Fitting the term structure of interest rates with smoothing splines

Mark Fisher*
Research and Statistics
Board of Governors of the Federal Reserve System

Douglas Nychka
Department of Statistics
North Carolina State University

and

David Zervos
Monetary Affairs
Board of Governors of the Federal Reserve System

September 1994
First version: January 1994

*Send correspondence to either Mark Fisher, Mail Stop 91, Federal Reserve Board, Washington, DC 20551; phone: 202–452–2258; e-mail: mfisher@frb.gov or David Zervos Mail Stop 74, Federal Reserve Board, Washington, DC 20551; phone: 202–736–5643; e-mail: dzervos@frb.gov.
We thank Robert Bliss, Jon Faust, Ludger Hentschel and Christian Gilles. The views expressed herein are the authors' and do not necessarily reflect those of the Board of Governors or the Federal Reserve System.

Abstract

We describe a technique for fitting the term structure of interest rates using smoothing splines, which incorporate a "roughness" penalty. An increase in the penalty reduces the effective number of parameters. We use generalized cross validation to choose adaptively the penalty and hence the effective number of parameters. We show how our technique can be used to spline an arbitrary transformation of the discount function, using a B-spline bases. Our Monte Carlo simulations and estimation results suggest that fitting a smoothing spline to the forward rate curve using generalized cross validation produces the best results.

1 Introduction

There has been considerable effort expended in searching for an accurate and well-behaved technique for estimating the term structure of interest rates from a cross-section of coupon bond prices. McCulloch (1971 and 1975) was the pioneer in this field with a spline-based estimation technique built upon a simple theory of bond pricing. Under the assumption that the price of a bond is equal to the present value of its future coupon payments, McCulloch parameterizes the present value function as a cubic spline and estimates the term structure via simple linear regression.

Following McCulloch, Vasicek and Fong (1982), Shea (1984), Jordan (1984), Chambers, Carleton and Waldman (1984), and Coleman, Fisher, and Ibbotson (1992), among others, extended the spline-based estimation technique to explore tax-related effects on bond pricing, to consider different parameterizations of the splines, and to analyze potential sources of heteroskedasticity in the residuals. With all the refinements, spline based estimation tends to generate accurate bond pricing, but does have problems producing well-behaved implied forward rates. In addition, the choice of both the number and placement of the *knot points* for the spline creates the potential for ad hoc parameterizations, especially as the spline changes through time.

Other authors have pursued alternative estimation techniques based on parsimonious parameterizations of the discount function. For instance, Nelson and Seigel (1987) and Bliss (1993) consider a functional form with only four unknown parameters. (In contrast, for a sample of 150 securities, McCulloch would typically choose a spline with 18 parameters.) This model forces forward rates to an asymptote, which has some appeal; however, it does not fit the data as well as the spline-based methods.[1]

In this paper we develop a technique that can both price bonds accurately and produce relatively stable forward rates. The technique retains the spline based structure, but unlike

[1] Although these models force the forward rates through an asymptote, this does not guarantee the positivity of the forward rates. In fact for some days the asymptote is approximately −10 percent. In our experiments with the Nelson and Siegel parameterization, the average absolute pricing errors were approximately 2.5 to 3 times larger than the pricing errors from spline-based techniques.

McCulloch, Jordan, or Shea, it places the spline directly on the forward rate function.[2] Also in contrast to previous studies, we fit smoothing splines instead of regression splines. Smoothing splines have a penalty for excess "roughness" and a single parameter that controls the size of the penalty.[3] An increase in the penalty reduces the effective number of parameters. Hence, a single value controls the entire parameterization of the spline. For regression splines, the number of parameters must be chosen exogenously. By contrast, we use "generalized cross validation" to choose adaptively the roughness penalty—and hence the effective number of parameters. In other words, we let the data determine the appropriate number of parameters.

For comparison, we present results from splining the discount function and the logarithm of the discount function. In each case we look at a set of regression splines (as in McCulloch) and smoothing splines with adaptively chosen parameterizations. In order to gauge the ability of the alternative estimation methods to accurately uncover the actual term structure, we turn to Monte Carlo simulation. We postulate a "true" term structure, subject the true bond prices to noise, and estimate with each technique. We then look at a variety of summary statistics to measure the biases and standard errors associated with the fitted term structure. Based on our simulations and our estimation results using daily data from December of 1987 though September 1994, splining the forward rate function with a smoothing spline and choosing the effective number of parameters via generalized cross validation produces, in general, the most accurate and least biased results.

2 Splining the term structure

The term structure of interest rates can be identified with any of a number of related concepts. For example, the discount function, $\delta(t,\tau)$, gives the price at time t of default-free zero-coupon bond that pays one unit at time τ. Hereafter, we assume that the current

[2]There is a mild similarity to Vasicek and Fong, however they use an exponential spline on the discount function which Shea (1985) shows to be effectively equivalent to a polynomial spline.

[3]There is a superficial similarity between the smoothing splines we present here and the maximum smoothness estimators proposed by Adams and Van Deventer (1994).

time is 0, suppress the first index, and write $\delta(\tau)$. The zero-coupon yield curve, $z(\tau) := -\log[\delta(\tau)]/\tau$, gives the yield-to-maturity on a zero-coupon bond that matures at time τ. The instantaneous forward rate curve, $f(\tau) := -d\log[\delta(\tau)]/d\tau$, gives the marginal return at maturity τ of extending one's investment. We will also be interested in $\ell(\tau) := -\log[\delta(\tau)] = \tau\, z(\tau)$.

The techniques described in this paper are designed to extract the term structure from a set of coupon bonds whose prices are largely determined by the present value of their stated payments.[4] Consider a set of n bonds. Let p_i be the price of bond i, c_{ij} be its j-th payment, paid at time τ_{ij}, and m_i be the number of remaining payments. Then[5]

$$p_i = \sum_{j=1}^{m_i} c_{ij}\, \delta(\tau_{ij}) + \varepsilon_i = {c_i}^\top \widetilde{\delta}(\tau_i) + \varepsilon_i \qquad (1)$$

where c_i is the vector of payments for bond i, τ_i is the vector of maturities of those payments, ε_i is a random variable[6] and

$$\widetilde{\delta}(\tau_i) := (\delta(\tau_{i1}), \cdots, \delta(\tau_{im_i}))^\top$$

is the $m_i \times 1$ column vector that results from applying δ to each element of τ_i.

2.1 Cubic B-spline basis

A cubic spline is a piecewise cubic polynomial joined at so-called knot points. At each knot point, the polynomials that meet are restricted so that the level and first two derivatives of each cubic are identical. Each additional knot point in the spline adds one independent parameter, as three of the four parameters of the additional cubic polynomial are constrained by the restriction. By increasing the number of knots, cubic splines provide increasingly flexible functional form. A simple, numerically stable parameterization of a cubic spline is

[4] Treasury securities that do not reasonably meet this criteria are callable bonds, "flower" bonds, and bonds "on special" in the repo market.

[5] "$\top$" denotes transpose. In addition, the prices, p_i, include accrued interest.

[6] We discuss this more below.

provided by a cubic B-spline basis.

Let $\{s_k\}_{k=1}^{K}$ denote the knot points, with $s_k < s_{k+1}$, $s_1 = 0$, and $s_K = M$, the maximum maturity of any bond in the sample.[7] The knot points define $K-1$ intervals over the domain of the spline, $[0,\, T]$. For the purpose of defining a B-spline basis, it is convenient to define an augmented set of knot points, $\{d_k\}_{k=1}^{K+6}$, where $d_1 = d_2 = d_3 = s_1$, $d_{K+4} = d_{K+5} = d_{K+6} = s_K$, and $d_{k+3} = s_k$ for $1 \le k \le K$.

A cubic B-spline basis is a vector of $\kappa = K+2$ cubic B-splines defined over the domain. A B-spline is defined by the following recursion, where $r = 4$ for a cubic B-spline and $1 \le k \le \kappa$:[8]

$$\phi_k^r(\tau) = \frac{\phi_k^{r-1}(\tau)\,(\tau - d_k)}{d_{k+r-1} - d_k} + \frac{\phi_{k+1}^{r-1}(\tau)\,(d_{k+r} - \tau)}{d_{k+r} - d_{k+1}}$$

for $\tau \in [0,\, T]$, with

$$\phi_k^1(\tau) = \begin{cases} 1, & \text{if } d_k \le \tau < d_{k+1} \\ 0, & \text{otherwise.} \end{cases}$$

To simplify notation, let $\phi_k(\tau) := \phi_k^4(\tau)$. The cubic B-spline basis, then, is the row vector

$$\phi(\tau) := (\phi_1(\tau), \cdots, \phi_\kappa(\tau)).$$

Over any interval between adjacent knot points, s_k and s_{k+1}, there are four non-zero B-splines, with adjacent intervals sharing three. This gives $\phi(\tau)$ a semi-orthogonal structure from which it gets its numerical stability. Any cubic spline can be constructed from linear combinations of the B-splines, $\phi(\tau)\,\beta$, where $\beta := (\beta_1, \cdots, \beta_\kappa)^\top$ is a vector of coefficients. Panel A of Figure 1 shows a cubic B-spline basis defined over six equally spaced knot points.

As is stands, $\phi(\tau)$ is a vector-valued function of a scalar argument, τ. In what follows, it will prove useful to have notation for a B-spline basis as a function of vector-valued argument, τ_i. To that end, define $\widetilde{\phi}_k(\tau_i) := (\phi_k(\tau_{i1}), \cdots, \phi_k(\tau_{im_i}))^\top$, an $m_i \times 1$ column

[7] In all cases, we distribute the knot points according to the distribution of the final maturities of the bonds. For example, with three knot points, we place the single interior knot point, s_2, at the median maturity.

[8] For a more detailed discussion of b-spline bases and their properties, see de Boor (1978).

vector, and $\widetilde{\phi}(\tau_i) := (\widetilde{\phi}_1(\tau_i), \cdots, \widetilde{\phi}_\kappa(\tau_i))$, an $m_i \times \kappa$ matrix.

2.2 Regression splines

We first consider splining an arbitrary function of the term structure, $h(\tau)$. The only restriction we impose on h is that there exist a functional transformation g such that

$$g(h(\,\cdot\,), \tau) \equiv \delta(\tau).$$

We will parameterize $h(\tau)$ as a cubic spline:

$$h_s(\tau, \beta) := \sum_{k=1}^{\kappa} \beta_k \, \phi_k(\tau) = \phi(\tau)\, \beta.$$

Define the splined discount function $\delta_s(\tau, \beta) := g(h_s(\,\cdot\,, \beta), \tau)$, and define the present value of a bond's payments according to $\delta_s(\tau, \beta)$

$$\pi_i(\beta) := {c_i}^\top \widetilde{\delta}_s(\tau_i, \beta) = {c_i}^\top \widetilde{g}(h_s(\,\cdot\,, \beta), \tau_i) = {c_i}^\top \widetilde{g}(\phi(\,\cdot\,)\, \beta, \tau_i),$$

where $\widetilde{g}(\phi(\,\cdot\,)\, \beta, \tau_i) := (g(\phi(\,\cdot\,)\, \beta, \tau_{i1}), \cdots, g(\phi(\,\cdot\,)\, \beta, \tau_{im_i}))^\top$, an $m_i \times 1$ vector.

Let P be an $n \times 1$ vector of bond prices, p_i, and $\mathit{\Pi}(\beta)$ be the corresponding vector of present values of the bonds, $\pi_i(\beta)$. Then $h_s(\tau, \beta^*)$ is a regression spline, where β^* solves

$$\min_{\beta} \, [(P - \mathit{\Pi}(\beta))^\top (P - \mathit{\Pi}(\beta))]. \tag{2}$$

In general, minimization problem (2) can be solved as a nonlinear least squares problem. Following Chow (1983), we linearize $\mathit{\Pi}(\beta)$ around an initial guess β^0,

$$\mathit{\Pi}(\beta) \approx \mathit{\Pi}(\beta^0) + (\beta - \beta^0) \left. \frac{\partial \mathit{\Pi}(\beta)}{\partial \beta^\top} \right|_{\beta = \beta^0},$$

and define $X(\beta^0) := \partial \mathit{\Pi}(\beta)/\partial \beta^\top \big|_{\beta=\beta^0}$ and $Y(\beta^0) := P - \mathit{\Pi}(\beta^0) + \beta^0 \, X(\beta^0)$. Rearranging

(2) using these definitions yields

$$\min_{\beta}\left[\left(Y(\beta^0)-X(\beta^0)\beta\right)^\top\left(Y(\beta^0)-X(\beta^0)\beta\right)\right]. \tag{3}$$

The minimizer for (3) is

$$\beta^1=\left(X(\beta^0)^\top X(\beta^0)\right)^{-1}X(\beta^0)^\top Y(\beta^0),$$

where β^1 is an updated β^0. We can use β^1 as the initial guess for the next iteration, obtaining β^2. We iterate until convergence. The solution is the fixed-point $\beta^* = (X(\beta^*)^\top X(\beta^*))^{-1}X(\beta^*)^\top Y(\beta^*)$.

2.3 Functional forms

In this section we examine three candidates for $h(\tau)$: $\delta(\tau)$, $\ell(\tau)$, and $f(\tau)$. The table below summarizes the relevant information regarding each case.

Case (i) is the sort of regression spline that McCulloch pioneered. Since $\partial\pi_i(\beta)/\partial\beta^\top$ does not depend on β in this case, (2) can be written

$$\min_{\beta}\,[(P-X\beta)^\top(P-X\beta)],$$

the solution to which is the OLS estimator $\beta^*=(X^\top X)^{-1}X^\top P$. In Case (ii), $g(h(\,\cdot\,),\tau)$ is not linear in $h(\,\cdot\,)$, so the nonlinear procedure must be followed. Case (iii) is identical to Case (ii) with the exception that

$$\psi(\tau):=\int_0^\tau \phi(s)\,ds,$$

replaces $\phi(\tau)$ everywhere.[9] This follows from

$$\delta_s(\tau,\beta) = g(h_s(\,\cdot\,,\beta),\tau) = \exp\left(-\int_0^\tau \phi(s)\beta\, ds\right) = \exp(-\psi(\tau)\beta).$$

Case	$h(\tau)$	$g(h(\,\cdot\,),\tau)$	$\pi_i(\beta)$	$\partial\pi_i(\beta)/\partial\beta^\top$
i	$\delta(\tau)$	$h(\tau)$	$c_i{}^\top\widetilde{\phi}(\tau_i)\,\beta$	$c_i{}^\top\widetilde{\phi}(\tau_i)$
ii	$\ell(\tau)$	$\exp(-h(\tau))$	$c_i{}^\top\exp(-\widetilde{\phi}(\tau_i)\,\beta)$	$-\pi_i(\beta)\,c_i{}^\top\widetilde{\phi}(\tau_i)$
iii	$f(\tau)$	$\exp(-\int_0^\tau h(s)\,ds)$	$c_i{}^\top\exp(-\widetilde{\psi}(\tau_i)\,\beta)$	$-\pi_i(\beta)\,c_i{}^\top\widetilde{\psi}(\tau_i)$

2.4 Smoothing splines

In a regression spline, the number of parameters is determined by the number of knot points. Either too few or too many knot points can lead to poor estimates. The strategy we follow is to use a large number of knot points but penalize excess variability in the estimated discount function. This has the effect of reducing the effective number of parameters since the penalty forces implicit relationships between the parameters of the spline. The penalty is defined as

$$\lambda\int_0^T h''(\tau)^2 d\tau,$$

a constant times the integral of the squared second derivative of the function being splined. For the time being we assume λ is fixed. The problem now consists of minimizing the residual sums of squares plus the penalty:

$$\min_{h(\tau)\in\mathcal{H}}\left[\sum_{i=1}^n\left(p_i - c_i{}^\top\widetilde{g}(h(\,\cdot\,),\tau_i)\right)^2 + \lambda\int_0^T h''(\tau)^2\,d\tau\right],$$

where $\mathcal{H}$ is the space of all functions defined on R_+ with squared second derivatives which integrate to a finite value. We restrict ourselves to this space since there are fairly well

[9] Panel B of Figure 1 shows the integral of a cubic B-spline basis defined over six equally spaced knot points.

known theoretical results on the solutions to such problems.[10,11]

In terms of the spline, $h_s(\tau, \beta)$, the penalty can be written as follows:

$$\lambda \int_0^T \left(\frac{\partial^2 h_s(\tau,\beta)}{\partial \tau^2} \right)^2 d\tau = \lambda \beta^\top \left(\int_0^T \phi''(\tau)^\top \phi''(\tau) d\tau \right) \beta = \lambda \beta^\top H \beta.$$

H is a $\kappa \times \kappa$ matrix that is band diagonal by the structure of a B-spline basis. Since any β that makes $h_s(\tau, \beta)$ linear in τ is not penalized, H has two zero eigenvalues. Also note that H is completely determined by the knot points.

The minimization problem can be stated as follows for a given λ:

$$\min_{\beta(\lambda)} \left[\left(P - \Pi(\beta(\lambda)) \right)^\top \left(P - \Pi(\beta(\lambda)) \right) + \lambda \, \beta(\lambda)^\top H \beta(\lambda) \right].$$

In general, the minimizer is found by nonlinear least squares as described in the previous section, iterating on

$$\beta^{i+1}(\lambda) = \left(X(\beta^i(\lambda))^\top X(\beta^i(\lambda)) + \lambda H \right)^{-1} X(\beta^i(\lambda))^\top Y(\beta^i(\lambda))$$

until convergence:

$$\beta^*(\lambda) = \left(X(\beta^*(\lambda))^\top X(\beta^*(\lambda)) + \lambda H \right)^{-1} X(\beta^*(\lambda))^\top Y(\beta^*(\lambda)).^{12} \tag{4}$$

The smoothing spline is given by $h_s(\tau, \beta^*(\lambda))$.[13]

---

[10] See Wahba (1990). Strictly speaking, the theoretical results only apply when $g(h(\,\cdot\,), \tau)$ is linear in $h(\,\cdot\,)$.

[11] In the macroeconomic literature this form of smoothing has been used for filtering economic aggregates. Hodrick and Prescott (1981) proposed a method for extracting the long-run component of a time series from its cyclical component through the an analogous minimization process. They consider a discrete time series $\{y_t\}_{t=1}^T$, and search for the smooth or long–run component by conducting the following minimization problem:

$$\min_{\{x_t\}_{t=1}^{T-1}} \left\{ \sum_{t=1}^{T} (y_t - x_t)^2 + \lambda \sum_{t=1}^{T-2} [(x_{t+2} - x_{t+1}) - (x_{t+1} - x_t)]^2 \right\},$$

where $\{x_t\}$ is the smoothed component and is analogous to a discrete time spline.

[12] When $g(h(\,\cdot\,), \tau)$ is linear in $h(\,\cdot\,)$, $\beta^*(\lambda) = (X^\top X + \lambda H)^{-1} X^\top P$.

[13] See Appendix A for a discussion of imposing the restriction $\delta_s(0, \beta) = 1$.

Formally, (4) is ridge regression estimator. By employing a roughness penalty, we can over-parameterize the spline, making $X(\beta^*(\lambda))^\top X(\beta^*(\lambda))$ nearly singular, and use the penalty to reduce the effective number of parameters. The penalty thus "solves" the multicollinearity problem. The advantage of this technique is that the shape of the spline is controlled by a single parameter, λ.

One common measure of the effective number of parameters is the trace of $A(\lambda)$, denoted $\mathrm{tr}(A(\lambda))$, where

$$A(\lambda) := X(\beta^*(\lambda))\Big(X(\beta^*(\lambda))^\top X(\beta^*(\lambda)) + \lambda H\Big)^{-1} X(\beta^*(\lambda))^\top.$$

Note that $A(\lambda)\,Y(\beta^*(\lambda))$ is the vector of fitted Y values which in the linear case is the vetcor of fitted prices. The extreme cases are $\mathrm{tr}(A(0)) = \kappa$ (with no penalty the number of effective parameters equals the number of B-splines), and $\mathrm{tr}(A(\infty)) = 2$ (with an infinite penalty the number of effective parameters equals 2).[14]

2.5 Generalized cross validation

In this section, we provide a technique for choosing the appropriate value for λ. We choose the value of λ that minimizes the "generalized cross validation" (GCV) value,[15]

$$\gamma(\lambda) := \frac{\Big((I - A(\lambda))Y(\beta^*(\lambda))\Big)^\top \Big((I - A(\lambda)))Y(\beta^*(\lambda))\Big)}{\Big(n - \theta\,\mathrm{tr}(A(\lambda))\Big)^2}. \tag{5}$$

The numerator of (5) is the residual sum of squares. When $\theta = 1$, the denominator is the squared effective degrees of freedom (the difference between the number of observations and

[14]See Appendix B.

[15]As the name "generalized cross validation" suggests, there is a criterion called cross validation. Intuitively, cross validation starts by looking at a "leave-one-out" estimator for each data point. The residual values from actual data point and the fitted data point from the leave-one-out estimation are averaged to construct the cross validation measure. With few parameters in the estimation the residuals will tend to be large, due to a poor overall fit, while with an interpolant, the perfect fit will tend produce spurious movements and hence large out of sample residuals. Somewhere in between is the lowest value of the cross validation measure and the "best" estimate. Generalized cross validation employs an alternative weighting scheme in the construction of the sum residuals. See Wahba (1990) for details.

the effective number of parameters). The parameter θ is called the cost. It controls the trade-off between goodness-of-fit and parsimony. In plain-vanilla GCV, $\theta = 1$. However, θ can be increased to reduce the signal extracted, thereby stiffening the spline.[16]

When $g(h(\cdot), \tau)$ is linear in $h(\cdot)$, $A(\lambda) = X(X^\top X + \lambda H)^{-1} X^\top$, and there is a simplified expression for $\gamma(\lambda)$ that can be minimized directly.[17] In general, however, a new $X(\beta^*(\lambda))$ matrix must be formed for each value of λ. Thus for each value of λ we test, we must solve for $\beta^*(\lambda)$ and then calculate $\gamma(\lambda)$. The overall solution is given by

$$\beta^*(\lambda^*) = \Big(X(\beta^*(\lambda^*))^\top X(\beta^*(\lambda^*)) + \lambda H\Big)^{-1} X(\beta^*(\lambda^*))^\top Y(\beta^*(\lambda^*)),$$

where λ^* minimizes $\gamma(\lambda)$. The resulting GCV smoothing spline is $h_s(\tau, \beta^*(\lambda^*))$.

2.6 Implementing the estimators

For smoothing splines, we need starting values for λ. However, λ is not free of units, thus it is not easy to know in advance what a good starting value is. At extremely large values of λ the GCV function becomes very flat and optimizers can get stuck at non-minimums. For the data we have examined, $\gamma(\lambda)$ is well-behaved for starting values between 10^{10} and 10^{20}; however, moving beyond 10^{25} can create serious precision problems.

When $g(h(\cdot), \tau)$ is not linear in $h(\cdot)$, we need starting values for β as well. Fortunately, good starting values for β are easy to calculate. One of the properties of B-splines is that $\sum_{k=1}^{\kappa} \phi(\tau) = 1$. As a consequence, the coefficients, β, track the value of the function, $\phi(\tau)\,\beta$. Thus any reasonable estimate of the function to be splined can be used to form starting values. For example, suppose a crude estimate of the function to be splined is $\widehat{h}(\tau)$. Let the starting value for β_k be $\beta_{k0} = \frac{1}{3}\sum_{i=k}^{k+2} \widehat{h}(d_i)$. With these starting values, the fixed point problem converges rapidly.[18]

---

[16] In our empirical work with U.S. Treasury bonds and our simulations, we set $\theta = 2$.

[17] See Appendix B.

[18] All of the estimators described in this paper are implemented in a set of *Mathematica* packages, available on request. These packages and a discussion of how they work will appear in Fisher and Zervos (forthcoming).

3 Monte Carlo simulations

In order to gauge the ability of the various estimation techniques to price bonds accurately and to uncover the true zero coupon and forward rate functions we resort to Monte Carlo simulation. We consider four separate functional forms for the "true" structure of forward rates and then generate the "true" prices for a large set of Treasury securities. By subjecting the prices to noise, we can apply a number of alternative estimation techniques repeatedly and construct explicit measures of the goodness-of-fit, biases, and standard errors for each of the estimators.

We will specify various "true" forward rate functions, $f_k(\tau)$. Associated with each $f_k(\tau)$ is a true discount function, $\delta_k(\tau)$. Given $\delta_k(\tau)$, we can calculate the "true" bond prices from a set of coupon payments, $\pi_{ik} := {c_i}^\top \widetilde{\delta}_k(\tau_i)$. To each of these true prices, we add random noise, producing "observed" bond prices, $p_{ikr} := \pi_{ik} + \varepsilon_{ikr}$, where ε_{ikr} is independently normally distributed with zero mean and standard error σ. For each $\delta_k(\tau)$, we produce R of these sets. We then fit the term structure to each of these sets of generated data using various functional forms, $h_s(\tau, \beta)$, and a variety of knot-point specifications and parameterizations. For each of these estimation techniques (indexed by α) we produce fitted bond prices, $\hat{p}^\alpha_{ikr}$, as well as fitted forward and zero functions, $\hat{f}^\alpha_{kr}(\tau)$, and $\hat{z}^\alpha_{kr}(\tau)$.

3.1 Criteria

We look at two sorts of criteria, one relating to how well we uncover the true bond prices and the other relating to how well we uncover the forward and zero curves.

For bond prices, we examine the mean absolute pricing error (MAPE) for the bonds in the sample:

$$\mathcal{A}(k, \alpha) := \frac{1}{R\,n} \sum_{r=1}^{R} \sum_{i=1}^{n} \mid \hat{p}^\alpha_{ikr} - \pi_{ik} \mid .$$

We also calculate the mean absolute pricing error for specific bonds that may or may not be in the sample:[19]

$$\mathcal{A}(k,\alpha,c,\tau) := \frac{1}{R}\sum_{r=1}^{R} \mid \hat{p}_{kr}^{\alpha}(c,\tau) - \pi_{ik}(c,\tau) \mid,$$

where $\pi_{ik}(c,\tau)$ is the true value of a bond with coupon rate c that matures at time τ and $\hat{p}_{kr}^{\alpha}(c,\tau)$ is the fitted value.

For forward and zero curves we examine three criteria. In what follows, let $\mathcal{F}$ denote either f or z. It is convenient to define the mean value of the fitted function at maturity τ:

$$\mathcal{M}_{\mathcal{F}}(k,\alpha,\tau) := \frac{1}{R}\sum_{r=1}^{R} \hat{\mathcal{F}}_{kr}^{\alpha}(\tau).$$

We now describe the criteria. First, the mean error (the bias) for a given maturity τ:

$$\mathcal{B}_{\mathcal{F}}(k,\alpha,\tau) := \mathcal{M}_{\mathcal{F}}(k,\alpha,\tau) - \mathcal{F}_k(\tau).$$

Second, the standard error for a given maturity τ:

$$\mathcal{S}_{\mathcal{F}}(k,\alpha,\tau) := \left(\frac{1}{R}\sum_{r=1}^{R}\left(\hat{\mathcal{F}}_{kr}^{\alpha}(\tau) - \mathcal{M}_{\mathcal{F}}(k,\alpha,\tau)\right)^2\right)^{\frac{1}{2}}.$$

Finally, the integrated mean absolute error (IMAE):

$$\mathcal{I}_{\mathcal{F}}(k,\alpha) := \frac{1}{M}\int_0^M \mid \mathcal{B}_{\mathcal{F}}(k,\alpha,\tau) \mid \, d\tau.$$

3.2 Our implementation

We choose to use the data from a typical day as our set our coupon payments for our simulations. The coupon payment structure does not vary significantly over our sample period (1988–1994), and we use April, 30 1993, when there are 163 bonds that meet our estimation criteria: non-callable, non-flower coupon bearing securities with maturities greater than

[19] In particular, U.S. Treasury data have a "gap" in the maturity structure after removing callable bonds. We are interested in how well the estimators would price non-callable bonds in the gap.

30 days. For the purposes of this analysis, we let $\sigma = 0.1$, the average residual standard deviation from our estimations during the more recent time periods.[20] We then generate 100 sets of observed prices ($R = 100$).

We explore the following functional forms for the true term structure of instantaneous forward rates:

$$
\begin{aligned}
f_1(\tau) &= 0.07305 \\
f_2(\tau) &= 0.05 + 1.461 \times 10^{-3}\,\tau \\
f_3(\tau) &= 0.04 + 4 \times 10^{-3}\,\tau - 1.33 \times 10^{-4}\,\tau^2 \\
f_4(\tau) &= 0.02 + 2.66 \times 10^{-3}\,\tau + 4.4 \times 10^{-4}\,\tau^2 \\
&\quad - 2.429 \times 10^{-5}\,\tau^3 + 2.37 \times 10^{-7}\,\tau^4 + 1.7 \times 10^{-3}\sin(0.566\,\tau)
\end{aligned}
$$

Each functional form is shown in figure 2 where the forward rate is in percentage terms and maturity in years.

In order to gauge both the appropriate placement of the spline as well as the as the efficacy of fixed knot point versus GCV parameterizations, we let the functional form $h(\tau)$ be $\delta(\tau)$, $\ell(\tau)$, or $f(\tau)$ and then estimate using 3, 6, and 10 knot point regression spline specifications, and finally the GCV based smoothing spline specification.[21] This gives us twelve alternative estimation methods to consider for each functional form. (These are the techniques indexed by α.)

3.3 Simulation results

The Monte Carlo simulation results are summarized in Tables 1–4. The central conclusion is that the best estimation technique is a smoothing spline specification on the forward rate function with the GCV parameterization, while the worst technique is a regression spline

[20] We have considered $\sigma = 0.22$ which is more representative of the early part of our sample. Although the larger standard error is reflected in the size of the errors and biases, the relative merits of one estimation technique versus another are unchanged.

[21] For GCV-based estimations, we choose the number of knots to be one-third of the sample size. Typically, for our data, this produces 50 to 60 knots. More knots slows down computation without changing the results.

on the discount function with 10 knots (similar to McCulloch's original technique).

Overall, there are a few recurring themes across all the simulations. First, the results demonstrate the importance of the penalty on the placement of the spline. The penalty tends to linearize the function that is being splined, especially in those regions where the cost of parsimony—the increase in the residual sum of squares—is small. This is true in the 25- to 30-year region where there are few data points. Thus the behavior of the curves at the long end reflects the action of the penalty on the functional form. In particular, when splining $\ell(\tau)$ with GCV, the forward rates tend to flatten out at the longer maturities, since the penalty tends to linearize the log of the discount function (*i.e.*, the derivative of the forward rate curve). When splining $f(\tau)$ the penalty acts acts to linearize the forward curve itself. When splining $\delta(\tau)$ the penalty tends to linearize the discount function, which translates into forcing the slope of the forward rate function to equal the square of its level. In this case, the forward rates tend to turn up at the longer maturities, producing a bias.

A second recurring theme across all the simulations is the similarity of the in-sample average absolute pricing errors. The pricing errors are large only when the curve is under-parameterized (*i.e.*, when the forward rates are given by $f_4(\tau)$ and only 3–6 knots are chosen). In many instances the in sample fit can be quite reasonable while the out of sample properties are quite bad.

Third, all methods tend to misprice bonds with maturities in the 13- to 21-year region, although the worst cases occur when splining $\delta(\tau)$ and $\ell(\tau)$ with fixed knot point specifications. The final property common to all simulations is the widening of the standard error bands as maturity increases. The precision of the estimates tends to drop off dramatically in the thirty year region. Below, we discuss the results for each functional form in more detail.

3.3.1 Simulation results for $f_1(\tau)$

Table 1 presents the resulting values for the test statistics when the true forward rate function is $f_1(\tau)$, a perfectly flat term structure. Notice that when placing the spline on $\delta(\tau)$, GCV yields 11.0 effective parameters while placement on $\ell(\tau)$ and $f(\tau)$ results in 1.0

and 2.0 effective parameters, respectively. The true function contains only 1 free parameter so that GCV actually chooses the "correct" number of parameters when the spline is placed on $\ell(\tau)$.[22] The over–parameterization in the case of splining $\delta(\tau)$ results in forward IMAE and zero IMAE values which are substantially greater than the zero values associated with splining $\ell(\tau)$ and $f(\tau)$. All methods, with the expectation of splining $\delta(\tau)$ with 3 knot points, result in about the same average absolute pricing error.

Turning to the biases and standard errors, all methods produce low zero rate biases and standard errors in the 2- to 10-year region. For the 30-year region, the potential inaccuracies become more serious with the an 11.6 basis point bias when splining $\delta(\tau)$ with 3 knots and a 4.6 basis point standard errors when splining $f(\tau)$ with 10 knots. For the forward rates, splining $\ell(\tau)$ and $f(\tau)$ with GCV results in essentially unbiased estimates throughout the entire curve and standard errors of no more than about 1 basis point. Splining $\delta(\tau)$ with any parameterization continues to produce biased estimates with fairly large standard errors.

Finally, looking the pricing errors for synthetic securities, using GCV and placing the spline on either $\ell(\tau)$ or $f(\tau)$ produces accurate results, even in the 15- to 20-year region where there is virtually no data. On the other hand, placing the spline on $\delta(\tau)$ yields relatively poor estimates, especially with only a few parameters.

3.3.2 Simulation results for $f_2(\tau)$

In this case splining $f(\tau)$ with GCV results in 2 effective parameters, the actual number of free parameters in $f_2(\tau)$, while splining $\delta(\tau)$ or $\ell(\tau)$ with GCV results in an over-parameterization. Again, the average absolute pricing errors are virtually the same across all methods, indicating again that this test statistic is not very useful for distinguishing the quality of the alternative methods. Also, the IMAE values are all still quite low, with the best results occuring when splining $f(\tau)$ with GCV and the worst results for splining $\delta(\tau)$ with either 3 knots or GCV.

The zero rate biases and standard errors are again rather small for all methods consid-

[22]Unrestricted GCV is constrained to pick at least 2 effective parameters. The restriction imposed on GCV estimator for $\ell(\tau)$ reduces the minimum to 1. See the Appendices.

ered, however they continue to become magnified at the 30-year region except when splining $f(\tau)$ with GCV. As for the forward rate biases and standard errors, there exists a much greater potential for inaccuracies. With 3 knot points all placements result in seriously based estimators for the 30-year rate, between 30 and 124 basis points, with standard errors between 20 and 94 basis points. The results are better with the four knot specification, but increasing the number of knots to eight results in significant bias and standard errors. With GCV, placement of the spline on $\delta(\tau)$ or $\ell(\tau)$ causes only moderate biases in the 2 to 10-year region, but substantial biases in the 30-year region. Splining $f(\tau)$ with GCV produces the most accurate results with virtually no bias anywhere on the curve, and standard errors well under a basis point.

The pricing of synthetic securities yields predictable results. Splining $f(\tau)$ with GCV produces the most dependable pricing for all maturities, while the other methods do reasonably well except in the 30-year region.

3.3.3 Simulation results for $f_3(\tau)$

The function $f_3(\tau)$ has the characteristic upward slope in the shorter maturities followed by a downward slope in the longer maturities, but it is still a relatively simple functional form. We consider a more complicated form in the final subsection.

Here, all GCV methods yield over-parameterizations, however, placing the spline on $f(\tau)$ is the most parsimonious. Again, the pricing errors are approximately the same except for splining $\delta(\tau)$ with 3 knot points. Looking at the IMAE values, the best results come from a spline placement on either $\ell(\tau)$ or $f(\tau)$ and choosing 3 knot points. For the GCV methods, placement on $f(\tau)$ yields the best results, however, all the GCV methods produce substantially inferior results to the low knot specifications. For all parameterizations, placing the spline on $\delta(\tau)$ continues to yield poor IMAE values.

The biases and standard errors for the zero coupon rates are once again quite small in the 2- to 10-year region with best performance for the fixed knot parameterizations with the spline on either $\ell(\tau)$ or $f(\tau)$. The same is true for the forward rate biases and standard errors with some fairly large biases in all the GCV methods.

Finally, the average absolute pricing errors are most pronounced in the 15 to 20-year region and the 30-year region.

3.3.4 Simulation results for $f_4(\tau)$

The last functional form under consideration produces a forward rate function that is typical of the shapes in recent data. In expectation—with all the twists and turns—it should be difficult for the under parameterized methods and the misspecified functional forms to produce reasonable fits. Here is where we expect GCV to fully parameterize the curve and produce a dependable estimate.

Notice first that the GCV methods choose between approximately 12 and 17 parameters depending on the placement. Also, the average absolute pricing errors are now substantially higher for all the fixed knot point specifications than for any of the GCV methods. This indicates that the parsimonious fixed knot specifications simply do not have enough degrees of freedom to accurately price the securities in this case. Problems with fixed knot specifications carry over to the IMAE values. Here, GCV produces the lowest errors and placing the spline on $f(\tau)$ gives the most accurate forward rate predictions while placing the spline on $\ell(\tau)$ gives the most accurate zero rate predictions.

In general, the zero and forward rate biases for the fixed knot specifications are higher than the biases associated with GCV, however, there are some exceptions.[23] The single knot point parameterizations simply do have enough flexibility to capture all the movements in the underlying functions, hence, each produces both heavily biased zero and forward rate estimates. The less parsimonious fixed knot specifications yield less biased results in the 2 to 10-year region, but become quite erratic in the 30 year region where forward rate biases vary between 150 and 870 basis points. Placing the spline on $f(\tau)$ and using GCV continues to provide an estimate which gives the least biased results.

All methods have trouble pricing a 15-year security, with the average error between $1.50

[23] The 10-year forward rate bias for splining $f(\tau)$ with GCV has a remarkably low bias and standard error, but in the 2- and 5-year it produces a thirty basis point bias while in the 30-year region the bias is 83 basis points.

and \$3.70 per one hundred dollars of face value. For the 20-year, the problems are less severe but are still quite significant in the fixed knot point specifications. Trouble with the 30-year security also arises in cases where the spline is placed on $\delta(\tau)$ or $\ell(\tau)$. The most accurate pricing tends to come from splining either $f(\tau)$ or $\ell(\tau)$ with GCV. In every case except the 15-year, these methods produce one of the lowest pricing errors, and in general are quite close to the lowest error. The other methods, which only produce reasonable estimates in only one or two cases (as with splining $\delta(\tau)$ with 3 knots at the 25-year maturity), tend to be much more erratic.

Finally, for this functional form we present plots of the actual zero coupon and implied forward rate curves along with the average point estimates and standard error bands for each of the methods. The results are contained in Figures 3 and 4. The figures give some idea of the character of the estimated curves across the various methods. For instance, in the upper right panel of Figure 3—the case of splining $\delta(\tau)$ with 6 knots—long maturity forward rates drop to nearly -10 percent and completely miss the curve after the five year horizon. Splining $f(\tau)$ with GCV tends to follow along with the curve everywhere except in the area with no data, while splining $\ell(\tau)$ with GCV creates an artificial flattening out of the forward rate function and splining $\delta(\tau)$ with GCV produces an upward sloping forward rate function at the 30-year maturity. Overall, splining $f(\tau)$ with GCV produces the most accurate estimates.

4 Estimation results

The Monte Carlo results in the previous section suggest that fixed knot point specifications, while accurate in some cases, are not flexible enough fit simple as well as complicated functional forms. With many knots, the potential for over-parameterization generates spurious movements in the estimated functional forms, while with few knots the potential for under-parameterization creates an inaccurate fit to all the twists and turns. The estimations using GCV provide better results across all types of functional forms, and hence in this section we only focus on the those methods.

Using daily data from December of 1987 until September of 1994 we construct GCV based estimates of the term structure with spline placements on $\delta(\tau)$, $\ell(\tau)$, and $f(\tau)$. On each day we use all coupon bearing Treasury securities which are non-callable, not flower bonds and have more than thirty days to maturity. The number of bonds in the sample varies from 160 in December 1987 to 180 in September 1994. As can be seen in the upper left panel of Figure 5, there exists a gap of acceptable bonds maturing between years 14 and 21. However, across the remainder of the maturity spectrum, there is a fairly even distribution of securities.

For an initial comparison, we present the fitted zero-coupon and forward rate curves for all three methods in the remaining panels of Figure 5. All methods produce similar zero-coupon and forward rate structures through the first five years, however, when splining $\delta(\tau)$ and $\ell(\tau)$ the forward rates move more abruptly in the 10-year and in the 20- to 25-year areas. In addition, when splining $\delta(\tau)$ and $\ell(\tau)$, the forward rates tend to flatten out in the 30-year region, which, based upon the results in Section 3, probably reflects bias: In every simulation in which the forward rates were not flat, placing the spline on $\ell(\tau)$ generated bias in the 30-year region.[24] In every simulation, placing the spline on $\delta(\tau)$ created bias in the 30-year region. In all likelihood, the two bottom panels of Figure 5 produce misleading results in long end of the yield curve.

Looking now at the estimation results across the entire sample, Figure 6 gives the average absolute pricing error, the effective number of parameters and the value of λ across each method for each day. Clearly, from the top panel, the average absolute pricing errors are virtually indistinguishable across methods. As noted in Section 3, average absolute pricing errors are not very useful in comparing alternative methods. Notice, however, that pricing has tended to become more accurate through time. Turning to the effective number parameters, the middle panel of Figure 6 indicates that except for early 1991 and early 1992 placing the spline on $\delta(\tau)$ tends to generate the greatest number of effective parameters while

[24]Recall that when penalizing movements in the squared second derivative of the logarithm of the discount function, we are effectively penalizing deviations in the forward rates from zero slope. Thus in regions with little data this creates a tendency to produce a flat forward rate function.

placing the spline on $f(\tau)$ results in the least. Interestingly, placing the spline on $\ell(\tau)$ and $f(\tau)$ produces a very parsimonious specification (only 2 to 3 effective parameters in some cases) in mid to late 1989, precisely when the yield curve was virtually flat. As the yield curved began to steepen in early 1991 and through 1992 the effective number of parameters began to increase, signifying a more complicated term structure. The bottom panel gives the values of the λ through time. Naturally, λ moves inversely to the effective number of parameters and there a sizeable differential in the magnitude when splining $f(\tau)$.

Finally, we present the evolution of the implied forward rates through time for each method by focusing only on the end of month values. Figures 7, 8 and 9 show the time series of forward rates when splining $f(\tau)$, $\ell(\tau)$ and $\delta(\tau)$, respectively. From Figure 7 we can clearly see the downward sloping term structure in mid 1989 followed by a gradual drop in short term rates starting in mid 1990. Most surprising is the downward sloping forward rate structure which begins to emerge precisely when short term rates begin to drop. The enormous drop of nearly 6 percentage points between the 24- and 30-year maturities since early 1991 has been one of so-called problems with spline based methods. Typically, forward rates are considered indicators of expected future spot rates and hence a 6 percentage drop in "expectations" between the 25 and 30 year maturities seems highly unlikely. However, recent work by Brown and Schafer (1994) and Gilles (1994) has shown that this is precisely what one would expect from forward rates due to the convexity bias. Basically, forward rates are a combination of expected future spot rates, a risk premium and a correction for convexity. The convexity term varies with the square of maturity and hence at long maturities can become quite large—even 6 percentage points. More unsettling than this extreme drop in forward rates in the lack of a drop in the earlier time periods as well as the remarkably flat long term forward rates throughout the entire sample when splining $\ell(\tau)$ in Figure 8. Based on the Monte Carlo results a good amount of this flattening could easily be artificial bias. The same holds true in Figure 9 when splining $\delta(\tau)$. With this technique, the forward rates actually tend to increase at the end. In general, the behavior of the long maturity forward rates and the results from the previous section suggest that splining $f(\tau)$ with GCV produces the least biased and most accurate estimates.

5 Conclusion

We have shown how to fit smoothing splines, which incorporate a penalty for roughness, to an arbitrary transformation of the term structure of interest rates, and we have shown how to choose the size of the penalty (and hence the effective number of parameters) by minimizing the generalized cross validation (GCV) value. Our Monte Carlo simulations indicate that fitting a smoothing spline to the instantaneous forward rate curve and using GCV to pick penalty provides the best estimator of the ones we examined. We estimated daily Treasury yield curves with seven years of data using our techniques and find the results reasonable in light of financial theory.

Appendices

A Imposing the restriction $\delta_s(0, \beta) = 1$

Since the value of a dollar today is a dollar, $\delta(0) = 1$. Consider imposing the restriction $\delta_s(0, \beta) = 1$ on the estimated spline. The content of the restriction depends on the functional form chosen for the spline, $h(\tau)$. When $h(\tau) = \delta(\tau)$, $\delta_s(0, \beta) = \phi(0)\,\beta = \beta_1$, since $\phi(0) = (1, 0, \cdots, 0)$. Thus the restriction amounts to $\beta_1 = 1$. When $h(\tau) = \ell(\tau)$, the restriction is $\delta_s(0, \beta) = \exp(\phi(0)\,\beta) = \exp(\beta_1)$. In this case the restriction amounts to $\beta_1 = 0$. Finally, when $h(\tau) = f(\tau)$, $\delta_s(0, \beta) = \exp(-\psi(0)\,\beta) = 1$ regardless of the value of β, since $\psi(0) = (0, 0, \cdots, 0)$. In this case, the restriction is automatically satisfied.

Linear restrictions on the coefficient vector β are typically expressed as $R\beta = r$. For the single restriction here, R is the row vector $(1, 0, \cdots, 0)$ of length k and either $r = 1$ or $r = 0$. For our purposes, it is convenient write β in terms of $\beta^\dagger = (\beta_2, \beta_3, \cdots, \beta_k)$, where the dagger indicates that the first element has been dropped. In particular, $\beta = \omega + \rho\beta^\dagger$, where $\omega = (r, 0, \cdots, 0)^\top$ is a vector of length κ and ρ is a $\kappa \times (\kappa - 1)$ matrix. The first row of ρ is $(0, 0, \cdots, 0)$ and the remaining rows form a $\kappa - 1$ order identity matrix.

Now we can impose the restriction on the minimization problem by substituting $\omega + \rho\beta^\dagger$ for β:[25]

$$
\begin{aligned}
&\min_{\beta^\dagger} \left[\Big(Y - X\,(\omega + \rho\beta^\dagger)\Big)^\top \Big(Y - X\,(\omega + \rho\beta^\dagger)\Big) + \lambda(\omega + \rho\beta^\dagger)^\top H(\omega + \rho\beta^\dagger) \right] \\
&= \min_{\beta^\dagger} \Big[(Y^\dagger - X^\dagger\beta^\dagger)^\top (Y^\dagger - X^\dagger\beta^\dagger) + \lambda(\omega + \rho\beta^\dagger)^\top H(\omega + \rho\beta^\dagger) \Big],
\end{aligned}
$$

where $Y^\dagger = Y - X\omega$ and $X^\dagger = X\rho$. The solution to this problem is

$$
\widehat{\beta}^\dagger = (X^{\dagger\top} X^\dagger + \lambda H^\dagger)^{-1} (X^{\dagger\top} Y^\dagger - \lambda\,\rho^\top H\,\omega), \tag{A.1}
$$

where $H^\dagger = \rho^\top H\rho$. As long as either $\lambda = 0$ or $r = 0$,[26] (A.1) has the same form as (4), and

[25] The dependence of X on β and of β on λ has been suppressed in this appendix for notational simplicity.
[26] When $r = 0$, $\omega = (0, 0, \cdots, 0)$.

the solution techniques described in the body of the paper can be applied. The only case that requires special treatment is fitting a smoothing spline on $h(\tau) = \delta(\tau)$. In that case, $r = 1$ and $\lambda \neq 0$.[27]

B Simplifications for $\mathrm{tr}(A(\lambda))$ and $\gamma(\lambda)$

In this appendix, we first derive a simplified expression for $\mathrm{tr}(A(\lambda))$, the effective number of parameters, and then we derive an algebraic simplification for $\gamma(\lambda)$.[28]

Letting $V := (X^\top X)^{-\frac{1}{2}}$ we can write

$$\begin{aligned} X^\top X + \lambda H &= V^{-2} + \lambda\, H \\ &= V^{-1}(I + \lambda\, V\, H\, V)V^{-1} \\ &= V^{-1}(I + \lambda U\, D\, U^\top)V^{-1} \\ &= V^{-1}U(I + \lambda\, D)U^\top V^{-1}, \end{aligned} \tag{B.1}$$

where $U\, D\, U^\top$ is the singular value decomposition of $V\, H\, V$. D is a diagonal matrix of eigenvalues, two of which are zero since H has two zero eigenvalues and V is nonsingular. Using (B.1), write

$$(X^\top X + \lambda H)^{-1} = V\, U(I + \lambda\, D)^{-1}U^\top V = G(I + \lambda\, D)^{-1}G^\top,$$

where $G := V\, U$. Now we can write

$$A(\lambda) = X\, G(I + \lambda\, D)^{-1}G^\top X^\top = \Omega(I + \lambda D)^{-1}\Omega^\top,$$

[27]Suppose $h(\tau) = f(\tau)$ and one wished to impose the restriction $f(0) = f_0$. This amounts to the restriction $\beta_1 = f_0$, and (A.1) applies with $\omega = (f_0, 0, \cdots, 0)$ and ρ unchanged.

[28]The dependence of X on β has been suppressed in this appendix for notational simplicity.

where $\Omega := X\,V\,U$. Notice that $\Omega^\top \Omega = I$.[29] Now we can write

$$\operatorname{tr}(A[\lambda]) = \operatorname{tr}(\Omega(I+\lambda D)^{-1}\Omega^\top) = \sum_{i=1}^{\kappa} \frac{1}{1+\lambda D_{ii}} = \iota^\top d(\lambda), \tag{B.2}$$

where $d(\lambda) := \operatorname{diag}(I+\lambda\,D)^{-1}$, the vector constructed from the main diagonal of $(I+\lambda\,D)^{-1}$, and $\iota := (1,1,\cdots,1)$. In (B.2) we can clearly see that $\operatorname{tr}(A(0)) = \kappa$ and $\operatorname{tr}(A(\infty)) = 2$.

Equation (B.2) can be used to simplify the denominator of $\gamma(\lambda)$. We now turn to simplifying the numerator. The numerator of $\gamma(\lambda)$, the residual sum of squares, can be written as

$$\begin{aligned}
&\Big((I-A(\lambda))Y\Big)^\top \Big((I-A(\lambda))Y\Big) \\
&= Y^\top Y - 2Y^\top A(\lambda)Y + Y^\top A(\lambda)^\top A(\lambda)Y \\
&= Y^\top Y - 2Y^\top \Omega(I+\lambda D)^{-1}\Omega^\top Y + Y^\top \Omega(I+\lambda D)^{-2}\Omega^\top Y \\
&= Y^\top Y - 2(y^2)^\top d(\lambda) + (y^2)^\top d(\lambda)^2,
\end{aligned}$$

where $y := \Omega^\top Y$ and $y^2 = y\,y$, with element-by-element multiplication. The computationally simplified expression for GCV is:

$$\gamma(\lambda) = \frac{Y^\top Y - 2(y^2)^\top d(\lambda) + (y^2)^\top d(\lambda)^2}{(n - \iota^\top d(\lambda))^2}. \tag{B.3}$$

When there is a restriction, the residual sum of squares is $(Y^\dagger - X^\dagger\,\hat\beta^\dagger)^\top (Y^\dagger - X^\dagger\,\hat\beta^\dagger)$, where $\hat\beta^\dagger$ is given in (A.1). This introduces additional terms in the numerator. In this case, the computationally simplified expression of GCV is

$$\gamma(\lambda) = \frac{(Y^\dagger)^\top Y^\dagger + 2(\lambda\,h\,y^\dagger - {y^\dagger}^2)^\top d^\dagger(\lambda) + (\lambda^2\,h^2 - 2\,\lambda\,h\,y^\dagger + {y^\dagger}^2)^\top d^\dagger(\lambda)^2}{(n - \iota^\top d^\dagger(\lambda))^2}, \tag{B.4}$$

where $h := U^\top V\,\rho^\top H\,\omega$ and $y^\dagger$ and $d^\dagger(\lambda)$ are defined in the obvious ways. Note that when there are no restrictions, $h = 0$, and (B.4) reduces to (B.3). Finally, note that expressions

[29] Also notice that $\Omega\Omega^\top Y$ is the OLS fit from the linear regression $Y = X\beta + \varepsilon$.

(B.3) and (B.4) are only useful when X—and hence D—is independent of β. But this is true only when $g(h(\cdot), \tau)$ is linear in $h(\cdot)$.

References

[1] Adams, K. J., and D. R. Van Deventer. "Fitting Yield Curves and Forward Rate Curves with Maximum Smoothness." *Journal of Fixed Income*, June (1994): 52–62.

[2] Bliss, R. R., Jr. "Testing Term Structure Estimation Methods." Working Paper, Indiana University (1993).

[3] Chambers, D., W. Carleton and D. Waldman. "A New Approach to Estimation of Term Structure of Interest Rates." *Journal of Financial and Quantitative Studies*, 19 No. 3 (1984): 233–252.

[4] Chow, G. C. *Econometrics* McGraw-Hill, New York (1983).

[5] Coleman, T. S., L. Fisher and R. Ibbotson. "Estimating the Term Structure of Interest From Data That Include the Prices of Coupon Bonds." *The Journal of Fixed Income*, September (1992): 85–116.

[6] de Boor, C., *A Practical Guide to Splines*, Springer-Verlag, (1978).

[7] Fisher, M., and D. Zervos. "Fitting the Term Structure." Forthcoming as a chapter in H. Varian, ed., *Economic and Financial Modeling with Mathematica*, vol. 2.

[8] Gilles, C. "Forward Rates and Expected Future Short Rates." Working paper, Federal Reserve Board (1994).

[9] Jordan, J.V. "Tax Effects in Term Structure Estimation." *Journal of Finance*, 39 No. 2 (1984): 393–406.

[10] McCulloch, J. H. "Measuring the Term Structure of Interest Rates." *Journal of Business*, 44 (1971): 19–31.

[11] McCulloch, J. H. "The Tax-Adjusted Yield Curve." *Journal of Finance*, 30 (1975): 811–830.

[12] Nelson, C. and A. Siegel. "Parsimonious Modeling of Yield Curves." *Journal of Business*, 60 No. 4 (1987): 473–489.

[13] Shea, G. "Pitfalls in Smoothing Interest Rate Term Structure Data: Equilibrium Models and Spline Approximations." *Journal of Financial and Quantitative Studies*, 19 No. 3 (1984): 253–269.

[14] Shea, G. "Interest Rate Term Structure Estimation with Exponential Splines: A Note." *Journal of Finance*, 40 No. 1 (1985): 319–325.

[15] Vasicek, O. and G. Fong. "Term Structure Estimation Using Exponential Splines." *Journal of Finance*, 38 (1982): 339–348.

[16] Wahba, G. *Spline Models for Observational Data* SIAM: Philadelphia (1990).

Table 1—Monte Carlo Results For $f_1(\tau)$

	Knots = 3			Knots = 6			Knots = 10			GCV		
Spline Placement	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$
Effective Params.	5	5	5	8	8	8	12	12	12	11.0	1.0	2.0
Avg. Abs. Price Err.	9.3	8.0	7.9	7.9	7.8	7.7	7.6	7.7	7.7	7.8	8.0	7.9
Fwd. IMAE	29.1	0.2	0.3	15.1	0.5	0.5	7.3	1.1	2.2	6.1	0.0	0.0
Zero IMAE	3.8	0.1	0.1	1.4	0.2	0.1	1.0	0.2	0.3	0.7	0.0	0.0
Zero Curve												
2-year Bias	1.6	0.1	0.0	–0.2	–0.3	–0.0	0.3	–0.1	0.3	0.3	0.0	–0.0
2-year Std.	0.6	0.7	0.6	1.0	0.9	1.0	1.0	1.2	1.2	0.8	0.2	0.4
5-year Bias	–1.7	0.0	–0.0	0.5	–0.1	0.0	0.4	0.1	0.2	–0.1	0.0	–0.0
5-year Std.	0.3	0.4	0.4	0.6	0.6	0.5	0.5	0.7	0.7	0.5	0.2	0.3
10-year Bias	4.4	–0.0	0.1	–0.0	0.2	–0.1	0.3	–0.2	0.1	0.1	0.0	–0.0
10-year Std.	0.4	0.5	0.7	0.8	0.9	0.8	0.9	0.9	0.9	0.9	0.2	0.2
30-year Bias	11.6	0.2	–0.1	6.5	–0.3	0.2	–2.7	–0.5	0.9	3.7	0.0	–0.0
30-year Std.	1.8	1.9	1.7	1.5	2.0	2.1	3.6	2.8	4.6	3.2	0.2	0.4
Forward Curve												
2-year Bias	–5.1	–0.0	–0.2	–2.0	0.2	–0.2	–0.0	0.0	0.5	1.6	0.0	–0.0
2-year Std.	0.4	0.7	1.3	2.7	3.0	3.3	6.4	6.7	5.1	2.1	0.2	0.3
5-year Bias	2.4	–0.1	0.1	–1.0	0.2	0.4	1.7	0.5	–0.0	–0.1	0.0	–0.0
5-year Std.	0.7	0.7	1.1	1.7	1.8	1.7	3.8	4.3	2.7	3.2	0.2	0.2
10-year Bias	14.4	–0.0	0.2	6.2	0.3	–0.6	5.7	–0.3	1.3	2.2	0.0	–0.0
10-year Std.	0.6	1.3	1.2	1.7	1.3	3.3	1.8	1.6	4.5	3.7	0.2	0.2
30-year Bias	294.5	0.8	–0.4	167.3	–1.4	–0.3	–67.9	–10.5	21.4	123.9	0.0	–0.1
30-year Std.	19.7	11.6	18.3	21.5	13.1	31.5	84.0	52.5	114.1	50.9	0.2	1.2

Average Absolute Pricing Errors For Synthetic 7% Coupon Securities

	Knots = 3			Knots = 6			Knots = 10			GCV		
Spline Placement	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$
2-year bond	3.4	0.3	0.0	0.3	0.5	0.5	0.5	0.2	0.5	0.2	0.0	0.0
5-year bond	6.3	0.1	0.2	1.9	0.3	0.1	1.4	0.4	0.8	0.5	0.1	0.0
10-year bond	23.6	0.2	0.3	0.7	0.8	0.3	1.4	0.9	0.5	0.3	0.1	0.1
15-year bond	41.6	0.1	0.6	15.5	1.0	1.3	13.1	0.8	2.7	5.0	0.1	0.1
20-year bond	16.8	0.1	0.3	10.8	0.5	1.0	4.9	0.2	1.3	0.6	0.2	0.2
25-year bond	14.8	0.5	0.1	7.1	0.2	0.0	2.2	0.5	0.7	0.3	0.2	0.2
30-year bond	48.5	1.0	0.5	25.5	1.0	0.4	7.4	2.1	3.5	13.4	0.2	0.3

Pricing errors in basis points, rate errors and biases in basis points per year.

Table 2—Monte Carlo Results For $f_2(\tau)$

	Knots = 3			Knots = 6			Knots = 10			GCV		
Spline Placement	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$
Effective Params.	5	5	5	8	8	8	12	12	12	8.3	7.1	2.0
Avg. Abs. Price Err.	8.1	7.9	7.7	7.9	7.8	7.7	7.8	8.0	7.5	7.9	7.9	7.8
Fwd. IMAE	12.5	5.7	3.0	0.3	0.2	0.5	0.5	0.5	1.1	10.0	5.7	0.0
Zero IMAE	1.6	0.5	0.4	0.1	0.1	0.1	0.1	0.1	0.2	0.5	0.5	0.0

Zero Curve

2-year	Bias	–0.6	0.0	–0.1	0.1	–0.2	–0.1	0.1	–0.1	–0.0	–0.1	0.0	–0.0
	Std.	0.5	1.1	0.9	0.6	0.8	1.3	0.6	1.0	1.4	0.6	0.8	0.3
5-year	Bias	0.7	–0.2	–0.3	–0.0	0.0	–0.1	0.0	–0.0	–0.0	–0.0	–0.1	–0.0
	Std.	0.4	0.5	0.5	0.4	0.4	0.5	0.3	0.4	0.7	0.4	0.5	0.3
10-year	Bias	–2.0	0.1	–0.2	0.0	–0.0	–0.0	–0.1	–0.1	–0.2	–0.1	–0.1	–0.0
	Std.	0.4	0.7	0.6	0.4	0.6	0.6	0.6	0.8	0.8	0.8	0.7	0.2
30-year	Bias	–4.9	–2.3	1.1	–0.3	0.0	–0.2	–0.1	–0.2	0.6	8.3	–4.6	0.0
	Std.	1.8	1.8	4.0	1.3	1.3	3.0	1.7	2.1	4.4	2.7	1.1	0.4

Forward Curve

2-year	Bias	2.4	0.9	–0.4	–0.1	–0.0	–0.1	0.2	0.1	–1.0	0.9	–0.9	–0.0
	Std.	0.5	3.0	6.6	0.7	2.9	6.1	1.2	3.5	4.9	1.2	1.4	0.3
5-year	Bias	–1.4	0.7	–0.2	–0.0	–0.0	0.5	–0.3	–0.3	0.4	–0.3	0.3	–0.0
	Std.	0.7	1.4	3.1	0.7	1.2	3.0	0.9	2.0	2.3	1.8	2.2	0.2
10-year	Bias	–6.0	–2.6	–2.5	0.1	0.1	–0.0	–0.2	0.1	–1.0	–0.8	–0.3	0.0
	Std.	0.5	1.3	1.2	0.9	0.8	1.3	1.1	3.5	4.8	2.3	1.6	0.2
30-year	Bias	–124.6	–61.6	30.0	–1.5	–0.7	–4.7	–2.9	–2.9	8.4	219.4	–62.6	0.0
	Std.	20.0	21.0	94.9	8.1	9.5	56.6	16.3	31.9	113.7	36.7	6.3	1.0

Average Absolute Pricing Errors For Synthetic 7% Coupon Securities

	Knots = 3			Knots = 6			Knots = 10			GCV		
Spline Placement	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$
2-year bond	1.4	0.1	0.2	0.1	0.5	0.1	0.2	0.3	0.0	0.2	0.1	0.0
5-year bond	3.0	0.7	1.1	0.1	0.1	0.5	0.2	0.2	0.1	0.1	0.4	0.1
10-year bond	12.5	1.3	0.8	0.1	0.4	0.2	0.9	0.9	0.9	0.7	0.3	0.1
15-year bond	21.1	6.7	6.8	0.4	0.1	0.3	0.8	0.1	2.7	2.8	0.5	0.0
20-year bond	8.3	4.7	2.7	0.3	0.6	0.2	0.0	0.6	1.7	0.5	1.3	0.0
25-year bond	6.9	3.1	1.1	0.2	0.7	0.0	0.3	0.5	0.3	1.8	2.7	0.1
30-year bond	22.4	9.5	2.9	1.0	0.4	1.1	0.6	0.7	1.4	29.5	17.4	0.1

Pricing errors in basis points, rate errors and biases in basis points per year.

TABLE 3—MONTE CARLO RESULTS FOR $f_3(\tau)$

	Knots = 3			Knots = 6			Knots = 10			GCV		
Spline Placement	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$
Effective Params.	5	5	5	8	8	8	12	12	12	8.8	9.0	5.4
Avg. Abs. Price Err.	14.3	7.9	7.9	8.2	7.9	7.8	7.9	8.0	7.6	7.8	7.8	7.9
Fwd. IMAE	45.3	0.2	0.1	16.7	0.4	0.4	11.6	0.3	0.9	9.8	8.0	4.1
Zero IMAE	7.5	0.0	0.1	2.0	0.1	0.1	1.6	0.1	0.1	0.9	0.7	0.5

ZERO CURVE

2-year	Bias	−3.6	0.0	0.2	−0.0	−0.0	−0.0	0.2	0.1	−0.1	0.3	−0.2	0.2
	Std.	0.5	0.8	0.6	0.9	0.9	0.8	0.9	1.2	1.2	0.6	0.8	0.7
5-year	Bias	3.9	0.0	0.0	−0.8	−0.1	−0.1	−0.9	−0.1	−0.0	−0.0	−0.1	−0.2
	Std.	0.4	0.3	0.3	0.5	0.5	0.5	0.5	0.6	0.6	0.4	0.5	0.4
10-year	Bias	−10.2	0.0	0.0	−0.2	0.2	0.1	−0.8	0.0	−0.1	−0.3	0.2	0.3
	Std.	0.3	0.3	0.6	0.8	0.6	0.6	0.7	0.6	0.8	0.7	0.7	0.5
30-year	Bias	−13.1	−0.2	−0.0	−5.1	−0.0	0.1	3.9	−0.1	0.5	6.3	5.2	2.0
	Std.	1.1	1.0	1.1	1.1	1.4	1.3	1.9	2.1	3.3	2.1	1.7	1.6

FORWARD CURVE

2-year	Bias	13.6	−0.0	−0.1	4.0	0.7	0.5	0.0	−0.5	−0.7	−1.8	0.0	−0.8
	Std.	0.4	0.5	1.0	2.5	2.4	2.9	5.9	7.0	4.0	1.3	2.3	0.7
5-year	Bias	−7.7	0.0	−0.0	2.4	−0.0	−0.2	−3.9	0.3	0.0	−0.2	0.1	−0.1
	Std.	0.6	0.6	0.9	1.2	1.2	2.2	2.9	3.7	2.3	1.4	2.4	0.8
10-year	Bias	−29.0	0.1	0.0	−12.3	0.2	0.5	−11.8	0.1	−0.0	−5.2	1.8	1.8
	Std.	0.4	0.8	1.0	1.5	1.0	2.8	1.3	1.2	4.0	1.8	2.7	1.8
30-year	Bias	−310.0	−0.9	0.3	−120.2	0.9	2.0	95.3	−1.4	13.2	152.7	111.6	43.6
	Std.	8.9	6.6	12.8	16.4	10.3	20.9	47.8	38.4	81.6	26.3	16.4	19.3

AVERAGE ABSOLUTE PRICING ERRORS FOR SYNTHETIC 7% COUPON SECURITIES

	Knots = 3			Knots = 6			Knots = 10			GCV		
Spline Placement	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$
2-year bond	8.1	0.0	0.4	0.2	0.0	0.1	0.4	0.3	0.2	0.9	0.2	0.5
5-year bond	16.5	0.0	0.2	3.3	0.4	0.6	3.8	0.6	0.2	0.1	0.3	0.7
10-year bond	65.9	0.1	0.2	0.0	1.1	0.3	4.6	0.1	0.7	1.3	1.3	1.8
15-year bond	101.1	0.2	0.1	31.5	0.8	1.2	29.3	0.1	1.1	13.1	5.7	5.3
20-year bond	34.0	0.0	0.1	17.5	0.1	0.7	9.1	0.2	0.1	0.5	3.7	2.6
25-year bond	30.4	0.4	0.3	10.0	0.2	0.1	3.9	0.2	0.3	1.5	4.3	2.6
30-year bond	90.3	1.2	0.3	32.5	0.1	0.6	15.2	0.3	2.6	31.1	28.1	11.7

Pricing errors in basis points, rate errors and biases in basis points per year.

TABLE 4—MONTE CARLO RESULTS FOR $f_4(\tau)$

	Knots = 3			Knots = 6			Knots = 10			GCV		
Spline Placement	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$
Effective Params.	5	5	5	8	8	8	12	12	12	15.4	16.4	12.8
Avg. Abs. Price Err.	41.2	51.6	23.4	14.7	13.8	12.7	11.1	11.3	10.9	7.8	7.7	7.8
Fwd. IMAE	97.8	181.6	93.7	162.5	91.0	106.7	129.1	72.6	100.7	67.4	55.8	48.4
Zero IMAE	20.1	26.4	18.1	19.5	10.5	17.0	17.7	10.9	15.4	11.0	7.8	8.7

ZERO CURVE

2-year	Bias	13.9	11.6	11.0	−0.2	−0.7	−3.6	0.2	−0.4	−0.9	0.1	0.1	−0.1
	Std.	0.5	0.6	0.6	0.9	0.9	1.0	1.0	1.0	1.1	1.1	1.4	0.8
5-year	Bias	−17.0	−15.6	−9.0	−1.9	1.6	−2.8	1.3	5.0	1.4	−0.2	−0.1	−0.1
	Std.	0.2	0.3	0.3	0.5	0.4	0.4	0.6	0.6	0.5	0.6	0.7	0.6
10-year	Bias	43.2	48.8	16.9	13.0	15.2	9.9	9.4	12.6	8.9	5.5	3.2	4.2
	Std.	0.3	0.4	0.5	0.7	0.6	0.7	0.5	0.7	0.8	0.7	0.9	0.8
30-year	Bias	8.31	87.9	9.6	−27.3	29.0	3.6	37.5	4.7	23.9	15.4	19.1	7.9
	Std.	1.5	1.6	1.7	1.4	1.7	1.9	3.6	2.8	4.6	3.9	2.7	4.1

FORWARD CURVE

2-year	Bias	−37.0	36.5	36.9	−0.2	−19.6	−8.5	−2.6	−7.2	3.0	−0.4	−1.4	0.3
	Std.	0.3	0.5	1.1	2.4	2.2	3.1	6.7	6.8	5.1	3.8	6.5	3.3
5-year	Bias	14.2	22.4	−30.2	−17.2	−38.4	−28.8	−5.8	2.4	−24.6	−0.8	0.4	0.2
	Std.	0.5	0.6	0.8	1.1	1.1	1.9	3.8	3.8	1.8	4.4	7.8	4.8
10-year	Bias	33.4	40.5	−4.3	−15.9	48.4	−24.3	−13.0	37.0	−14.6	−2.8	19.9	13.1
	Std.	0.4	1.0	0.8	1.2	1.1	2.7	1.2	1.8	3.7	3.5	3.6	5.7
30-year	Bias	−86.3	1027.0	−83.0	−870.9	396.4	−244.8	666.7	−149.4	304.0	252.5	339.5	−24.4
	Std.	12.3	9.5	18.3	16.6	12.0	25.9	88.7	60.1	111.2	74.3	37.5	86.3

AVERAGE ABSOLUTE PRICING ERRORS FOR SYNTHETIC 7% COUPON SECURITIES

	Knots = 3			Knots = 6			Knots = 10			GCV		
Spline Placement	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$	$\delta(\tau)$	$\ell(\tau)$	$f(\tau)$
2-year bond	30.7	25.9	19.6	0.5	1.2	6.8	0.3	0.7	1.6	0.4	0.3	0.1
5-year bond	73.7	68.3	37.2	7.7	7.2	11.1	5.6	20.7	6.1	0.9	0.3	0.6
10-year bond	313.0	355.9	119.8	94.9	102.8	71.4	67.3	85.7	63.1	39.5	22.1	28.8
15-year bond	139.8	109.7	283.8	370.1	160.9	345.2	357.1	202.1	319.0	259.6	178.1	206.2
20-year bond	127.6	169.8	105.0	180.9	53.9	114.9	129.6	77.9	95.1	4.8	22.3	26.2
25-year bond	1.1	47.4	25.5	69.6	4.7	38.6	37.3	16.7	27.7	10.6	6.8	7.5
30-year bond	20.9	304.3	2.1	174.1	91.4	36.9	87.6	10.0	45.9	29.2	51.5	4.7

Pricing errors in basis points, rate errors and biases in basis points per year.