proposed basically on the assumption that the models to be compared are estimated by the method of maximum likelihood. Konishi and Kitagawa (1996) extended AIC to the cases where the models are not estimated by ML, and proposed a framework of Generalized Information Criteria (GIC). Model selection among regularlized (nonlinear) regressions comes within the range of GIC. Derivation of information criteria reduces to how we estimate the asymptotic bias of a certain statistical functional, that is a penalized likelihood in our problem. Under the assumption that the specified family of probability distributions does not necessarily contain the true model, Konishi and Kitagawa (1996) derived the asymptotic bias as a function of the empirical influence function of the estimator and the score function of the parametric model.

Theorem 2.1 in Konishi and Kitagawa (1996) can be restated in the following way depending on the problem we consider. Whether the model is f_B or f_E, GIC will be given in the following form,

$$
\mathrm{GIC}(m, \lambda) = n \log(2\pi\hat{\sigma}^2) + n + 2\mathrm{tr}\left(I_G J_G^{-1}\right)
$$

where $\hat{\sigma}^2$ is the estimate of residual variance. I_G and J_G are the $(m+2) \times (m+2)$ matrices, and I_G is basically the product of the empirical influence function and the score function, while J_G is the matrix of second derivative of the penalized likelihood.

If we fit the f_B-class, then the solution to the model (5) by maximizing penalized likelihood (10) leads to (11). As for the explicit form of the penalty term for this case, please refer to Imoto and Konishi (2003). Note that, in our problem, it is the case of 'approximating the discount function', hence GCV is also valid as well as GIC. However, when it comes to the evaluation of f_E-class models (9), the regression functional is not linear with respect to unknown parameters. Hence model selection by GCV the lacks theoretical background for its applicability, while we can still construct GIC for the regularized estimation of the f_E-class models.

When the f_E-class models are estimated by the maximum penalized likelihood, the matrices I_G and J_G appear in the bias term take the following form,

$$\begin{aligned} I_G &= \frac{1}{n\hat{\sigma}^2}\begin{pmatrix} \Phi'\Lambda/\hat{\sigma}^2 - \lambda K\hat{\boldsymbol{w}}\mathbf{1}_n' \\ \boldsymbol{q}' \end{pmatrix}(\,\Lambda\Phi, \hat{\sigma}^2\boldsymbol{q}\,), \\ J_G &= \frac{1}{n\hat{\sigma}^2}\begin{pmatrix} \Phi'\Phi - D + n\hat{\sigma}^2\lambda K & \Phi'\Lambda\mathbf{1}_n/\hat{\sigma}^2 \\ \mathbf{1}_n'\Lambda\Phi/\hat{\sigma}^2 & n/2\hat{\sigma}^2 \end{pmatrix}, \end{aligned}$$

where $K = D_2'D_2$, Λ is given by

$$\Lambda = \mathrm{diag}\left[y_1 - \boldsymbol{c}_1'\boldsymbol{\delta}(\boldsymbol{t}_1;\hat{\boldsymbol{w}}), ..., y_n - \boldsymbol{c}_n'\boldsymbol{\delta}(\boldsymbol{t}_n;\hat{\boldsymbol{w}})\right].$$

$\mathbf{1}_n = (1, 1, ..., 1)'$, and $\boldsymbol{q}$ is an n-dimensional vector with ith element

$$\left(y_i - \boldsymbol{c}_i'\boldsymbol{\delta}(\boldsymbol{t}_i;\hat{\boldsymbol{w}})\right)^2/2\hat{\sigma}^4 - 1/2\hat{\sigma}^2.$$

Φ and D are $n\times m$ and $m\times m$ matrices which depend on particular choice of the regression functional δ. If we decide to place spline on zero coupon yield, our choice of δ is (6). Then (i,j)th element Φ_{ij} and D_{ij} is given by

$$\begin{aligned} \Phi_{ij} &= \sum_{k=1}^{L_i} c_i\delta(t_k^i;\hat{\boldsymbol{w}})f_j(t_k)t_k, \\ D_{ij} &= \sum_{\alpha=1}^{n}[(y_\alpha - \boldsymbol{c}_\alpha'\boldsymbol{\delta}(\boldsymbol{t}_\alpha;\hat{\boldsymbol{w}})) \\ &\times (\sum_{k=1}^{L_\alpha} c_\alpha\delta(t_k^\alpha;\hat{\boldsymbol{w}})f_i(t_k^\alpha)f_j(t_k^\alpha)(t_k^\alpha)^2)]. \end{aligned}$$

If we turn to place spline on the forward rate curve, this means we employ (8). Then (i,j)th element Φ_{ij} and D_{ij} is given by

$$\begin{aligned} \Phi_{ij} &= \sum_{k=1}^{L_i} c_i\delta(t_k^i;\hat{\boldsymbol{w}})f_j(t_k), \\ D_{ij} &= \sum_{\alpha=1}^{n}[(y_\alpha - \boldsymbol{c}_\alpha'\boldsymbol{\delta}(\boldsymbol{t}_\alpha;\hat{\boldsymbol{w}})) \\ &\times (\sum_{k=1}^{L_\alpha} c_\alpha\delta(t_k^\alpha;\hat{\boldsymbol{w}})f_i(t_k^\alpha)f_j(t_k^\alpha)]. \end{aligned}$$

The values of the smoothing parameter λ and the number of basis functions m are determined as the minimizers of the GIC.

Note that for fitting the smoothing spline for $\delta(t)$, the procedure suggested by Fisher, Nychka and Zervos (1995) is entirely valid, and there is no problem with the use of GCV. However, the use of GCV for fitting an exponential spline or comparing curves to be splined is no longer theoretically justified. We inevitably resort to GIC when the regression functionals take non-linear forms such as (6) and (8). Of course, GIC can also be constructed in the linear functional case (4), making it possible to compare the linear with the non-linear models directly.

4 MONTE CARLO EXPERIMENTS

We start by specifying the true functional form of the forward rate curve $f(t)$ in the experiments. The following functional form is set as the true term structure of the instantaneous forward rate;

$$\begin{aligned} f(t) = \beta_0 + \beta_1 \exp\left(-\frac{t}{\tau}\right) \\ +\beta_2\left[\frac{t}{\tau}\exp\left(-\frac{t}{\tau}\right)\right]. \end{aligned} \tag{12}$$

This parameterization was proposed by Nelson and Siegel (1987). In the simulation, we set $\beta_0 = 0.02$, $\beta_1 = -\beta_0$, $\beta_2 = 0.2$ and $\tau = 10$. Figure 2 shows $f(t)$ under this setting. By (1) and (2), we can derive from $f(t)$ the discount function $\delta(t)$ and zero coupon yield $\eta(t)$. As we omit the explicit form of $\delta(t)$ to save space, please refer to Anderson et al. (1996, p.41). Given $\delta(t)$, random samples were generated from the true bond price model $p_\alpha = R_\alpha\delta(t_\alpha) + \varepsilon_\alpha$ for $t_\alpha = 30\times((\alpha-1)/(n-1))$ and $\alpha = 1, ..., n$. For the error term ε_α, we consider an independent normal distribution case, $\varepsilon_\alpha \sim \mathrm{N}(0,\sigma^2)$, where $\sigma = 0.1$. The redemption payment R_α is assumed to be 100, considering that the face value of Japanese Governmental Bonds is ¥100.

The maturity interval $[0, 30]$ is divided into equally spaced intervals, and 100 time points are chosen: $\{t_\alpha\}$ with $t_1 = 0$ and $t_n = 30$. These time points are fixed throughout the experiments. The price of artificial zero coupon bonds is then generated according to the bond equation. All yield curves ($\hat{\delta}$, $\hat{\eta}$, and $\hat{f}$) were estimated in the end, regardless of the curve fitted, and the bias from the true curves (δ, η, f) were measured at the fixed time points. The squares of the biases over maturity were then averaged, and the mean-squared error (MSE) of the ith experiment is defined as

$$D_i^f = n^{-1}\sum_{\alpha=1}^{n}(\hat{f}^{(i)}(t_\alpha) - f(t_\alpha))^2.$$

The overall Monte Carlo mean $\tilde{D}^f = M^{-1}\sum_{i=1}^{M} D_i^f$ for M Monte Carlo trials and its standard deviation was then determined. $\tilde{D}^\delta$ and $\tilde{D}^\eta$ were calculated in the same way.

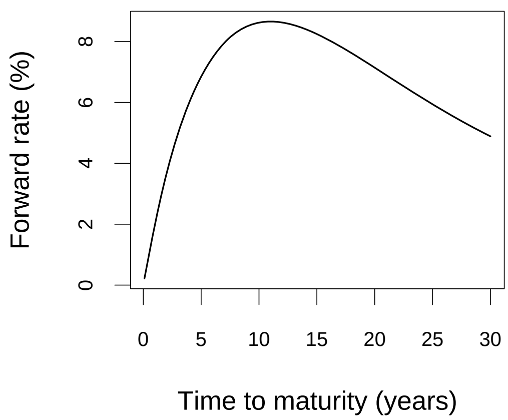

Figure 2. Nelson-Siegel type forward rate

Target func.	$\delta(t)$	$\eta(t)$	$f(t)$
f/B	2.92	1.36	7.67
(std. err.)	(0.52)	(0.25)	(1.45)
η/B	19.22	4.87	154.96
(std. err.)	(0.53)	(0.23)	(0.92)
δ/B	39.81	10.20	511.10
(std. err.)	(0.56)	(0.46)	(0.82)

Table 1. Simulation results for selection of smoothing parameter λ and number of basis functions m

Target func.	$\delta(t)$	$\eta(t)$	$f(t)$
f/B	4.60	1.80	29.7
(std. err.)	(0.62)	(0.44)	(2.73)
η/B	19.31	5.12	189.62
(std. err.)	(0.54)	(0.24)	(1.12)
δ/B	41.91	11.01	512.02
(std. err.)	(0.58)	(0.50)	(0.82)

Table 2. Simulation results for selection of smoothing parameter λ only (m set to one-third the sample size)

The results of the Monte Carlo simulations are summarized in Table 1. The simulation results were obtained by averaging over $M = 100$ repeated Monte Carlo trials. The standard deviations (SDs) are given in parentheses below the means. In the table, f/B indicates that B-spline bases are placed on the forward rate $f(t)$. Similarly, η/B and δ/B indicate placement of the B-spline on the zero coupon yield ($\eta(t)$) and the discount function ($\delta(t)$). The specification of the B-spline here follows that of Steeley (1991), and Eilers and Marx (1996), for example, and differs from the definition given in Fisher et al. in that extra knots are placed outside the actual maturity interval and knots are not overlapped at the ends. All the models are estimated by the maximum penalized likelihood method stated in Section 2, and the smoothing parameter λ is chosen according to GIC derived in Section 3.MSE values in the table are read as follows: for example, the MSE for the estimation of $f(t)$ via forward-rate-splining (f/B) is 7.67. Hence, on average, the bound for the estimated forward rate curve is approximately $\pm$ 2.77 basis points.

The simulation results here support one of the findings in Fisher, Nychka and Zervos (1995); fitting a smoothing spline for the forward rate curve (with B-spline bases) provides the best performance, and it is not recommended to estimate the discount function first and then derive other yield curves. Although the use of GCV in Fisher et al. has no theoretical foundation, their findings were indeed correct on that point. It should be noted, however, that GIC gives a theoretically justified route to compare the various yield curve models using a roughness penalty.

The GIC is used here to choose both λ and m. The introduction of λ was aimed at resolving the ill-posed nature of the regression spline, and choosing the optimal number of basis function m originates from the experience that introducing excess basis functions often leads to an unacceptable shape of the forward rate curve, even though the discount function may be reasonably shaped. On the other hand, in terms of fitting the smoothing spline, one might suspect that choosing the number of basis functions (m) may be unnecessary because the large roughness penalty value may automatically reduce the effective number of parameters.

In light of this argument, similar experiments were performed without choosing m. Instead, a fixed number of basis functions equal to one-third of the sample size was chosen; 33 in the experiments here. The results in Table 2 show that all the MSE values are larger than when an appropriate number of basis functions are chosen (Table 1). Most notably, choosing the number of basis function results in a significant reduction of the MSE for $f(t)$, particularly when $f(t)$ is splined directly (29.7 $\rightarrow$ 7.67 for normal independent error). It therefore appears that in all cases choosing the number of basis functions improves the estimation.

5 CONCLUSION

A penalized likelihood approach was proposed for estimation of the term structure of interest rates from a set of coupon data. In the penalized likelihood approach, the method for choosing the smoothing parameter is important. If the nonparametric regression functional is linear in its parameters, then GCV can be used. However, if we want to spline the zero coupon yield or the forward rate, we inevitably have to estimate exponential spline models, for which GCV loses its theoretical basis. It was shown that a customized version of the generalized information criterion (GIC) can be constructed even for nonlinear spline problems. Monte Carlo studies and analysis of real data clearly showed that B-splining the forward rate with a roughness penalty provides the most accurate estimation of yield curves, confirming the findings of Fisher, Nychka and Zervos (1995) by a theoretically valid route. It was also verified that choosing the optimal number of basis functions rather than letting the single (smoothing) parameter control the number of bases reduces the estimation error.

6 ACKNOWLEDGEMENT

The first author would like to thank Japan Society for the Promotion of Science for its support by Grant-in-Aid for Scientific Research 17500189, 17200020 and 14380127.

7 REFERENCES

Akaike, H. (1974) A New Look at the Statistical Model Identification, *IEEE Trans. Autom. Contr.* **AC-19**, 716–723.

Anderson, N., F. Breedon, M. Deacon, A. Derry and G. Murphy. (1996) *Estimating and Interpreting the Yield Curve*, John Wiley and Sons, Chichester.

Eilers, P. H. and Marx, B. D. (1996) Flexible Smoothing with B-splines and Penalties, *Statistical Science*, **11**, 89–102.

Fisher, M. E, Nychka, D and Zervos, D. (1995) Fitting the Term Structure of Interest Rates with Smoothing Splines, Federal Reserve Bank Finance and Economics Discussion Paper 95-1, January.

Imoto, S. and Konishi, S. (2003) Selection of Smoothing Parameters in B-spline Nonparametric Regression Models Using Information Criteria, *Annals of the Institute of Statistical Mathematics*, **55**, 671-687.

Konishi, S. and Kitagawa, G. (1996) Generalised Information Criteria in Model Selection, *Biometrika*, **83**, 875–890.

Langetieg, T. C. and Smoot, J. S. (1989) Estimation of the Term Structure of Interest Rates, *Research in Financial Services*, **1**, 181–222.

Mastronikola, K. (1991) Yield Curves for Gilt-Edged Stocks: A New Model, Bank of England Discussion Paper (Technical Series), No. 49.

McCulloch, J. H. (1971) Measuring the Term Structure of Interest Rates, *Journal of Business*, **44**, No. 1 (January), 19–31.

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

Nelson, C. R. and Siegel, A. F. (1987) Parsimonious Modeling of Yield Curves, *Journal of Business*, **60**(4), 473–89.

Schaefer, S. M. (1981) Measuring a Tax-Specific Term Structure of Interest Rates in the Market for British Government Securities, *The Economic Journal*, **91**, 415–38.

Steely, J. M. (1991) Estimating the Gilt-Edged Term Structure: Basis Splines and Confidence Intervals, *Journal of Business, Finance and Accounting*, **18**, No. 4 (June), 512–29.

Vasicek, O. A. and Fong, H. G. (1982) Term Structure Modeling Using Exponential Splines, *Journal of Finance*, **37**, No.2 (May), 339–56.