

Estimating the Interest Rate Term Structures of Treasury and Corporate Debt with Bayesian Penalized Splines

Min Li¹ and Yan Yu²

¹California State University, Sacramento and ²University of Cincinnati

Abstract: This paper provides a Bayesian approach to estimating the interest rate term structures of Treasury and corporate debt with a penalized spline model. Although the literature on term structure modeling is vast, to the best of our knowledge, all methods developed so far belong to the frequentist school. In this paper, we develop a two-step estimation procedure from a Bayesian perspective. The Treasury term structure is first estimated with a Bayesian penalized spline model. The smoothing parameter is naturally embedded in the model as a ratio of posterior variances and does not need to be selected as in the frequentist approach. The corporate term structure is then estimated by adding a credit spread to the estimated Treasury term structure, incorporating knowledge of the positive credit spread into the Bayesian model as an informative prior. In contrast to the frequentist method, the small sample size of the corporate debt poses no particular difficulty to the proposed Bayesian approach.

Key words: Coupon bonds, credit spreads, forward rate, Markov chain Monte Carlo, penalized splines, Treasury bonds.

1. Introduction

This paper presents a term structure estimation method using Bayesian penalized splines. The term structure of interest rates describes how the interest rate of bonds evolves over time. This interest rate is implicitly determined by the price of the bonds through any of the discount function $D(0, T)$, the yield curve $y(0, T)$, or the forward rate $f(0, T)$, where T is time to maturity from today (time 0).

To simplify the notation, we consider current time to be fixed at 0 and omit 0. The discount function $D(T)$ represents today's price of a *zero coupon bond* (a bond that pays no interest or principal until maturity and then pays a fixed amount called the par value) that pays one dollar at maturity time T . The yield curve $y(T) = \int_0^T f(s)ds/T$ gives the average of the forward rate between today (time 0) and the maturity date T . The *forward* rate $f(s)$ is a variable interest

rate (changing over time s). It is the interest rate one can lock in today for future borrowing or lending at time s . The relationships among these three functions are:

$$D(T) = \exp\{-Ty(T)\} = \exp\left\{-\int_0^T f(s)ds\right\}. \quad (1.1)$$

Availability of any one of the three determines the rest.

To estimate the term structure, a spline may be used to model the discount function $D(T)$, the yield curve $y(T)$, or the forward rate $f(s)$. However, Fisher, Nychka, and Zervos (1995) conclude that modeling the forward rate using a spline gives the most accurate estimation. Moreover, modeling the forward rate $f(s)$ with a spline naturally embeds the constraint that a dollar paid today is worth a dollar: $D(0) = 1$. On the other hand, modeling the discount function with splines requires the constraint $D(0) = 1$ to be imposed. Shea (1985) notes serious problems when modeling the discount function $D(T)$ with splines: negative forward rates and instability at the long maturities. Moreover, for monetary policy purposes, estimating the forward rate may be the objective (Anderson and Sleath 1999). Therefore, only the term structure of the forward interest rate is considered in this paper.

The term structure of government debt such as US Treasury bonds is of interest to monetary authorities as it contains information about current macroeconomic conditions and market participants' expectation of future economic conditions. Treasury bonds are virtually risk free as they are backed by the taxing power of the government. The term structure of such bonds provides a benchmark for analysts and traders to assess other securities. The term structure can also be used to price financial products such as bonds, swaps, etc. To maximize returns, fixed-income money managers vary their portfolio based on the term structure as it reveals the volatility in the interest rate along the time line.

Corporate bonds bear credit risk in that a company may default on its debt. Due to increased trading in instruments with credit risk, credit derivatives were created to partially or fully offset the credit risk of a business deal. Estimation of the corporate term structure is becoming increasingly important in financial markets as the market for credit derivatives grows explosively. The size of the credit derivatives market in 2001 reached around 835.5 billion dollars. Pricing models for both corporate debt and credit derivatives require the corporate term structure as input (see Jarrow and Turnbull 1995; Duffie and Singleton 1999).

Much research has been done in term structure estimation. McCulloch (1971, 1975) and Shea (1984) estimate the discount function with regression splines. Vasicek and Fong (1982) and Shea (1985) estimate the discount function with exponential splines. Chambers, Carleton, and Waldman (1984) estimate the yield curve with polynomials. Adams and Van Deventer (1994), Fisher *et al.* (1995), Tanggaard (1997), and Waggoner (1997) estimate term structure using a

smoothing spline model, minimizing a penalized sum-of-squared-errors criterion. Schwartz (1998) adopts a robust criterion in eliminating outliers and estimates the term structure of AAA credit class corporate debt by fitting the forward rate with a piecewise constant curve. Linton, Mammen, Nielsen, and Tanggaard (2001) present kernel methods in term structure estimation and establish asymptotic properties. JRY (Jarrow, Ruppert, and Yu 2004) estimate the term structure of individual corporate bonds when the sample size is very small (typically 4 or 5). They adopt a semi-parametric model to estimate the individual corporate debt term structure. They estimate the Treasury term structure with a nonparametric penalized spline model by minimizing the sum of the least squares criterion and a roughness penalty term. Due to the limited number of individual corporate bonds available, they obtain the individual corporate debt term structure by adding a credit spread to the Treasury term structure.

To the best of our knowledge, all of the term structure modeling approaches that have appeared in the literature belong to the frequentist school. In this paper, we provide a Bayesian approach to estimating Treasury and corporate term structures with a penalized spline (P-splines) model. Eilers and Marx (1996), Ruppert and Carroll (1997), and Ruppert, Wand, and Carroll (2003) present penalized splines in detail. The models considered in these works are all linear in nature. The term structure models we estimate are *nonlinear* since the forward rate term appears exponentially in the discount function. We develop a Bayesian penalized spline model and employ Markov chain Monte Carlo (MCMC) techniques in our term structures estimation.

We model the forward rate curve f with a spline $f = \boldsymbol{\delta}'B$, where $\boldsymbol{\delta}$ is a vector of spline coefficients and B is a matrix of spline basis functions such as power basis or B -splines. We use penalized splines to estimate the forward rate term structure of Treasury and corporate debt. The classical penalized splines minimize the sum of the lack of fit of the data and a roughness penalty measure. The forward rate $f = \boldsymbol{\delta}'B$ is estimated by minimizing

$$\frac{1}{n} \sum_{i=1}^n [P_i - \mu_i(\boldsymbol{\delta})]^2 + \alpha \rho(\boldsymbol{\delta}) \quad (1.2)$$

where P_i is the observed bond price, $\mu_i(\boldsymbol{\delta})$ is the model price from the penalized spline model, α is the smoothing parameter, and $\rho(\boldsymbol{\delta})$ is a roughness penalty measure. The roughness penalty measure is a function of the spline coefficients. The smoothing parameter α is chosen to control the trade-off between the lack of fit and the roughness measure.

The roughness measure can also be treated as the prior information on the spline coefficients within a Bayesian framework. It represents our subjective knowledge about the spline coefficients while the lack of fit term is our attempt

to estimate the spline coefficients with the existing data. This is the approach we take.

We also estimate the corporate term structure by adding a credit spread to the estimated Treasury term structure as in JRY. A credit spread represents the excess return from a corporate bond over the return from an equivalent Treasury bond. It can be used to assess credit quality for related uses in risk management procedures (see Jarrow 2001). Credit quality assessment is essential for bond portfolio management, value at risk computations, and even FDIC insurance premium calculation¹. An estimate of the credit spread gives an approximation of the term structure of the individual corporate bond. This term structure can be used to price and hedge corporate debt, a common risk management practice in business (see Jarrow 2002).

There are usually quite many Treasury bonds available but far fewer individual corporate bonds in trading. Borrowing strength from the Treasury term structure to estimate corporate term structures is natural. A Bayesian approach facilitates this “borrowing strength” nicely. We include the information that this credit spread should be positive because corporate securities are riskier than Treasury securities. The knowledge that this credit spread is positive can be easily incorporated into the Bayesian model as prior information. On the other hand, within the frequentist framework, forcing the credit spread to be positive is equivalent to adding n constraints to the minimization criterion aforementioned, where n is the number of corporate bonds available. It is possible to solve this constrained nonlinear optimization problem. However, the small sample size of corporate bonds available on a given day is 4 or 5 on average (JRY). This makes it difficult to justify the frequentist statistical inference procedures as they rely on the assumption of large samples. On the other hand, the Bayesian approach does not rely on this kind of assumption.

Compared with the frequentist approach, the Bayesian approach has several advantages. First, frequentist large sample asymptotic statistical procedures often break down when the sample size is small. With a Bayesian approach, the small sample size of individual corporate bond data poses no problem in statistical inference. Second, we are able to incorporate important information of positive credit spreads between corporate and Treasury securities into the model through the adoption of an informative prior distribution. Third, the role of the smoothing parameter is diminished. The smoothing parameter is used to control the trade-off between lack of fit and overfitting. A common criterion to select the smoothing parameter in the frequentist approach is generalized cross validation (GCV) or cross validation (C_p) through grid search. With the Bayesian approach, the smoothing parameter is obtained as a by-product and does not need to be

¹See: <http://www.fdic.gov/deposit/insurance/initiative/OptionPaper.html>.

selected.

The remainder of the paper is organized as follows. We describe the fixed income data base in Section 2 and present a Bayesian approach to estimating the Treasury term structure with penalized splines in Section 3. We then illustrate the Bayesian estimation of the corporate term structure with informative priors in Section 4. Some concluding remarks follow in Section 5.

Table 1: A Sample of five of the 117 US Treasury STRIPS on December 31, 1995.

Current Date	Issue Date	Maturity Date	Market Price
19951231	19850215	19960515	98.155
19951231	19850215	19961115	95.717
19951231	19850215	19970515	93.215
19951231	19850215	19971115	90.852
19951231	19850215	19980515	88.481

Table 2: AT&T Bonds on December 31, 1995. Coupon listed is the semi-annual dollar coupon payment.

Current Date	Issue Date	Maturity Date	Market Price	Coupon
19951231	19920114	20020115	109.458	7.125
19951231	19940324	20040401	106.284	6.750
19951231	19940601	20060601	111.436	7.500
19951231	19950228	20070301	115.509	7.750
19951231	19950512	20050515	107.659	7.000

2. Data

The University of Houston Fixed Income database contains over 28,000 instruments, including publicly traded non-convertible debt with principal value no less than one million dollars. Warga (1995) gives a detailed description of the database. The bond data that make up the Lehman Brothers Bond Indices are reported with month-end flat prices, accrued interest, coupon, yields, current date, issuance date, maturity date, S&P and Moody's ratings, and option-like features. The market price of a corporate bond equals the quoted flat price plus the accrued interest. The data we use are US Treasury STRIPS (Separate Trading of Registered Interest and Principal of Securities) and AT&T Bonds from April 1994 to December 1995. The US Treasury STRIPS (coupon and principal

STRIPS) are zero coupon bonds that are synthesized from the coupon and principal payments of Treasury bonds. Table 1 shows the market prices of five of the 117 US Treasury STRIPS on December 31, 1995 in the fixed income data base.

Table 2 shows the market prices of all five AT&T bonds available on December 31, 1995:

There are usually few individual corporate bonds available in a given month. JRY observe that on average there were only 4.3 AT&T bonds available per month during the period of April 1994 to December 1995. It is difficult to obtain a meaningful estimate of the corporate term structure based on these few observations. On the other hand, the average number of US Treasury STRIPS per month during the same period was 117, ranging from 115 to 120. Borrowing strength from other sources such as US Treasury STRIPS becomes necessary in estimating the term structure of individual corporate bonds.

For estimation, time-to-maturity and the coupon payment times need to be converted to the same unit scale. These can be easily accomplished with the Matlab finance toolbox functions `days365()` and `cfdates()`. `Days365()` counts the number of days between dates based on 365-day year and `cfdates()` gives cash flow dates for a fixed-income security. We use `days365()` and `cfdates()` based on conventional actual/365 day count.

3. A Bayesian Model for the Term Structure of Treasury Bonds

3.1 Model

In this section, we estimate the term structure of US Treasury bonds using Bayesian penalized splines. Let $P_i, i = 1, \dots, n$, be the market price of the i -th bond at time 0. Each bond pays fixed coupons and principal $C_i(t_{ij})$ due on dates t_{ij} , where $j = 1, \dots, z_i$ and z_i is the total number of coupon and principal payments for the i th bond. Thus, t_{i1} is the first coupon payment date and t_{iz_i} is the maturity date of bond i . According to the theory of term structure, the model price for the coupon bond, μ_i , is related to the forward rate f through the discount function:

$$\mu_i = \sum_{j=1}^{z_i} C_i(t_{ij})D(t_{ij}) = \sum_{j=1}^{z_i} C_i(t_{ij}) \exp\left\{-\int_0^{t_{ij}} f(s)ds\right\}, \quad (3.1)$$

where the current time is assumed to be 0. For zero-coupon bonds such as the US Treasury STRIPS data we use, there is no coupon payment and only the principal is due on the maturity date. Equation (3.1) reduces to

$$\mu_i = L \exp\left\{-\int_0^{T_i} f(s)ds\right\}, \quad (3.2)$$

where L is the principal and T_i is time until maturity for the i -th bond. However, many Treasury bonds pay coupons.

The forward rate curve f is modeled as splines $f(s) = \boldsymbol{\delta}'\mathbf{B}(s)$, where $\mathbf{B}(s)$ is a vector of spline basis functions and $\boldsymbol{\delta}$ is the coefficient vector. We use the d -th degree power basis:

$$f(s) = \delta_0 + \delta_1 s + \cdots + \delta_d s^d + \sum_{k=1}^K \delta_{d+k} (s - t_k)_+^d.$$

Here

$$\boldsymbol{\delta} = [\delta_0, \delta_1, \dots, \delta_{d+K}]^t$$

and

$$\mathbf{B}(s) = \left[1, s, \dots, s^d, (s - t_1)_+^d, \dots, (s - t_K)_+^d \right],$$

where $(s - t_k)_+ = \max(0, s - t_k)$ and $\{t_k\}_{k=1}^K$ are K fixed spline knots. Power basis has the advantage of being simple and allows easy modeling of the credit spread for corporate term structure in the later portion of the paper. Moreover, setting some polynomial coefficients to zero allows convenient modeling of sub-models. However, other basis, such as B-spline basis (de Boor 1978), can also be used.

The model relating the observed market price P_i and the model price μ_i for the i th bond is:

$$P_i = \mu_i + \epsilon_i,$$

where the disturbances $\epsilon_i, i = 1, \dots, n$, are assumed to be normal with zero mean and constant variance. JRY estimate $\boldsymbol{\delta}$ by minimizing a penalized sum-of-squared-errors criterion (1.2), where the roughness penalty measure $\rho(\boldsymbol{\delta}) = \alpha \boldsymbol{\delta}'\mathbf{G}\boldsymbol{\delta}$ and \mathbf{G} is a symmetric, positive semi-definite penalty matrix. If we choose the penalty matrix \mathbf{G} to be a diagonal matrix with its last K diagonal elements equal to one and all others zero as in Ruppert and Carroll (1997), the roughness penalty measure $\boldsymbol{\delta}'\mathbf{D}\boldsymbol{\delta}$ quantifies the closeness of the spline coefficients $\delta_{d+1}, \delta_{d+2}, \dots, \delta_{d+K}$ to zero. The positive smoothing parameter α controls the trade-off between the two terms in the minimization criterion. A large α shrinks the spline coefficients toward zero and the lack of fit term plays an insignificant role in the minimization criterion, yielding large model errors. On the other hand, a small α puts more emphasis on the lack of fit term, making the model errors small. Thus, the smoothing parameter α controls the trade-off between lack of fit and overfitting the sample.

As we have noted in Section 1, the roughness measure can also be treated as the prior information on the spline coefficients within a Bayesian framework. Here we assume the disturbances follow a normal distribution with mean zero and

constant variance σ^2 . We impose normal prior distributions on the polynomial coefficients with $\delta_0, \delta_1, \dots, \delta_d$ distributed as $N(0, \sigma_1^2)$ and the knot coefficients $\delta_{d+1}, \delta_{d+2}, \dots, \delta_{d+K}$ distributed as $N(0, \sigma_2^2)$. If σ_1^2 is infinity, the posterior distribution of the coefficients is proportional to

$$\exp \left\{ \left(-\frac{1}{2} \right) \left[\left(\frac{1}{\sigma^2} \sum_{i=1}^n (P_i - \mu_i)^2 + \left(\frac{1}{\sigma_2^2} \boldsymbol{\delta}' G \boldsymbol{\delta} \right) \right) \right] \right\}. \quad (3.3)$$

If we let the smoothing parameter α be σ^2/σ_2^2 , the mode of this posterior distribution (3.3) minimizes the criterion in (1.2), as pointed out by Ruppert and Carroll (1997) for the linear penalized spline model. Thus, minimizing (1.2) is equivalent to finding the mode of (3.3). The reason for σ_1^2 being set to infinity is that the polynomial coefficients $\delta_0, \delta_1, \dots, \delta_d$ are not penalized. Only the spline coefficients $\delta_{d+1}, \delta_{d+2}, \dots, \delta_{d+K}$ are penalized and they are assumed to be exchangeable. The idea of treating the smoothing parameter as a variance ratio was first proposed by Lindley and Smith (1972) in the scenario of exchangeability within multiple regression equations. Since the variances σ^2 and σ_2^2 in (3.3) are unknown, we develop a hierarchical Bayesian approach to estimate the coefficients.

We consider the following model:

$$\mathbf{P} = \boldsymbol{\mu} + \mathbf{e}$$

where

$$\mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}), \quad \mathbf{P} = [P_1, P_2, \dots, P_n]', \quad \boldsymbol{\mu} [\mu_1, \mu_2, \dots, \mu_n]',$$

and

$$\mu_i = \sum_{j=1}^{z_i} C(t_{ij}) \exp \left\{ - \int_0^{t_{ij}} \boldsymbol{\delta}' \mathbf{B}(s) ds \right\},$$

and n is the total number of Treasury bonds available. By assuming the disturbances to be normal, we are in fact minimizing the sum of the least squares criterion and a roughness penalty term as in (1.2). To simplify the notation, we integrate the forward rate:

$$\int_0^{t_{ij}} f(s) ds = \int_0^{t_{ij}} \boldsymbol{\delta}' \mathbf{B}(s) ds = \boldsymbol{\delta}' \int_0^{t_{ij}} \mathbf{B}(s) ds = \boldsymbol{\delta}' B^I(t_{ij}),$$

where

$$\begin{aligned} \mathbf{B}^I(t_{ij}) &= \int_0^{t_{ij}} \mathbf{B}(s) ds \\ &= \left[t_{ij}, \frac{t_{ij}^2}{2}, \dots, \frac{t_{ij}^{d+1}}{d+1} \frac{(t_{ij} - t_1)_+^{d+1}}{d+1} + \dots + \frac{(t_{ij} - t_K)_+^{d+1}}{d+1} \right]^t. \end{aligned}$$

Then the model price for the coupon bond can be expressed as

$$\mu_i = \sum_{j=1}^{z_i} C(t_{ij}) \exp\{-\boldsymbol{\delta}'\mathbf{B}^I(t_{ij})\}.$$

Note that this model is nonlinear in nature since the spline coefficients appear in the exponential term in the discount function.

We adopt the usual conjugate priors for the coefficient vector $\boldsymbol{\delta}$ and the error variance σ^2 and develop an MCMC sampling scheme to find the parameter estimates. According to the prior specification above, the prior distribution for the coefficient vector $\boldsymbol{\delta}$ is multivariate normal $N(\mathbf{0}, \mathbf{C})$, where

$$\mathbf{C} = \begin{bmatrix} \sigma_1^2 \mathbf{I}_{d+1} & \mathbf{0} \\ \mathbf{0} & \sigma_2^2 \mathbf{I}_K \end{bmatrix}.$$

The error variance σ^2 has an Inverted-Gamma distribution: $IG(a, b)$. We impose a second-stage prior on σ_2^2 : an Inverted-Gamma distribution, $IG(a_2, b_2)$. Here a, b, a_2 and b_2 are constants. Let $[\]$ denote probability densities. The joint density of \mathbf{P} , $\boldsymbol{\delta}$, and σ^2 can be written as

$$\begin{aligned} [\mathbf{P}, \boldsymbol{\delta}, \sigma^2] &= [\mathbf{P} | \boldsymbol{\delta}, \sigma^2][\boldsymbol{\delta}][\sigma^2] \\ &= N_n(\mathbf{P} | \boldsymbol{\mu}, \sigma^2 \mathbf{I}) \times N_{d+K+1}(\boldsymbol{\delta} | \mathbf{0}, \mathbf{C}) \times IG(\sigma^2 | a, b) \end{aligned}$$

The posterior distribution of the coefficient vector $\boldsymbol{\delta}$ is

$$\begin{aligned} [\boldsymbol{\delta} | \mathbf{P}, \sigma^2, \sigma_1^2, \sigma_2^2] &\propto [\mathbf{P} | \boldsymbol{\delta}, \sigma^2, \sigma_1^2, \sigma_2^2][\boldsymbol{\delta}] \\ &\propto N_n(\mathbf{P} | \boldsymbol{\mu}, \sigma^2 \mathbf{I}) \times N_{d+K+1}(\boldsymbol{\delta} | \mathbf{0}, \mathbf{C}) \\ &\propto \exp \left[\left(-\frac{1}{2} \right) \left(\frac{(\mathbf{P} - \boldsymbol{\mu})'(\mathbf{P} - \boldsymbol{\mu})}{\sigma^2} + \boldsymbol{\delta}'\mathbf{C}^{-1}\boldsymbol{\delta} \right) \right], \end{aligned} \quad (3.4)$$

where \propto means proportional to. This posterior distribution is not of any known type and the Metropolis-Hastings algorithm (Hastings 1970) is needed to draw random variates.

The posterior distributions of σ^2 and σ_2^2 are shown below:

$$\begin{aligned} [\sigma^2 | \mathbf{P}, \boldsymbol{\delta}, a, b] &\propto [\mathbf{P} | \boldsymbol{\delta}, \sigma^2][\boldsymbol{\delta}][\sigma^2] \\ &\propto N_n(\mathbf{P} | \boldsymbol{\mu}, \sigma^2 \mathbf{I}) IG(a, b) \\ &\propto (\sigma)^{-(n+2a+2)} \exp \left[\left(-\frac{1}{\sigma^2} \right) \left(\frac{(\mathbf{P} - \boldsymbol{\mu})'(\mathbf{P} - \boldsymbol{\mu})}{2} + \frac{1}{b} \right) \right] \end{aligned}$$

This is an Inverted-Gamma distribution:

$$IG \left(\frac{n}{2} + a, \left(\frac{(\mathbf{P} - \boldsymbol{\mu})'(\mathbf{P} - \boldsymbol{\mu})}{2} + \frac{1}{b} \right)^{-1} \right).$$

Similarly, the posterior distribution for σ_2^2 can be found to be an Inverted-Gamma distribution:

$$IG \left(a_2 + \frac{K}{2}, \left(\frac{\delta_{d+1}^2 + \dots + \delta_{d+K}^2}{2} + \frac{1}{b_2} \right)^{-1} \right).$$

Unlike the posterior distribution for $\boldsymbol{\delta}$ the random variates for these two posterior distributions of σ^2 and σ_2^2 are readily available. If we allow a and a_2 to be small while b and b_2 to be large, the impact of our prior distributions for σ^2 and σ_2^2 on the corresponding posterior distributions will be minimal. Doing so allows us to adopt the usual diffuse (vague) priors. Also, the smoothing parameter, α , in the classical penalized splines, is the ratio of the modes of these two Inverted-Gamma distributions. Thus, the smoothing parameter α is automatically obtained within the Bayesian framework. We describe the MCMC computation algorithm in more detail in the next section.

3.2 Algorithm

The conditional distributions of the variances are Inverted-Gamma and can be drawn directly. To draw the coefficient vector $\boldsymbol{\delta}$, we implement the Metropolis-Hastings algorithm. First, at iteration J we draw $\boldsymbol{\delta}$ from a proposal density, a multivariate normal distribution $N(\boldsymbol{\delta}^{(J-1)}, \Sigma)$ where Σ is tuned so that the acceptance rate of the algorithm is around 10%. Second, compute the ratio $\lambda = [\boldsymbol{\delta}]/[\boldsymbol{\delta}^{(J)}]$ where $[\]$ is the posterior density of $\boldsymbol{\delta}$ without the normalizing constant. The normalizing constant is not needed in the ratio due to cancellation. Third, generate a random number ν between 0 and 1. If $\nu \leq \lambda$, $\boldsymbol{\delta}^{(J)} = \boldsymbol{\delta}$. Otherwise, $\boldsymbol{\delta}^{(J)} = \boldsymbol{\delta}^{(J-1)}$. To facilitate computation, the ratio λ can be shown to be

$$\exp \left(\frac{1}{2} \right) \left[\frac{\boldsymbol{\mu}'\boldsymbol{\mu} - \boldsymbol{\mu}'_*\boldsymbol{\mu}_* + 2\mathbf{P}(\boldsymbol{\mu}_* - \boldsymbol{\mu})}{\sigma^2} + \boldsymbol{\delta}'\mathbf{C}^{-1}\boldsymbol{\delta} - \boldsymbol{\delta}'_*\mathbf{C}^{-1}\boldsymbol{\delta} \right]$$

where $\boldsymbol{\delta}$ is generated from the proposal distribution and $\boldsymbol{\mu}$ is computed with the generated $\boldsymbol{\delta}$ as the input. The simulation algorithm uses both the Metropolis algorithm and Gibbs sampling and is the so-called Metropolis within Gibbs algorithm. A detailed presentation of these sampling schemes can be found in Robert and Casella (1999). The algorithm is summarized below:

Step 1. Choose arbitrary starting values for parameter estimates of $\boldsymbol{\delta}$.

Step 2. Set a and a_2 to be small positive numbers. Set b and b_2 to be large positive numbers.

Step 3. Draw from Inverted Gamma

$$IG\left(\frac{n}{2} + a, \left(\frac{(\mathbf{P} - \boldsymbol{\mu})'(\mathbf{P} - \boldsymbol{\mu})}{2} + \frac{1}{b}\right)^{-1}\right).$$

Step 4. Draw σ_2^2 from Inverted Gamma

$$IG\left(a_2 + \frac{K}{2}, \left(\frac{\delta_{d+1}^2 + \dots + \delta_{d+K}^2}{2} + \frac{1}{b_2}\right)^{-1}\right).$$

Step 5. Draw $\boldsymbol{\delta}$ with the Metropolis-Hastings algorithm outlined above.

We note that when a, a_2 are small and b, b_2 are large, the priors for σ^2 and σ_2^2 have minimal impact on their posterior distributions respectively. The smoothing parameter equals the ratio of the modes of these two posterior distributions. Thus, our priors have little effect on the smoothing parameter and it is chosen automatically by the data.

We apply this Metropolis within Gibbs algorithm to the US Treasury STRIPS data available on December 31, 1995 (see Table 1). The knots carefully chosen by Schwartz (1998) are used here. These knots are located at times to maturity of 1 year, 2 years, 3 years, 4 years, 6 years, 8 years, 10 years, and 18 years. We also adopt the automatic quantile knots used in penalized splines (Ruppert and Carroll 1997). The knots can be equally spaced or placed at uniformly spaced quantiles of maturity time T_i where the knot t_k is the $(k/(K + 1))$ -th sample quantile of the T_i 's. A quadratic spline ($d = 2$) is adequate for the data. It is also easy to use splines with different degrees. For Step 5, a burn-in phase of 100,000 simulation draws was sufficient to achieve convergence and 10,000 draws were saved for computing the estimates. Convergence in Steps 3 and 4 can be achieved much more rapidly.

We report the mean absolute deviation (MAD) and the root mean squared error (RMSE) in Table 3 along with those from JRY and Schwartz (1998), using both the fixed 8 knots in Schwartz (1998) and the 8 automatic quantile knots. Both MAD ($MAD = (1/n) \sum_{i=1}^n |P_i - \mu_i|$) and RMSE ($RMSE = \sqrt{(1/n) \sum_{i=1}^n (P_i - \mu_i)^2}$) measure the goodness-of-fit. The MADs and RMSEs from the Bayesian approach are similar to the ones given by JRY under the 8 fixed knots. They are much smaller than the MAD and RMSE from Schwartz's piecewise constant method.

Table 3: Results for US Treasury STRIPS on December 31, 1995. The piecewise constant fit is as in Schwartz (1998), with $d = 0$ and $\alpha = 0$; the quadratic spline fit is as in JRY, with $d = 2$ and α chosen by GCV; the Bayesian quadratic spline fit uses $d = 2$ and α is the ratio of two posterior variances. Spline knots are either fixed and located at times to maturity of 1 year, 2 years, 3 years, 4 years, 6 years, 8 years, 10 years, and 18 years, as in Schwartz (1998), or are placed at uniformly spaced quantiles of maturing times.

Type of Methods	Schwartz $d = 0$	JRY $d = 2$, GCV	Bayesian, $d = 2$	Normal
Knots	fixed (8)	fixed (8) auto (8)	fixed (8)	
MAD	0.20	0.04 0.04	0.04	
RMSE	0.25	0.06 0.06	0.06	

The method is then applied 21 times independently to the 21 monthly data sets from April 30, 1994 to December 31, 1995. Figure 2 (see Section 4) displays the fitted forward rate from April 30, 1994 to December 31, 1995. The STRIPS data are sparse after 20 years to maturity and only the forward curves up to 20 years to maturity are plotted. We do not recommend using the fitted values after 20 years to maturity. By graphing the 21 fitted curves in one figure, we show both the evolution of the end-of-month forward rate from April 30, 1994 to December 31, 1995 and the rate based on years to maturity from 0 to 20 years. If maturity is fixed, the forward rate is a function of time. The evolution of the term structure displayed in this figure has important applications such as pricing interest rate derivatives (Jarrow 2002).

Remark: The Treasury STRIPS are zero-coupon bonds as stated in (3.2):

$$\mu_i = L \exp\left\{-\int_0^{T_i} f(s)ds\right\}.$$

A monotonic log transformation may be adopted (JRY). An advantage is that $\log(\mu_i)$ of equation (3.2) is now linear in the parameters so that a linear Bayesian penalized spline procedure (Ruppert *et al.*, 2003) can be used. That is, the conditional posterior distribution of the coefficient vector is a multivariate normal, not the indeterminate type in (3.4). Gibbs sampling can be used to easily draw random variates from the three conditional posterior distributions. We implemented this linear model and did not find noticeable differences from the previous nonlinear approach. However, for general coupon bonds, the proposed nonlinear procedure needs to be applied.

4. Bayesian Estimation of the Corporate Term Structure with Informative Priors

We have estimated the term structure of Treasury bonds using Bayesian penalized splines. In this section, we illustrate Bayesian estimation of the corporate term structure with informative prior distributions on the parameters.

Since the number of corporate bonds available is very small, we need to estimate the corporate term structure by borrowing strength from the Treasury term structure. As in JRY, we model the corporate term structure $f_c(s)$ by adding a spread to the estimated Treasury term structure $\hat{f}_{Trea}(s)$ from Section 3:

$$f_c(s) = \hat{f}_{Trea}(s) + \text{spread}.$$

For a constant spread, $f_c(s) = \hat{f}_{Trea}(s) + \beta_0 = \boldsymbol{\delta}'_c \mathbf{B}(s)$ where

$$\boldsymbol{\delta}_c = [\hat{\delta}_0 + \beta_0, \hat{\delta}_1, \hat{\delta}_2, \dots, \hat{\delta}_{d+K}]'.$$

For a linear spread,

$$\boldsymbol{\delta}_c = [\hat{\delta}_0 + \beta_0, \hat{\delta}_1 + \beta_1, \hat{\delta}_2, \dots, \hat{\delta}_{d+K}]'.$$

For a quadratic spread,

$$\boldsymbol{\delta}_c = [\hat{\delta}_0 + \beta_0, \hat{\delta}_1 + \beta_1, \hat{\delta}_2 + \beta_2, \hat{\delta}_3, \dots, \hat{\delta}_{d+K}]'.$$

We first consider estimation of a constant spread. Since the Treasury securities are backed by the taxing power of the US government, corporate debt should be riskier than Treasury securities. As a result, the forward rate for corporate debt should be higher than those for Treasury securities, as a premium is required to compensate for the risk. Consequently, we incorporate the knowledge of positive credit spreads into our Bayesian model as informative priors. We assume that the spread is between 0 and a positive constant prior to our estimation, where the prior for a constant spread is supposed to follow a uniform distribution $(0, g)$:

$$[\beta_0] = \frac{1}{g} I_{(0,g)}(\beta_0) \propto I_{(0,g)}(\beta_0)$$

The posterior distribution for the constant spread is:

$$[\beta_0 | \mathbf{P}, \sigma_2] \propto [\beta_0][\mathbf{P} | \beta_0, \sigma^2] \propto I_{(0,g)}(\beta_0) \exp\left(-\frac{\sum_{i=1}^{n_c} (P_i - \mu_i)^2}{2\sigma^2}\right),$$

where n_c is the number of corporate bonds, P_i is the observed price of the i -th corporate bond, and $\mu_i = \sum_{j=1}^{z_i} C_i(t_{ij}) \exp\{-\int_0^{t_{ij}} f_c(s) ds\}$ (each bond pays fixed

coupons and principal $C_i(t_{ij})$ due on dates t_{ij} , where $j = 1, \dots, z_i$ and z_i is the total number of coupon and principal payments for the i -th bond). This is not of any known distribution and we again apply the Metropolis algorithm. The posterior distribution of the error variance is an Inverted-Gamma distribution:

$$[\sigma^2 | \mathbf{P}, \beta_0, a, b] \sim IG \left(\frac{nc}{2} + a, \left(\frac{\sum_{i=1}^{nc} (P_i - \mu_i)^2}{2} + \frac{1}{b} \right)^{-1} \right).$$

To estimate the linear spread, we adopt a flat prior and add a constraint: $0 < \beta_0 + \beta_1 s < g$. The prior distribution can be written as: $[\beta_0, \beta_1] \propto I_{(0 < \beta_0 + \beta_1 s < g)}(\beta_0, \beta_1)$. For the quadratic spread, the constraint becomes: $0 < \beta_0 + \beta_1 s + \beta_2 s^2 < g$. The prior becomes: $[\beta_0, \beta_1, \beta_2] \propto I_{(0 < \beta_0 + \beta_1 s + \beta_2 s^2 < g)}(\beta_0, \beta_1, \beta_2)$. One needs to choose g to complete the prior specification. This prior information can come from one's subjective knowledge. For example, a trader who has previously traded AT&T bonds before has some idea regarding the difference between US STRIPS and AT&T bonds. The prior information can also come from previous empirical studies on the AT&T bonds credit spread. We suppose g to be 0.2 as we believe it is reasonable to assume that the credit spread is between 0 and 0.2 prior to estimation.

We note that the priors we have adopted are equivalent to constraints in a nonlinear optimization problem with a frequentist approach. The average sample size of the individual corporate bonds on a given day is only 4 to 5. Within the frequentist framework, large-sample theories cannot be applied. JRY use a parametric bootstrap procedure for inference. With the Bayesian approach, statistical inference does not rely on the assumption of large samples and can be easily performed. The Bayesian approach allows one to obtain the posterior interval, a $100(1 - \alpha)\%$ central interval, from MCMC sample draws. The estimation results are summarized in Table 4.

We notice that under the quadratic spread, the probability that the constant spread lies in the interval (0.0045, 0.0075) given the observed bond prices is 0.95 while the confidence intervals for the linear spread and the quadratic spread include 0. Under the linear spread case, the confidence interval for the linear spread also contains 0. The confidence intervals for the constant spread coefficient clearly do not contain 0. There is a 0.95 probability that the constant spread is between 0.0038 and 0.0044. Hence, the simple model of a constant spread is supported by the data. The results are not surprising. A constant spread (one parameter) is sufficient to model a sample of 5 bonds. However, for other individual corporate bonds with a larger sample size, including both the linear and the quadratic spreads may be needed. Figure 1 provides the fitted forward rate curves for both the US STRIPS and the AT&T bonds on December 31, 1995 for comparison.

Table 4: Estimation results of AT&T bonds on December 31, 1995. Three types of spreads are considered: constant spreads, linear spreads, and quadratic spreads. One parameter needs to be estimated for the constant spread case. Two parameters, the intercept and the linear coefficient, need to be estimated for the linear spread. Three parameters, the intercept, the linear coefficient, and the quadratic coefficient, need to be estimated for the quadratic spread. The root mean squared error, parameter estimates (posterior means), and their 95% posterior intervals are reported for the three types of spreads.

	Constant spread	Linear spread	Quadratic spread
RMSE	0.1728	0.11	0.1176
Parameter estimates and 95% CI:	0.004 (.0038, .0044)	0.0051 (.0034, .0063)	0.0063 (.0045, .0075)
Constant spread			
Parameter estimates and 95% CI:		-.00025 (-.00051, .0015)	-.00083 (-.0013, .000038)
Linear spread			
Parameter estimates and 95% CI:			.00005 (-.000024, .000086)
Quadratic spread			

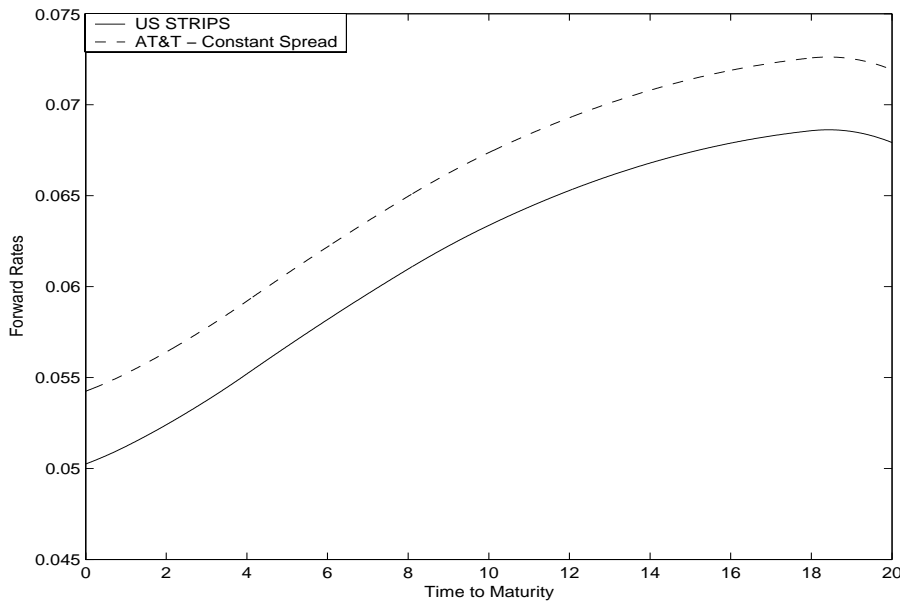


Figure 1: Fitted forward rate curves by Bayesian quadratic penalized splines on US STRIPS and AT&T bonds on December 31, 1995.

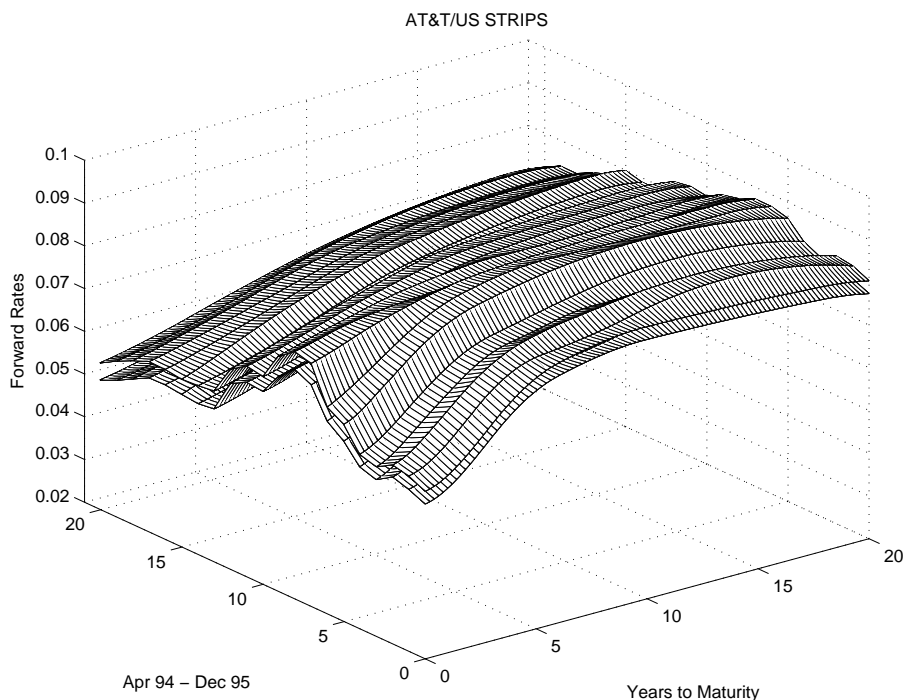


Figure 2: The fitted forward rate curves for AT&T (upper sheet) and US STRIPS (lower sheet) over the 21-month period of April 1994 to December 1995. They are obtained by adding a constant spread to the estimated Treasury term structures.

Next we estimate independently the term structures of AT&T bonds over the 21-month period of April 1994 to December 1995 by adding a constant spread to the fitted forward rate for US STRIPS over the same 21 months from Section 3.3. The fitted forward rates are plotted in Figure 2. The interpretation and application of these 21 curves are described at the end of Section 3.2.

5. Conclusions

We have proposed a Bayesian approach to estimating Treasury and corporate term structures with a penalized spline model. We first estimate the Treasury term structure. Due to the small sample size of corporate bonds, credit spreads are added to the estimated Treasury term structure to obtain the corporate term structure. The fact that the positive credit spreads between the default-free Treasury securities and the risky corporate bonds is too important to ignore. The Bayesian approach naturally takes this into account as a prior distribution and also gives valid statistical inference.

References

- Adams, K. J. and Van Deventer, D. R. (1994). Fitting yield curves and forward rate curves with maximum smoothness. *Journal of Fixed Income* **4**, 52-56.
- Anderson, N. and Sleath, J. (1999). New estimates of the UK real and nominal yield curves. *Bank of England Quarterly Bulletin*, November, 384-392.
- Chambers, D. R., Carleton, W. T., and Waldman, D. W. (1984). A new approach to estimation of the term structure of interest rates. *Journal of Financial and Quantitative Analysis* **19**, 233-252.
- de Boor, C. (1978). *A Practical Guide to Splines*. Springer Verlag.
- Duffie, D. and Singleton, K. (1999). Modeling term structures of defaultable bonds. *Review of Financial Studies* **12**, 197-226.
- Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* **11**, 89-121.
- Fisher, M., Nychka, D., and Zervos, D. (1995). Fitting the term structure of interest rates with smoothing splines. FEDS 95-1, Federal Reserve Board, Washington DC. (available at www.markfisher.net/~mefisher/papers/downloadable_papers.html or request at Mark.Fisher@atl.frb.org).
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97-109.
- Jarrow, R. (2001). Default parameter estimation using market prices. *Financial Analysts Journal* **57**, 75-92.
- Jarrow, R. (2002). *Modeling Fixed-Income Securities and Interest Rate Options*. Stanford University Press.
- Jarrow, R., Ruppert, D., and Yu, Y. (2004). Estimating the term structure of corporate debt with a semiparametric penalized spline model. *Journal of the American Statistical Association* **99**, 57-66.
- Jarrow, R. and Turnbull, S. (1995). Pricing derivatives on financial securities subject to credit risk. *Journal of Finance* **50**, 53-85.
- Lindley, D. V., and Smith, A. F. M. (1972). Bayes estimates for the linear model (with discussion). *Journal of the Royal Statistical Society, Series B* **34**, 1-41.
- Linton, O., Mammen, E., Nielsen, J. P., and Tanggaard, C. (2001). Yield curve estimation by kernel smoothing methods. *Journal of Econometrics* **105**, 185-223.
- McCulloch, J. H. (1971). Measuring the term structure of interest rates. *Journal of Business* **44**, 19-31.
- McCulloch, J. H. (1975). The tax-adjusted yield curve. *Journal of Finance* **30**, 811-830.
- Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer-Verlag.

- Ruppert, D. and Carroll, R. (1997). Penalized regression splines. Working paper, School of Operations Research and Industrial Engineering, Cornell University (available at www.orie.cornell.edu/~davidr/papers).
- Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press.
- Schwartz, T. (1998). Estimating the term structure of corporate debt. *Review of Derivatives Research* **2**,193-230.
- Shea, G. (1984). Pitfalls in smoothing interest rate term structure data: Equilibrium models and spline approximations. *Journal of Financial and Quantitative Studies* **19**, 253-269.
- Shea, G. (1985). Interest rate term structure estimation with exponential splines: A note. *Journal of Finance* **40**, 319-325.
- Tanggaard, C. (1997). Nonparametric smoothing of yield curves. *Review of Quantitative Finance and Accounting* **9**, 251-267.
- Vasicek, O. A. and Fong, H. G., (1982). Term structure modeling using exponential splines. *Journal of Finance* **37**, 339-348.
- Waggoner, D. (1997). Spline methods for extracting interest rate curves from coupon bond prices, Working Paper, No 97-10, Federal Reserve Bank of Atlanta.
- Warga, A. (1995). Fixed income securities database. Fixed Income Research Program, University of Houston. (Request information at warga@uh.edu).

Received March 2, 2004; accepted May 4, 2004.

Min Li
Management Information Science Department
California State University, Sacramento
6000 J St
Sacramento, CA 95819-6088, USA
limin@csus.edu

Yan Yu
Department of Quantitative Analysis and Operations Management
PO Box 210130
Cincinnati, OH 45221-0130, USA
Yan.Yu@uc.edu