Estimating the Interest Rate Term Structure of Corporate Debt With a Semiparametric Penalized Spline Model

Article *in* Journal of the American Statistical Association · March 2004

DOI: 10.1198/016214504000000070 · Source: RePEc

Robert Alan Jarrow
Cornell University

David Ruppert
Cornell University

Estimating the Interest Rate Term Structure of Corporate Debt with a Semiparametric Penalized Spline Model

Robert Jarrow, David Ruppert, and Yan Yu *

Abstract

This paper provides a new methodology for estimating the term structure of corporate debt using a semiparametric penalized spline model. The method is applied to a case study of AT&T bonds. Typically, very little data is available on individual corporate bond prices, too little to find a nonparametric estimate of term structure from these bonds alone. This problem is solved by "borrowing strength" from Treasury bond data. More specifically, we combine a nonparametric model for the term structure of Treasury bonds with a parametric component for the credit spread. Our methodology generalizes the work of Fisher, Nychka, and Zervos (1995) in several ways. First, their model was developed for only Treasury bonds and cannot be applied directly to corporate bonds. Second, we more fully investigate the problem of choosing the smoothing parameter, a problem that is complicated because the forward rate is the derivative $-\log\{D(t)\}$, where the discount function D is the function fit to the data. In our case study estimation of the derivative requires substantially more smoothing than selected by generalized cross-validation (GCV). Another problem for smoothing parameter selection is possible correlations of the errors. We compare three methods of choosing the penalty parameter: linearized GCV, the residual spatial autocorrelation (RSA) method of Ellner and Seifu (2002), and a modification of Ruppert's (1997) EBBS. Third, we provide approximate sampling distributions based on both large-sample and small-noise asymptotics. The latter are novel and are motivated by the application to corporate bond prices where the sample sizes are small but the noise is very low. Confidence bands and tests of interesting hypotheses, e.g., about the functional form of the spreads, are also discussed.

Key Words: Credit Spreads; EBBS; Forward Rate; GCV; Roughness Penalty; Small-sample asymptotics, Treasury Bonds.

*Robert Jarrow is R. P. and S. E. Lynch Professor of Investment Management, Johnson Graduate School of Business, Cornell University, Ithaca, NY, 14853; David Ruppert is Andrew Schultz, Jr., Professor of Engineering, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY, 14853, email: davidr@orie.cornell.edu; and Yan Yu is Assistant Professor of Quantitative Analysis and Operations Management at the University of Cincinnati, PO BOX 210130, Cincinnati, OH, 45221, email: yuy@statqa.cba.uc.edu.

1 Introduction

This paper contains a case study in statistical finance as well as several methodological innovations that also should be of interest in other areas of statistics. In particular, we suggest a new method of choosing the smoothing parameter when estimating a derivative of a regression function using a spline. (In this paper we discuss mathematical derivatives as well as financial derivatives. The meaning should be clear from the context.)

The prices of bonds determine an implied interest rate. Consider a *zero-coupon bond* paying no interest or principal until maturity, then paying a fixed amount called the par. Suppose that $P(t)$ is the current price, as a fraction of par, of a zero-coupon bond maturing in t years. This price is consistent with a fixed continuous-compounding "forward" interest rate f such that

$$P(t) = \exp(-ft). \qquad (1)$$

Figure 1(a) shows typical price data for a zero coupon bond. There are maturities from 0 to 30 years, spaced nearly quarterly. The rough exponential decay may appear consistent with (1), but here, and in general, there is no single fixed rate f that prices correctly for all t. To appreciate this, one can look at Figure 1(b). The "empirical forward" rate in that figure is $\Delta\{-\log(P)\}/\Delta t$, where $\Delta\{-\log(P)\}$ is the differenced series of negative log prices and Δt is the differenced series of maturities. The EBBS and GCV estimates of the forward rate in that figure are explained below. The key point is that the difference quotients exhibit both random variation and systematic deviation from a constant rate. However, bond prices are also consistent with a variable interest rate $f(t)$ such that

$$P(t) = \exp\left\{-\int_0^t f(s)ds\right\}. \qquad (2)$$

One can find a single smooth rate function $f(t)$, called the *forward* rate, for which (2) holds for all t, except possibly for some small errors. The financial significance of $f(t)$ is that it is the rate one can lock in today for future borrowing or lending at time t. The errors can be attributed to staleness of the price data; the price of a bond of a given maturity was determined at the time of the last trade, so the prices are not exactly concurrent.

The dependence of $f(t)$ on t is called the *term structure*. The term structure can only be inferred from observable bond prices. Although the literature studying the estimation of Treasury term structure is voluminous (see McCulloch 1971, Vasicek and Fong 1982, Shea 1984, Chambers, Carleton and Waldman 1984, Adams and Van Deventer 1994, and Fisher, Nychka and Zervos 1995), the literature studying corporate term structure estimation is almost non-existent (see Schwartz

1998 and references therein). The problem is that for any individual corporation, there are bond prices at only a few maturities so determination of $f(t)$ for all t is challenging. This appears to be the first paper to estimate the term structure for individual corporate bonds.

There are many reasons why estimation of f is of interest. Suppose one were offering to buy or to sell a bond of a maturity not traded recently. Estimation of f allows one to interpolate prices from other maturities and determine such a "fair" price.

There are other more complex and interesting applications of the term structure. Corporate bonds are a classical example of an instrument bearing credit risk, the risk that an agent fails to fulfill contractual obligations. Increased trading in instruments subject to credit risk has led to the creation of credit derivatives, instruments that partially or fully offset the credit risk of a deal. Given the recent explosive growth in the market for credit derivatives (see Risk Magazine, 2002) and the regulatory-induced need to account for credit risk in the determination of equity capital (net worth of a business raised from owners), e.g., Jarrow and Turnbull (2000), the estimation of corporate term structures has become of paramount interest. To put this in perspective, the size of the credit derivatives market in 2001 (as measured in notional amounts outstanding) was estimated to be 835.5 billion dollars.

The most traded credit derivatives include default swaps, credit spread options, credit linked notes, and collateralized default obligations (CDOs). For example, a credit call (put) option gives its owner the right to buy (sell) a credit-risky asset at a predetermined price, regardless of credit events which may occur before expiration of the option. A full treatment of credit derivatives can be found in Bielecki and Rutkowski (2002). The primary inputs to pricing models for these credit derivatives are the corporate term structures (see Jarrow and Turnbull 1995, Duffie and Singleton 1999, Bielecki and Rutkowski 2002). These term structures can also be used to infer the market's assessment of credit quality for related uses in risk management procedures (see Jarrow 2001). Credit quality assessment is essential for value at risk (VaR) (Dowd, 1998; Jorion, 2000) computations, bond portfolio management, corporate loan considerations, and even FDIC insurance premium calculations (see FDIC 2000).

In the estimation of the Treasury term structure hundreds of bond prices are normally available on any given month, but for corporate term structures only a handful usually exist. (This problem is observed in the Fixed Income data base, Warga 1995.) Consequently, corporate bonds require special estimation procedures.

Fisher, Nychka and Zervos's (1995) (F-N-Z) penalized spline model is non-parametric and as such it requires numerous bond price observations. The F-N-Z model applies to Treasury bonds where prices at many maturities are available on any date, but it is problematic when applied directly to corporate debt. We generalize the F-N-Z model to corporate debt by modeling the corporate term structure as a Treasury term structure plus a parametric credit spread. The credit spread is the extra interest investors demand to buy risky and less liquid corporate bonds instead of Treasury bonds. For the Treasury term structure, we use F-N-Z's non-parametric model. We find that a credit spread that is constant in time, thus requiring only a single parameter, fits our data well. In other situations, a spread that is linear in time might be used.

We extend F-N-Z's work by: (i) providing a comparison of F-N-Z's linearized generalized cross validation (GCV), Ruppert's (1997) EBBS method, and Ellner and Seifu's (2002) residual spatial autocorrelation (RSA) method for choosing penalty parameters, (ii) deriving asymptotic sampling distributions for the term structure estimates which, enable us (iii) to compute confidence bands for the term structure estimates. We also introduce "low noise" asymptotics that justify linearizing a nonlinear regression model even when the sample size is small; this theory is motivated by the application to corporate bonds where the sample size might be as small as 5 but the noise is quite low.

The term structure of interest rates can be identified by any of the discount function, yield curve, forward rate curve, or the definite integral of the forward rate, each of which determines the others. The forward curve has already been discussed. The discount function, $D(t)$, gives the price of a zero coupon bond that pays one dollar at maturity time t, so that $D(t) = P(t)$ is given by (2). The yield curve, $y(t)$, is the average of $f(s)$ between 0 and t: $y(t) = t^{-1}\int_0^t f(s)ds$. The definite integral of f is $F(t) = ty(t)$. The relationships among these functions are:

$$P(t) = D(t) = \exp\{-F(t)\} = \exp\{-ty(t)\} = \exp\left\{-\int_0^t f(s)ds\right\}. \tag{3}$$

Each of f, D, F, and y depends on two time parameters, the current time and maturity, but current time is considered fixed at 0 and not included in the notation.

Equation (3) holds only for zero-coupon bonds, but many bonds including the AT&T bonds in our case study have coupons. To price a coupon bond, we can view that coupon as a bundle of zero-coupon bonds, one for each payment. Payments can be priced by (3) and then summed.

Let $P_1, \cdots, P_n$ denote the current (time 0) observed market prices of n bonds from which the interest rate term structure is to be inferred. Bond i, $i = 1, \cdots, n$, has z_i fixed payments $C_i(t_{i,j})$ due on dates $t_{i,j}, j = 1, ..., z_i$. The payment, $C_i(t_{i,j})$, consists of coupons and principal at maturity

t_{i,z_i}. The model price for the ith coupon bond is

$$\widehat{P_i} = \sum_{j=1}^{z_i} C_i(t_{i,j}) D(t_{i,j}) = \sum_{j=1}^{z_i} C_i(t_{i,j}) \exp\{-t_{i,j} y(t_{i,j})\} = \sum_{j=1}^{z_i} C_i(t_{i,j}) \exp\left\{-\int_0^{t_{i,j}} f(s) ds\right\}. \quad (4)$$

Should one use a spline model for the forward rate f or for some other function such as $D(t)$? F-N-Z consider spline modeling of f, F, and D and conclude that modeling f results in the most accurate estimation. If D is modeled as a spline, then the model is linear in the spline coefficients, which is obviously attractive. However, there are advantages to modeling f itself as a spline. The constraint that a dollar paid today is worth a dollar, i.e., that $D(0) = 1$, is then embedded in this model. In contrast, when fitting splines to D or F, the constraint $D(0) = 1$ or $F(0) = 0$ must be imposed. Also, Shea (1984) noticed serious problems fitting splines to D, such as negative forward rates and instability at the long maturities.

There are three basic approaches to spline estimation: smoothing splines, regression splines, and penalized splines (P-splines). Smoothing splines (e.g., Wahba, 1990; Eubank, 1999) are defined for linear estimation problems. They require a knot at every distinct value of the independent variable. A roughness penalty prevents overfitting, and because of fast algorithms the large number of knots is not a difficulty for linear estimation problems. Exactly how smoothing splines generalize to nonlinear estimation is not clear, but certainly placing a knot at every observation would cause difficulties. Regression splines use a small number of knots placed judiciously, and the spline coefficients are estimated by ordinary (unpenalized) least squares. Shea (1984) found that unrealistic term structure shapes could be caused by the subjective specification of regression spline knots. Perhaps adaptive data-driven knot placement algorithms (Friedman and Silverman, 1989; Friedman, 1991) could ameliorate this difficulty with regression splines, but we are unaware of research on adaptive knot placement for estimation of term structure. "Adaptive" means that the knots locations depend on the response and are chosed to obtain a good but parsimonious fit. In contrast, "automatic" knot placement used in this paper, e.g., placing the knots at quantiles of the independent variable, is done independently of the response values.

P-splines, the approach taken in this paper, bypass the problem of knot placement. P-splines require the user to choose only the number of knots, not their locations. Once the number of knots is selected, the knots are located at equally-spaced points as in Eilers and Marx (1996) or, as in Ruppert and Carroll (2000) and in Section 7, at equally-spaced quantiles of the independent variable. A relatively large number, K, of knots is used, but still far less than for a smoothing spline, e.g., a P-spline may use $K = 20$ for $n = 200$. This makes P-splines more suitable for nonlinear problems such as estimating f. Because the roughness penalty prevents overfitting, the value of K is not crucial, provided that more than a minimum value is used; see Ruppert (2002).

P-splines, like smoothing splines, minimize the sum of a goodness-of-fit statistic plus a roughness penalty. We model the spline as $f(s) = \boldsymbol{\delta}'\mathbf{B}(s)$, where $\mathbf{B}$ is a vector of spline basis functions and $\boldsymbol{\delta}$ is a vector of spline coefficients. The roughness penalty is $\lambda\boldsymbol{\delta}'\mathbf{G}\boldsymbol{\delta}$ where $\lambda > 0$ is a smoothing parameter and $\mathbf{G}$ is a symmetric, positive semidefinite matrix. Possible choices of $\mathbf{G}$ are discussed in Section 3.

Proper selection of λ to control the trade off between goodness-of-fit and smoothness is crucial but complicated by three difficulties. The first, that GCV uses the trace of the smoother matrix defined only for linear smoothers, is solved by F-N-Z's approximation based upon a Taylor linearization.

A second problem is that the choice of the smoothing parameter depends on the function estimated. We are estimating $f(t) = (d/dt)[-\log\{D(t)\}]$, but since least-squares compares $D(t)$ to prices, GCV will choose the λ best for estimating D, not f. We address this problem by a modification of the EBBS method of Ruppert (1997). EBBS minimizes an estimate of the mean square error of f averaged across observed maturities.

A third problem is that GCV and related methods such as cross-validation (CV) assume independent errors. This assumption is suspect in our case. Because of possible correlation, an alternative method of smoothing parameter selection due (Ellner and Seifu; 2002) based on RSA is considered. However, in the case study we find that RSA and GCV seriously undersmooth while EBBS works better.

We present a case study of US Treasury STRIPS and AT&T bonds over the 21 month period from April 1994 to December 1995. A Treasury STRIPS (Separate Trading of Registered Interest and Principal of Securities) is a synthetic zero-coupon bond constructed from Treasury bonds and issued by the Federal Reserve. The AT&T bonds bear coupons. The data are from the University of Houston Fixed Income data base (Warga 1995). One could estimate corporate term structure using either a one-step and two-step procedures. The one-step method simultaneously estimates the Treasury term structure and the spread for a single corporation. In the two-step procedure, first one estimates the non-parametric Treasury term structure and then, with that term structure fixed, estimates the parametric credit spread. The two-step procedure is motivated by the application at hand. Although only one Treasury term structure exists, there are thousands of different corporate term structures, one for each company issuing debt. It makes sense to estimate the Treasury term structure only once, so we recommend the two-step procedure.

The remainder of the paper is organized as follows. Section 2 describes the fixed income data base. Section 3 introduces P-splines and presents a spline model for Treasury bonds. Section 4 discusses the GCV, RSA, and EBBS criteria for selecting the penalty parameter. Section 5 describes the two-step estimation procedure that we recommend. Asymptotics, confidence bands and tests about the spread model are presented in Section 6. The case study is presented in Section 7. Section 8 discusses some alternatives and potential implementation.

2 Data

The University of Houston Fixed Income data base includes over 28,000 instruments and covers virtually every firm that has outstanding publicly traded non-convertible debt with principal value of at least one million dollars. Information on individual bonds that make up the Lehman Brothers Bond Indices are reported including 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 data for our case study consists of all US Treasury STRIPS and all AT&T bonds. Market prices are available for five AT&T bonds on December 31, 1995. All have semi-annual coupons with different maturities and with no option embedded features, e.g., the right to prepay, for which our price model does not apply. Each price is obtained from the quoted flat price plus the corresponding accrued interest.

Issue and maturity are given in year-month-day format. We need the time-to-maturity and the coupon payment times, $t_{i,j}$, on the same scale. The MATLAB finance toolbox can easily handle date conversions using, for instance, the functions `days365(·)` and `days360(·)`, for dates based on 365 or 360 days a year; 30-day months or 360 days per year is a convention used for some types of bonds, but not those in our case study. The coupon payment time can then be calculated by the function `cfdates(·)`. These calculations can also be easily implemented if the day counts need to exclude holidays and weekends. We use MATLAB functions `days365(·)` and `cfdates(·)` based on conventional actual/365 day count.

Table 2 lists the summary statistics for the numbers of US Treasury STRIPS and AT&T bonds available over the 21 month period of April 1994 to December 1995 and demonstrates that far fewer AT&T bonds are available than US Treasury STRIPS.

Table 1: AT&T Bonds on December 31, 1995. Dates and first coupon payment time $t_{i,1}$ are converted to numbers in units of one year using MATLAB functions `days365(·)` and `cpndaten(·)` based on actual/365 day count. The current date is set to time 0.

Date(yr)	Issue(yr)	Maturity(yr)	First Coupon(yr)	Coupon	Price
0	−3.9616	6.0411	0.0411	7.1250	109.4580
0	−1.7726	8.2493	0.2493	6.7500	106.2840
0	−1.5836	10.4164	0.4164	7.5000	111.4360
0	−0.8384	11.1644	0.1644	7.7500	115.5090
0	−0.6384	9.3699	0.3699	7.0000	107.6590

Table 2: Summary statistics of number of bonds available per month for period of April 1994 – December 1995.

Bond Number	Average	Min	Quantile(25%)	Quantile(75%)	Max
US Treasury STRIPS	117.7	115	116.75	119	120
AT&T	4.3	3	4	5	5

3 A Spline Model for the Term Structure of Treasury Bonds

The Treasury forward rate curve, denoted f_{Tr}, will be approximated by a spline $f_{Tr}(t) = \boldsymbol{\delta}'\mathbf{B}(t)$. Here $\mathbf{B}(t)$ is a vector of spline basis functions, e.g., (truncated) power basis functions or B-splines, and $\boldsymbol{\delta}$ is the coefficient vector. We will use the p-th degree power basis functions, with $\mathbf{B}(t) = \left(1 \; t \; \cdots \; t^p \; (t-\kappa_1)_+^p \; \cdots \; (t-\kappa_K)_+^p \right)'$ and $\boldsymbol{\delta} = (\delta_0, \delta_1, \ldots, \delta_{p+K})'$, where $\left\{\kappa_k\right\}_{k=1}^{K}$ are spline knots and $(t-\kappa_k)_+^p = (t-\kappa_k)^p$ if $s \geq \kappa_k$ and is 0 otherwise. The power basis is convenient because polynomial submodels can be defined by setting certain coefficients to 0. This basis known to be poorly conditioned, but the use of a penalty improves conditioning considerably. Moreover, if conditioning becomes a problem, one can compute in a different basis and then transform the results back to the power basis; see Ruppert (2002).

Then $f_{Tr}(t)$ is estimated by minimizing

$$Q_{n,\lambda}(\boldsymbol{\delta}) = \frac{1}{n}\sum_{i=1}^{n}\{P_i - \widehat{P}_i\}^2 + \lambda\boldsymbol{\delta}'\mathbf{G}\boldsymbol{\delta}. \tag{5}$$

There are several sensible choices for $\mathbf{G}$. One choice is given by Ruppert and Carroll (2002), where the power basis functions are used and $\mathbf{G}$ is a diagonal matrix with its last K diagonal elements equal to one and all others zero. This $\mathbf{G}$ penalizes jumps at the knots in the pth derivative of the spline. As $\lambda \to \infty$ the fit converges to a pth degree polynomial fit. This penalty can be viewed as a penalty on the $p+1$th derivative where that derivative is a generalized function. Schwartz

(1998) uses piecewise constant ($p = 0$) splines with no penalty ($\lambda = 0$). In our numerical work in Section 7.1 we will use this penalty with p equal to 2.

A second choice, the quadratic penalty on the dth derivative, $\int \{f^{(d)}(s)\}^2\, ds$, for $d \leq p$, uses $G_{ij} = \int B_j^{(d)}(s) B_k^{(d)}(s) ds$, where $B_j(t)$ is the jth element of $\mathbf{B}(t)$. P-splines with this choice of $\mathbf{G}$ were proposed by O'Sullivan (1986). Estimates using this penalty with $d = p = 2$ are similar to our numerical results in Section 7.1. The choice $d = 2$, the usual choice for smoothing splines, penalizes any deviation from linearity. In their maximum smoothness approach, Adams and Van Deventer (1994) use $d = 2$. Frishling and Yamamura (1997) use $d = 1$, which penalizes deviations from a constant function. If a quadratic integral penalty is used, then as $\lambda \to \infty$ the estimated forward rate converges to the $d - 1$th degree polynomial fit. If $\lambda = 0$ then the fit is a non-penalized regression spline.

We can rewrite expression (5) in a more revealing form. Because splines are piecewise polynomials, it is easy to compute their integrals. In our model, $\int_0^t f(s)ds = \boldsymbol{\delta}' \int_0^t \mathbf{B}(s)ds$. Denote

$$\mathbf{B}^I(t) := \int_0^t \mathbf{B}(s)ds = \begin{pmatrix} t & \cdots & \frac{t^{p+1}}{p+1} & \frac{(t-\kappa_1)_+^{p+1}}{p+1} & \cdots & \frac{(t-\kappa_K)_+^{p+1}}{p+1} \end{pmatrix}'.$$

We can simplify notation by expressing the model price in terms of $\mathbf{B}^I$ so that (5) becomes

$$Q_{n,\lambda}(\boldsymbol{\delta}) = \frac{1}{n} \sum_{i=1}^{n} \Big\{ P_i - \sum_{j=1}^{z_i} C_i(t_{i,j}) \exp\{-\boldsymbol{\delta}' \mathbf{B}^I(t_{i,j})\} \Big\}^2 + \lambda \boldsymbol{\delta}' \mathbf{G} \boldsymbol{\delta}. \tag{6}$$

4 Selection of the knots and Smoothing Parameter

4.1 Choosing the knots

An advantages of our P-spline approach is that the knots can be chosen automatically; following Ruppert and Carroll (2000) and Ruppert (2002) the knot κ_k is the $\frac{k}{(K+1)}$th sample quantile of the t_{i,z_i}'s. Ruppert (2002) has a detailed study of the choice of K. We recommend that K be sufficiently large, say 8 or more, to accommodate nonlinearity of f_{Tr}, but a larger K does not cause overfitting provided λ is suitably chosen.

4.2 Linearized Generalized Cross Validation

A smoother is linear if $\widehat{\mathbf{P}}$ and $\mathbf{P}$ are related by $\widehat{\mathbf{P}} = \mathbf{A}(\lambda)\mathbf{P}$ for some "smoother" matrix $\mathbf{A}(\lambda)$ independent of $\mathbf{P}$. GCV is an approximation to cross-validation (CV) where λ is chosen by minimizing

$$GCV(\lambda) = \frac{n^{-1} \sum_{i=1}^{n} \Big\{ P_i - \widehat{P}_i \Big\}^2}{\{1 - n^{-1} \mathrm{tr} \mathbf{A}(\lambda)\}^2}. \tag{7}$$

GCV applies to linear smoothers, but following F-N-Z we can Taylor expand the model about $\widehat{\boldsymbol{\delta}}$ to linearize. Let the model price of the ith bond be $m_i(\boldsymbol{\delta}) := \widehat{P}_i = \sum_{j=1}^{z_i} C_i(t_{i,j}) \exp\{-\boldsymbol{\delta}'\mathbf{B}^I(t_{i,j})\}$. Also $\mathbf{m}^{(1)}(\widehat{\boldsymbol{\delta}}) := \left(m_1^{(1)}(\widehat{\boldsymbol{\delta}}(\lambda)) \quad m_2^{(1)}(\widehat{\boldsymbol{\delta}}(\lambda)) \quad \cdots \quad m_n^{(1)}(\widehat{\boldsymbol{\delta}}(\lambda)) \right)'$, where $m_i^{(1)}(\widehat{\boldsymbol{\delta}}(\lambda)) = \frac{\partial m_i(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}'}\Big|_{\widehat{\boldsymbol{\delta}}(\lambda)} = -\sum_{j=1}^{z_i} C_i(t_{i,j}) \exp\{-\widehat{\boldsymbol{\delta}}'\mathbf{B}_i^I(t_{i,j})\}\mathbf{B}_i^I(t_{i,j})$. Define

$$\boldsymbol{\Sigma}_n = n^{-1}\left[\{\mathbf{m}^{(1)}(\widehat{\boldsymbol{\delta}})\}'\mathbf{m}^{(1)}(\widehat{\boldsymbol{\delta}})\right]. \tag{8}$$

The approximate linearized smoothing matrix operator is then $\mathbf{A}(\lambda) = \mathbf{m}^{(1)}(\widehat{\boldsymbol{\delta}})\ \{n(\boldsymbol{\Sigma}_n + \lambda\mathbf{G})\}^{-1}$ $\{\mathbf{m}^{(1)}(\widehat{\boldsymbol{\delta}})\}'$. We use the trace of $\mathbf{A}(\lambda)$ as the effective degrees of freedom (Hastie and Tibshrani 1990) and then λ minimizes GCV in equation (7). The computation of $\mathrm{DF}(\lambda) = \mathrm{tr}\{\mathbf{A}(\lambda)\}$ is rapid; one only needs to compute a $(1+p+K)\times(1+p+K)$ matrix because $\mathrm{tr}\{\mathbf{A}(\lambda)\} = \mathrm{tr}\left\{(\boldsymbol{\Sigma}_n + \lambda\mathbf{G})^{-1}\boldsymbol{\Sigma}_n\right\}$.

In our case study, the performance of GCV was not entirely satisfactory, which led us to search for alternatives, first RSA and then EBBS.

4.3 Residual Spatial Autocorrelation

The RSA method of Ellner and Seifu (2002) applies Moran's index of spatial autocorrelation I to the residuals to chooses the λ giving the least deviation of I from its expectation under a random permutations. We refer the reader to Ellner and Seifu (2002) for details.

4.4 EBBS

EBBS (Empirical Bias Bandwidth Selection) developed by Ruppert (1997) for choosing the bandwidth for local regression can be extended to other smoothing parameters. EBBS models the bias as a function of the smoothing parameter. The variance of the estimated function can be estimated by an asymptotic formula; see (10). To estimate MSE, the estimated bias is squared and added to the estimated variance. When applied to the f_{Tr}, we have $\mathrm{MSE}(\widehat{f}_{Tr}; t, \lambda)$, the estimated MSE of $\widehat{f}_{Tr}$ at t and λ. $\mathrm{MSE}(\widehat{f}_{Tr}; t, \lambda)$ can be averaged over maturities t_{i,z_i}, $i = 1, \ldots, n$ and then minimized over λ.

The key issue is how to estimate bias. Let $\widehat{f}_{Tr}(t, \lambda)$ be $\widehat{f}_{Tr}$ depending on maturity and λ. To estimate bias at λ, assume $\widehat{f}_{Tr}(t, \lambda_0)$ for some small λ_0 is unbiased. Then $\widehat{f}_{Tr}(t, \lambda) - \widehat{f}_{Tr}(t, \lambda_0)$ estimates the bias at $\lambda > \lambda_0$. The variance of $\widehat{f}_{Tr}(t, \lambda)$ is estimated by (9) and (10) below. The choice of λ_0 is discussed in Section 7.

5 The Two-Step Estimation Procedure

Because there is only one Treasury curve but many corporate bond types (and credit spreads), we recommend the two-step procedure, now discussed in more detail:

Step 1: Nonparametric P-spline fitting of a forward rate to US Treasury bonds.

The Treasury forward rate curves f_{Tr} or, equivalently, $\boldsymbol{\delta}$ is estimated by minimizing $Q_{n,\lambda}(\boldsymbol{\delta})$ in (6) and λ is chosen by GCV, RSA, or EBBS as discussed in Section 4. Then $\widehat{f}_{Tr}(t) = \widehat{\boldsymbol{\delta}}'\mathbf{B}(t)$, where $\widehat{\boldsymbol{\delta}}$ are the estimated spline coefficients.

Step 2: Parametric estimation to obtain the forward rate curve for a corporation's bonds.

The forward rate of a corporation's bonds is modelled as $f_C(t) = f_{Tr}(t) +$ polynomial spread, with $f_{Tr} = \widehat{f}_{Tr}$ from the first step.

We adopt polynomial spreads of low degree for several reasons. There are only five AT&T bond prices, so a simple parametric model is necessary. As can be seen Table 1, the maturities of AT&T bonds are between 6 and 11.2 years, so estimation of the spread for $t > 11.2$ is extrapolation, with well-known dangers, and there also is relatively little information about the spread for $t < 0$. Using a simple parametric model of the spread will cause some bias, but this bias should be small. Our reasoning is as follows. The estimated forward rate is between .05 and .07 but the estimated spread is an order of magnitude smaller, about .005; see Figure 5. Moreover, the change in the linear spread between 6 and 11.2 is less than .002. It seems reasonable that the bias using a constant spread is of order .002 or less, at least for $t \in [6, 11.2]$. As reported later, tests that the spread is constant accept this hypothesis. The point is not that the null hypothesis is exactly true, but rather that any deviation from it is likely to be too small to detect or model with the available data. Therefore, the spread will be modelled by a constant term α (constant spread), by $\alpha + \beta s$ (linear spread), or by $\alpha + \beta s + \gamma s^2$ (quadratic spread). The spread parameters can be estimated by parametric nonlinear least-squares with $\boldsymbol{\delta}$ fixed at $\widehat{\boldsymbol{\delta}}$ from step 1.

6 Asymptotic Properties and Inference

In this section we develop asymptotic properties needed for inference and to justify the linearized GCV in Section 4.2.

In the following, n is the number of Treasury bond prices. We only study large-sample asymptotics for Treasury prices. The number of corporate bonds is usually so small that large-sample theory seems pointless. For corporate bonds, we use low-noise asymptotics motivated by the low noise seen in Figure 1.

Asymptotics could be developed with $K \to \infty$ as $n \to \infty$, but fixed-K asymptotic is most relevant for applications where large sample theory provides approximate distributions. These approximations should be most accurate if K is held at the value used in an application.

Since K is fixed, consistency will mean convergence of $\widehat{\boldsymbol{\delta}}$ to $\boldsymbol{\delta}_0$ defined as follows. Assume that the empirical distribution of $\{t_{i,z_i}\}_{i=1}^n$ converges weakly to some limiting distribution F_X. Consider the space of splines with knots equal to $F_X^{-1}\{\ell/(K+1)\}$, $\ell = 1\ldots, K$. Then let $\boldsymbol{\delta}_0$ be the coefficients of the spline that best approximates f_{Tr} in $L^2(F_X)$. Ruppert (2002) results suggest that for smooth f_{Tr}, the best L^2 approximation is quite close to f_{Tr} and the bias due to approximating f_{Tr} by a spline is negligible compared to the standard deviation and bias due to the penalty.

6.1 Large-Sample Asymptotics with $\lambda_n \to 0$

Denote λ by λ_n. The variance of $\widehat{\boldsymbol{\delta}}$ goes to 0 as n tends to ∞ whether or not λ_n tends to 0. However, if $\lambda_n \to 0$ as $n \to \infty$, then the bias also tends to 0 and consistency can be established. The assumptions of the following two theorems are in the appendix. The proofs are similar to those in Yu and Ruppert (2002) and are omitted.

Theorem 1 *Let $(\widehat{\boldsymbol{\delta}}_{n,\lambda_n})$ be a sequence of penalized least squares estimators minimizing (6). Under assumptions 1 and 2, if the smoothing parameter λ_n is $o(n^{-1/2})$, then $\widehat{\boldsymbol{\delta}}_n$ is a (strongly) consistent estimator of true parameter $\boldsymbol{\delta}_0$.*

Theorem 2 *Let $(\widehat{\boldsymbol{\delta}}_{n,\lambda_n})$ be a sequence of penalized least squares estimators of equation (6). Under assumptions* 1 *through* 4*, if the smoothing parameter λ_n is $o(n^{-1/2})$, then $\sqrt{n}(\widehat{\boldsymbol{\delta}}_{n,\lambda_n} - \boldsymbol{\delta}_0) \xrightarrow{D} N(0, \sigma_0^2\Omega^{-1}(\boldsymbol{\delta}_0))$, where $\Omega(\boldsymbol{\delta}_0) := \lim_n \boldsymbol{\Sigma}_n$, is defined in equation (8).*

6.2 Large-Sample Asymptotics with λ fixed and the sandwich formula

The asymptotic variance in Theorem 2 does not involve λ since λ goes to 0. In finite samples this asymptotic variance will over-estimate the variance of $\widehat{\boldsymbol{\delta}}$ which is decreasing in λ, so for inference we give the asymptotic distribution of $\widehat{\boldsymbol{\delta}}$ when λ is fixed.

Using estimating equations, e.g., in Carroll, Ruppert, and Stefanski (1995), we can derive the "sandwich formula" for the asymptotic variance matrix of $\widehat{\boldsymbol{\delta}}(\lambda)$. From (6), $\widehat{\boldsymbol{\delta}}(\lambda)$ is the solution to the estimating equation $0 = \frac{\partial}{\partial \boldsymbol{\delta}} Q_{n,\lambda}(\boldsymbol{\delta}) = \sum_{i=1}^n \psi_i(\boldsymbol{\delta}, \lambda, \mathbf{G})$, where $\psi_i(\boldsymbol{\delta}, \lambda, \mathbf{G}) = -\{P_i - m_i(\boldsymbol{\delta})\}\, m_i^{(1)}(\boldsymbol{\delta})$ $+\lambda \mathbf{G}\boldsymbol{\delta}$. The sandwich formula for the asymptotic variance matrix of $\boldsymbol{\delta}(\lambda)$ is $\widehat{\mathrm{Var}}\left\{\widehat{\boldsymbol{\delta}}(\lambda)\right\} = n^{-1}\boldsymbol{\mathcal{B}}_n^{-1}$

$\mathcal{A}_n\ \mathcal{B}_n^{-1}$, where $\mathcal{A}_n = n^{-1}\sum_{i=1}^n E\{\psi_i(\boldsymbol{\delta},\lambda,\mathbf{G})\psi_i(\boldsymbol{\delta},\lambda,\mathbf{G})'\} = \sigma^2\boldsymbol{\Sigma}_n$, and, with $\boldsymbol{\Sigma}_n$ as in (8), $\mathcal{B}_n = \frac{\partial}{\partial\boldsymbol{\delta}'}\ n^{-1}\sum_{i=1}^n E\{\psi_i(t,T,\boldsymbol{\delta},\lambda,\mathbf{G})\} = \boldsymbol{\Sigma}_n + \lambda\mathbf{G}$. Therefore

$$\widehat{\mathrm{Var}}\{\widehat{\boldsymbol{\delta}}(\lambda)\} = \frac{\sigma^2}{n}\left[\{\boldsymbol{\Sigma}_n+\lambda\mathbf{G}\}^{-1}\boldsymbol{\Sigma}_n\{\boldsymbol{\Sigma}_n+\lambda\mathbf{G}\}^{-1}\right]. \tag{9}$$

Note that as $\lambda \to 0$, $\widehat{\mathrm{Var}}\{\widehat{\boldsymbol{\delta}}(\lambda)\}$ converges to $n^{-1}\sigma^2\boldsymbol{\Sigma}_n^{-1}$ as given in Section 6.1.

6.3 Confidence bands for f_{Tr}

Since the estimated Treasury forward rate at time t for a Treasury STRIPS is $\widehat{\boldsymbol{\delta}}'\mathbf{B}(t)$, a standard error for this forward rate is

$$\mathrm{sd}\{\widehat{f}_{Tr}(t)\} = \sqrt{\mathbf{B}(t)'\left[\widehat{\mathrm{Var}}\{\widehat{\boldsymbol{\delta}}(\lambda)\}\right]\mathbf{B}(t)}, \tag{10}$$

where $\widehat{\mathrm{Var}}\{\widehat{\boldsymbol{\delta}}(\lambda)\}$ is given by (9). By a delta method calculation, the standard error of the estimated discount function, $\widehat{D}_T(t) = \exp(-\mathbf{B}^I(t)\widehat{\boldsymbol{\delta}})$ is easily obtained. From these standard errors, pointwise $(1-\alpha)$ confidence band for $f_{Tr}(t)$ and $D_T(t)$ are obtained in the usual manner.

6.4 Low-Noise Asymptotics

The usual linear model approximation to a nonlinear regression model uses the law of large number for consistency and then a Taylor series linearization and the central limit theorm for asymptotic normality. Since the number of corporate bond prices is generally very low, e.g., 5 in our case study, large-sample theory is of dubious value. However, one can justify the linear model approximation by "low-noise" asymptotics where the variance of the errors goes to 0 with the sample size fixed. Of course, the central limit theory does not apply but it is not necessary if the errors are normally distributed. Low-noise asymptotics are very simple, but as far as we are aware have not be discussed in the context of nonlinear regression. Simultaneous low-noise *and* large-sample asymptotics have been applied to other problems, e.g., transformation models (Bickel and Doksum, 1981; Carroll and Ruppert, 1981) and measurement error (Stefanski, 1985).

We will work with a general nonlinear model $y_i = m(\mathbf{x}_i,\boldsymbol{\theta}_0) = \sigma\epsilon_i$, where $\{\epsilon_i\}_{i=1}^n$ are iid $N(0,1)$. Define $S(\boldsymbol{\theta}) = \sum_{i=1}^n\{m(\mathbf{x}_i,\boldsymbol{\theta}) - m(\mathbf{x}_i,\boldsymbol{\theta}_0)\}^2$. Assume that for all $\Delta > 0$,

$$\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>\Delta} S(\boldsymbol{\theta}) > 0. \tag{11}$$

Define $H(\boldsymbol{\theta}) = \sum_{i=1}^n\{y_i - m(x_i,\boldsymbol{\theta})\}^2$. Then

$$H(\boldsymbol{\theta}) = S(\boldsymbol{\theta}) + 2\sigma\sum_{i=1}^n\epsilon_i\{m(\mathbf{x}_i,\boldsymbol{\theta}_0) - m(\mathbf{x}_i,\boldsymbol{\theta})\} + \sigma^2\sum_{i=1}^n\epsilon_i^2. \tag{12}$$

Let $\widehat{\boldsymbol{\theta}}$ be the least-squares estimator. Assume that for each fixed $\mathbf{x}_i$, $m(\mathbf{x}_i, \boldsymbol{\theta})$ is bounded over the parameter space $\boldsymbol{\Theta}$; this assumption holds in our case study since the discount function takes values on [0,1]. Then by (12) with n fixed as $\sigma \to 0$, $T(\boldsymbol{\theta}) \to S(\boldsymbol{\theta})$ uniformly on $\boldsymbol{\Theta}$, so by (11) $\widehat{\boldsymbol{\theta}} \to \boldsymbol{\theta}_0$ as $\sigma \to 0$.

Therefore, as $\sigma \to 0$, $y_i - \{m(\mathbf{x}_i, \widehat{\boldsymbol{\theta}}) + m^{(1)}(\mathbf{x}_i, \widehat{\boldsymbol{\theta}})'(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \epsilon_i\} = o(\sigma)$. This gives us a linear model in the limit. To use this linear model approximation, let $\boldsymbol{\epsilon}$ be the vector with ith element ϵ_i and let $\mathbf{X}$ be the matrix with ith row equal to $m^{(1)}(\mathbf{x}_i, \widehat{\boldsymbol{\theta}})'$. Then $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0 + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\epsilon} + o(\sigma)$. If follows that as $\sigma \to 0$, then $(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)/\sigma$ converges in distribution to $N(0, (\mathbf{X}'\mathbf{X})^{-1})$, so that $\widehat{\boldsymbol{\theta}} \approx N(\boldsymbol{\theta}_0, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$.

Moreover, F-test statistics for linear hypotheses about $\boldsymbol{\theta}$ will converge in distribution to F-distributions as $\sigma \to 0$. Since $\widehat{\sigma}$ is not consistent in that $\widehat{\sigma}/\sigma$ does not converge to 1 as $\sigma \to 0$, Wald test statistics will not have chi-square limits so F rather than Wald tests should be used.

6.5 Inference about the spread parameters

For inference about the spread, we use f_{Tr} equal $\widehat{f}_{Tr}$ from the first step and treated as fixed. This gives us a parametric nonlinear regression model in the spread parameters. Using the low-noise asymptotics of Section 6.4, this model can be linearized. Confidence intervals and F-tests about the spread parameters are then standard.

7 The AT&T Case Study

In this section, we return to our example of the term structure for AT&T bonds.

7.1 Estimation Results

We applied the two-step procedure to the STRIPS and AT&T bond prices from April 1994 through December 1995. First, we concentrated on STRIPS prices on December 31, 1995, which are plotted in Figure 1. Since the STRIPS are zero coupon bonds, the prices in Figure 1 follow the discount function D.

We used prices on this date as a test bed for our methods, especially of choosing the smoothing parameter. Our main conclusions were:

- The GCV and RSA methods of smoothing parameter selection are unstable for estimation of f_{Tr} with $\widehat{f}_{Tr}$ depending heavily on the number of knots.

- GCV or RSA could be used if the number of knots is chosen carefully, though this introduces subjectivity.

- EBBS, however, is stable with the fitted curve independent of the number of knots.
  - if EBBS and a quadratic spline are used, then anywhere between 5 and 80 knots, and perhaps more, will work well, with $\widehat{f}_{Tr}$ depending very little on K.

- the residuals are autocorrelated.
  - however, autocorrelation does not affect EBBS much, and EBBS can be corrected for autocorrelation.
  - standard errors that assume independence are too small, but can be corrected.

Figure 2 plots the GCV function and the EBBS estimate of the $\sum_{i=1}^{n} \text{MSE}(\widehat{f}_{Tr};\, t, \lambda)$ for 12 values of λ with $\log_{10}(\lambda)$ equally spaced between 0 and 11. Quadratic splines with $K = 25$ knots were used. GCV is minimized at the second smallest value of λ with $\text{DF}(\lambda)$ approximately 24.7, but $\sum_{i=1}^{n} \text{MSE}(\widehat{f}_{Tr};\, t, \lambda)$ is minimized at a much larger λ with $\text{DF}(\lambda)$ near 3.7, effectively 21 parameters less than GCV.

Figure 3 compares the GCV and EBBS estimates of f_{Tr} and D_T, plots of residuals against maturity, the sample autocorrelation function of the residuals, and normal plots of the residuals. The GCV estimate of f_{Tr} is very rough. The problem is that, as discussed in Section 1, GCV is sensitive to $\widehat{D}_T$, not the derivative $\widehat{f}_{Tr}$. The EBBS, RSA and GCV estimates of f_{Tr} are also plotted in Figure 4(a) along with the empirical estimates of f_{Tr} by differencing. GCV tracks the empirical estimates closely while EBBS smooths away local variation. GCV is even more extreme in tracking the data if K is larger. For example, with $K = 80$, GCV chooses $\text{DF}(\lambda) \approx 66$ and the fit oscillates widely attempting to interpolate the empirical estimates. Figure 3 suggests that the residuals are autocorrelated when λ is large but not when λ is small. There are two very different explanations. One is that the "true" f_{Tr} has fine detail that is being estimated correctly with the small value of λ selected by GCV and a large number of knots. If this explanation is true, that the autocorrelation seen in the residuals when λ is large is due to bias. A second explanation is that the fine detail is not "real" but is an indication of autocorrelated errors. In this case, the spline fits the forward rate plus the errors, thus removing the correlation from the residuals. Recall that a principal source of error is price "staleness." If bonds of similar maturity were last traded at about the same time, then staleness would be correlated as would the errors. The data alone cannot decide between the two explanations, since nonparametric regression with correlated errors is a non-identified problem — one cannot distinguish between a high frequency component in the regression function and correlated noise. As with all non-identified problems, subject matter knowledge is essential to decide between the two explanations. In this case, we know of no financial reasons why the true f_{Tr} should have very fine detail so we accept the second explanation.

If the errors are, in fact, positively correlated then the variance of $\widehat{f}_{Tr}$ is somewhat larger than indicated by our standard errors, which assume independence. To correct the standard errors, first we modeled the autocorrelation function (ACF). Parsimonious ARMA models did not fit well, but as can be see in Figure 3, the sample ACF (SACF) decays nearly linearly from 1 to 0 as the lag increases from 0 to 10. Therefore, we modeled the ACF as $\rho(k) = \max\{0, (1 - |k|/10)\}$. This is the ACF of the moving average process $\epsilon_t = u_t + \cdots + u_{t-9}$, where $\{u_t\}$ is white noise. This ACF is plotted in Figure 3, fourth row of the "EBBS" column as "linear SACF." Notice the close agreement with the SACF. Let R_n the be correlation matrix of $\boldsymbol{P}$ under this ACF. The corrected sandwich formula is

$$\widehat{\mathrm{Var}}\{\widehat{\boldsymbol{\delta}}(\lambda)\} = \frac{\sigma^2}{n}\left[\{\boldsymbol{\Sigma}_n + \lambda\mathbf{G}\}^{-1}\boldsymbol{C}_n\{\boldsymbol{\Sigma}_n + \lambda\mathbf{G}\}^{-1}\right]. \tag{13}$$

where $\boldsymbol{C}_n$ is the adjustment of $\boldsymbol{\Sigma}_n$ in (8) for autocorrelation: $\boldsymbol{C}_n = n^{-1}\left[\{\mathbf{m}^{(1)}(\widehat{\boldsymbol{\delta}})\}'\boldsymbol{R}_n\{\mathbf{m}^{(1)}(\widehat{\boldsymbol{\delta}})\}\right]$. What are the implications of autocorrelation for smoothing parameter selection? If the errors are correlated, then the premise behind RSA is false, so its ability to select a correct amount of smoothing is dubious. Also, cross-validation and GCV are known to perform poorly in the presence of correlated noise, since undersmoothing causes the fit to estimate the regression funciton plus the interpolated noise and reduces the CV and GCV criteria (Hart, 1991). EBBS may also undersmooth since it underestimates variance, but is apparently much less susceptible to undersmoothing than GCV and RSA. Also, we modified EBBS method by using (13) as the variance estimate. Figure 2 shows both the uncorrrected and corrected estimates of variance and mse. The corrected EBBS chooses DF(λ) equal to 3.1, compared to 3.7 for uncorrected EBBS. Such a small change in the amount of smoothing has no noticeable effect on the fit.

We experimented with using other values of K besides 25, in particular, values between 5 and 80. The EBBS fitted curves varied little with K. As mentioned above, GCV is even more undersmoothed with large K, say $K = 80$, than with $K = 25$. The GCV or RSA fitted curves are relatively smooth if $K = 8$ and, in fact, for $K = 8$ the GCV, RSA, and EBBS curves are similar; see Figure 4(bottom panel). Thus, one can use GCV if K is chosen appropriately. However, the appropriate choice of K may depend on the particular application, so we recommend EBBS which is stable over a wide range of K. If one is interested only in D, then, besides EBBS, GCV and RSA also appear suitable for selecting λ. For comparison purposes, we will report results for GCV, RSA, and EBBS, but we recommend EBBS.

Since EBBS is choosing between 3 and 4 degrees of freedom, K could be quite small; K sets a maximum of $3 + K$ on DF(λ) when a quadratic spline is used. Five knots is probably sufficient. We do not recommend 25 knots. Rather, we initially used 25 knots to illustrate the problems of GCV and how EBBS avoids them.

We initially used a very wide range of value of λ corresponding to a range for DF(λ) from approximately 3 to 27.2; the maximal range is 3 to 28. However, we used this wide range for exploratory purposes and to show the problems with GCV. In practice we recommend that the range of λ correspond to approximately 3 to 12 degrees of freedom, and even an upper bound of 12 may be unnecessarily high. EBBS uses fit with the largest value of DF(λ) to estimate bias. Fortunately, EBBS seems very insensitive to the choice of the baseline. When the large value of DF(λ) was decreased from 27.2 to approximately 12, the amount of smoothing selected by EBBS did not change.

Schwartz's (1998) method which uses $p = 0$ and no penalty and subjectively chosen knots is the main competitor to the F-N-Z method and our extension thereof, since other spline approaches do not estimate f_{Tr}. Figure 4 (a) compares the fitted forward rates by Schwartz's (1998) method with quadratic splines with λ chosen by EBBS. Schwartz (1998) used the 8 knots located at 1, 2, 3, 4, 6, 8, 10, and 18 years. We used knots at equally-spaced quantiles. In this example, GCV and RSA give essentially the same fit.

The normal plots in Figure 3 shows the GCV residuals to be heavy-tailed. The EBBS residuals are light-tailed though more variable than the GCV residuals.

7.2 Modeling the Spread Function

The F-tests in Section 6 can test hypotheses of economic interest about the spread. A simple model that the spread is constant is tested by testing that $\beta = \gamma = 0$ where the spread is $\alpha + \beta s + \gamma s^2$. The F-statistic for this hypothesis is .098 with a p-value of .91. If instead one tests a constant spread versus the alternative of a linear spread, then the p-value is .20.

The risk of default by AT&T immediately after time $t = 0$ is negligible, and the spread at $t = 0$ is due to liquidity risk, not credit risk. AT&T bonds are less liquid than Treasury bonds, so there is no guarantee that an AT&T bond holder could sell the bond immediately if that were necessary. If cash were needed quickly, the bond holder might need to sell at a discount to find an immediate buyer. This is liquidity risk. In the linear or quadratic spread models, the intercept α of the spread can be interpreted as liquidity risk. The test that $\alpha = 0$ in the quadratic spread model has an F-statistic of 39.35 with a p-value of .024.

In the constant spread model, $\alpha = 0$ corresponds to no spread so one can test for the existence of spread by testing that $\alpha = 0$. For this test, the F-statistic is 996 and the p-value is almost zero. There is extremely strong evidence that a spread exists.

Mallow's C_p was computed for five models: constant, linear, quadratic, linear spread without an intercept, and quadratic spread without an intercept. The C_p values were: 3.05, 2.21, 3.00, 64.92, and 40.35.

In summary, the constant spread model is supported by the data but there is little evidence that more complex models are needed, though the linear spread minimizes C_p. Models without an intercept, which correspond to no liquidity risk, seem contradicted by the data.

Figure 5 graphs the fitted forward rates on AT&T bonds on December 31, 1995, with constant, linear and quadratic spread terms.

Until now, we have only used end of the month data for December 1995. However, modeling the evolution of the term structure is an important problem in finance and is necessary, for example, to price interest rate derivative (Jarrow, 2001). To study this evolution, we fit the AT&T bond prices separately for each month over the 21 month period of April 1994 to December 1995. Figure 6 shows the evolution of the end-of-month forward rates estimated by quadratic P-splines and EBBS, with a constant spread. If we fix maturity, and observe the forward rate as a function of time, then we see a rough curve. This is to be expected, since interest rates move randomly and abruptly. This is why we did *not* use a bivariate smooth in maturity and time.

8 Discussion

We have shown how to use a semiparametric model to estimate the forward rates of individual corporate debt with limited data. When GCV and RSA are used, the number of knots must be chosen carefully, which negates one advantage of P-splines that they are rather insensitive to the number of knots. If EBBS is used, then fitted forward curve is satisfactory regardless of the number of knots, so that EBBS is recommended. The application of EBBS to spline estimation is new and should be of interest in other situations where a derivative is being estimated. Sometimes GCV works satisfactorily for that purpose (Ruppert, Wand, Carroll, 2003), but as we have seen that GCV fails in our case study, probably because of the very low noise and the correlated errors. Low noise and correlated errors are the characteristic distinguishing this example from data sets where we have seen GCV work well for derivative estimation.

We modeled the autocorrelation function and implemented a corrected EBBS. The amount of smoothing selected by correlation-corrected EBBS was indistinguishable from the amount that uncorrected EBBS selected.

By applying our methodology separately to each of a series of months, we can observe the evolution of the term structure. Modeling this evolution is of paramount importance for, inter alia, pricing interest rate derivatives.

A Assumptions

The following assumption is needed for the proof of (strong) consistency.

Assumption 1 *The parameter space* $\mathbf{\Theta}$ *is compact. The mean function* $m(\cdot)$ *is continuous on* $\mathbf{\Theta}$*,* $\frac{1}{n}\sum_{i=1}^{n}\left\{m_i(\boldsymbol{\delta}) - m_i(\tilde{\boldsymbol{\delta}})\right\}^2$ *converges uniformly to some limit in* $\boldsymbol{\delta}, \tilde{\boldsymbol{\delta}} \in \mathbf{\Theta}$*, and* $Q(\boldsymbol{\delta}) = \lim \frac{1}{n}\sum_{i=1}^{n}\left\{m_i(\boldsymbol{\delta}_0) - m_i(\boldsymbol{\delta})\right\}^2$ *exists and has a unique minimum at* $\boldsymbol{\delta} = \boldsymbol{\delta}_0$*.*

Under the following additional assumption, asymptotic normality can be established.

Assumption 2 *The true parameter vector* $\boldsymbol{\delta}_0$ *is an interior point of* $\mathbf{\Theta}$*, the mean function* $m(\cdot)$ *is twice continuously differentiable in a neighborhood of* $\boldsymbol{\delta}_0$ *and* $\Omega(\boldsymbol{\delta}_0) := \lim \frac{1}{n}\sum_{i=1}^{n} m_i^{(1)}(\boldsymbol{\delta}_0) m_i^{(1)}(\boldsymbol{\delta}_0)'$ *exists and is non-singular, where* $m_i^{(1)}(\boldsymbol{\delta}_0) = \left.\frac{\partial m_i(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}'}\right|_{\boldsymbol{\delta}_0}$*. Furthermore,* $\frac{1}{n}\sum_{i=1}^{n} m_i^{(1)}(\boldsymbol{\delta}) m_i^{(1)}(\boldsymbol{\delta})'$ *and* $\frac{1}{n}\sum_{i=1}^{n} \left.\frac{\partial^2 m_i(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}_j \partial \boldsymbol{\delta}_k{}'}\right|_{\boldsymbol{\delta}}$*,* $j, k = 1, ..., \dim(\boldsymbol{\delta})$*, converge uniformly in* $\boldsymbol{\delta}$ *in an open neighborhood of* $\boldsymbol{\delta}_0$*.*

References

Adams, K. J. and D. R. Van Deventer (1994), "Fitting Yield Curves and Forward Rate Curves with Maximum Smoothness", *Journal of Fixed Income*, 52-56.

Bickel, P., and Doksum, K., (1981), "An analysis of transformations revisited," *Journal of the American Statistical Association,* 76, 296–311.

Bielecki, T. and M. Rutkowski (2002), *Credit Risk: Modeling, Valuation and Hedging*, New York: Springer Velag.

Carroll, R.J., and Ruppert, D. (1981), "On prediction and the power transformation family," *Biometrica*, 68, 609–615.

Carroll, R.J., Ruppert, D. and L.A. Stefanski (1995), *Measurement Error in Nonlinear Models*, New York: Chapman & Hall.

Chambers, D. R., W. T. Carleton and D. W. Waldman (1984), "A New Approach to Estimation of the Term Structure of Interest Rates", *Journal of Financial and Quantitative Analysis*, 19, 233–252.

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

De Boor, C. (1978), *A Practical Guide to Splines*, New York: Springer-Verlag.

Dowd, K. (1998), *Beyond Value at Risk: The New Science of Risk Management*, New York: Wiley.

Duffie, D. and K. Singleton (1999), "Modeling Term Structures of Defaultable Bonds", *Review of Financial Studies*, 12, 197–226.

Ellner, S. P. and Y. Seifu (2002), "Using Spatial Statistics to Select Model Complexity", *Journal of Computational and Graphical Statistics*, 11, 348–369.

Eubank, R. (1999). *Nonparametric Regression and Spline Smoothing*, New York: Marcel Dekker.

FCIC (Federal Deposit Insurance Corporation) (2000), "Options Paper", August issue.

Fisher, M., D. Nychka and D. Zervos (1995), "Fitting the Term Structure of Interest Rates with Smoothing Splines", Working Paper, Finance and Economics Discussion Series, Federal Reserve Board. (available upon request at Mark.Fisher@atl.frb.org)

Friedman, J. (1991), "Multivariate adaptive regression splines (with discussion)," *The Annals of Statistics*, 19, 1–141.

Friedman, J., and Silverman, B., (1989), "Flexible parsimonious smoothing and additive modeling (with discussion)," *Technometrics*, 31, 3–39.

Frishling, V. and J. Yamamura (1997), "Fitting a Smooth Forward Rate Curve to Coupon Instruments", *Journal of Fixed Income*, 97–103.

Green, P. J. and B. W. Silverman (1994), *Nonparametric Regression and Generalized Linear Models*, London: Chapman & Hall.

Hart, J. (1991), "Kernel regression estimation with time series errors," *Journal of the Royal Statistical Society, Series B,* 53 , 173–187.

Hastie, T. J. and Tibshirani, R. (1990), *Generalized Additive Models*, London: Chapman & Hall.

Heath, D., Jarrow R. and Morton D. (1992), "Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claim Valuation", *Econometrica*, 60, 77–105.

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, 2nd edition*, Stanford University Press: Stanford, CA.

Jarrow, R. and S. Turnbull (1995), "Pricing Derivatives on Financial Securities Subject to Credit Risk", *Journal of Finance*, 50, (1): 53–85.

Jarrow, R. and S. Turnbull (2000), *Derivative Securities*, 2nd edition, South-Western Pub.: Cincinnati, OH.

Jarrow, R. and F. Yu (2001), "Counterparty Risk and the Pricing of Defaultable Securities", *Journal of Finance*, 56, 1765–1799.

Jorion, P. (2000), *Value at Risk: The New Benchmark for Managing Financial Risk*, McGraw-Hill.

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

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

O'Sullivan, F. (1986). "A statistical perspective on ill-posed inverse problems (with discussion)," *Statistical Science*, 1, 505–527.

Risk Magazine (2002), "Credit Derivatives Survey: A Risk Special Report", February 2002 issue.

Ruppert, D. (1997), "Empirical-bias Bandwidths for Local Polynomial Nonparametric Regression and Density Estimation," *Journal of the American Statistical Association*, 92, 1049-1062.

Ruppert, D. (2002), "Selecting the number of knots for penalized splines", *Journal of Computational and Graphical Statistics*, to appear. (available at www.orie.cornell.edu/~davidr/papers)

Ruppert, D., and R. J. Carroll (2000), "Spatially-Adaptive Penalties for Spline Fitting", *Australian and New Zealand Journal of Statistics*, 42, 205–223.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2002), *Semiparametric Regression*, Cambridge University Press, to appear.

Schwartz, T. (1998), "Estimating the Term Structures of Corporate Debt", *Review of Derivatives Research*, 193–230.

Shea, G. (1985), "Interest Rate Term Structure Estimation with Exponential Splines: A Note", *Journal of Finance*, 40, 319–325.

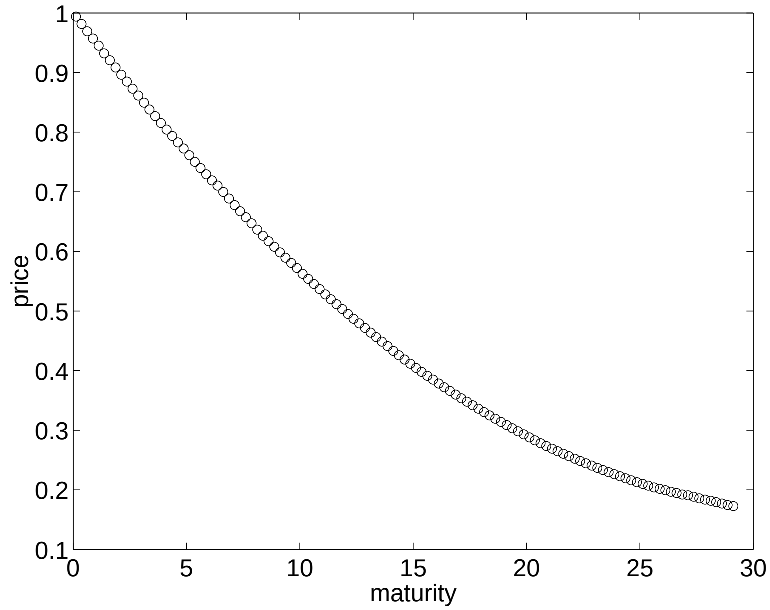
Stefanski, L. (1985), "The effects of measurement error on parameter estimation," *Biometrika*, 72, 583–592.

Vasicek, O. A. and H. G. Fong (1982), "Term Structure Modeling Using Exponential Splines", *Journal of Finance*, 37, 339–356.

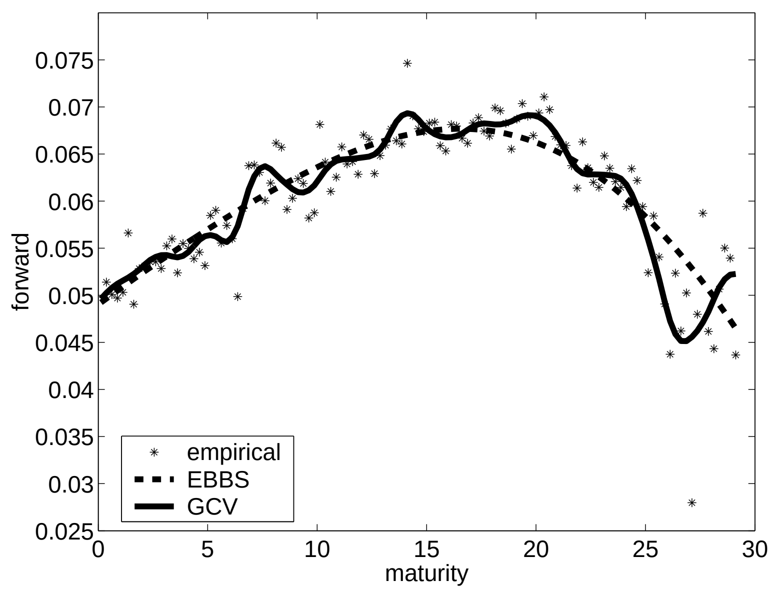
Wahba, G. (1990), *Spline Models for Observational Data*, SIAM: Philadelphia.

Warga, A. (1995), "Fixed Income Securities Database", Fixed Income Research Program, University of Houston. (Request information at www.uh.edu/~awarga)

Yu, Y. and D. Ruppert (2002), "Penalized Spline Estimation for Partially linear Single-Index Models", *Journal of the American Statistical Association*, to appear in December, 2002 issue. (available at www.statqa.cba.uc.edu/~yuy/research.htm)

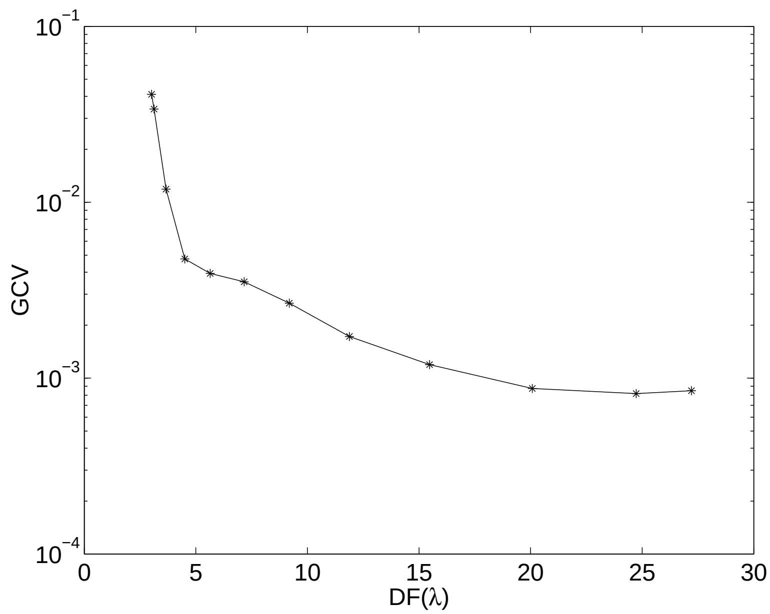


(a)

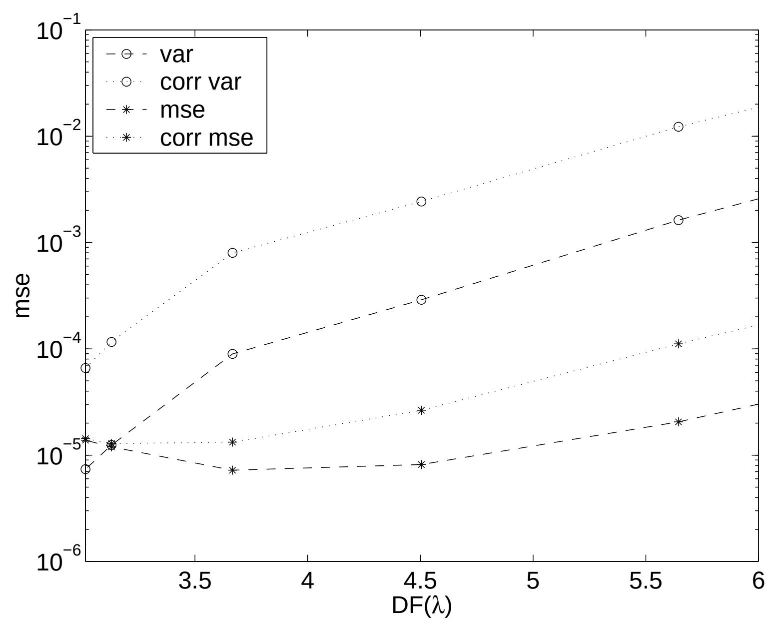


(b)

Figure 1: (a) Price versus maturity for the US Treasury STRIPS. Note the very low noise. (b) Forward rate estimates with 25 knot quadratic splines. The empirical estimate is the time series of finite different quotients, that is, the ratios of changes in minus the log-prices to changes in maturity times when the data are order by maturity time; see text on page 1. P-splines have λ estimated by EBBS and GCV.



(a)



(b)

Figure 2: (a) GCV as function of $\mathrm{DF}(\lambda) = \mathrm{tr}\{\mathbf{A}(\lambda)\}$. (b) Estimated sum of MSE's and sum of variances for estimated forward rates as function of $\mathrm{DF}(\lambda)$. Quadratic spline with 25 knots.

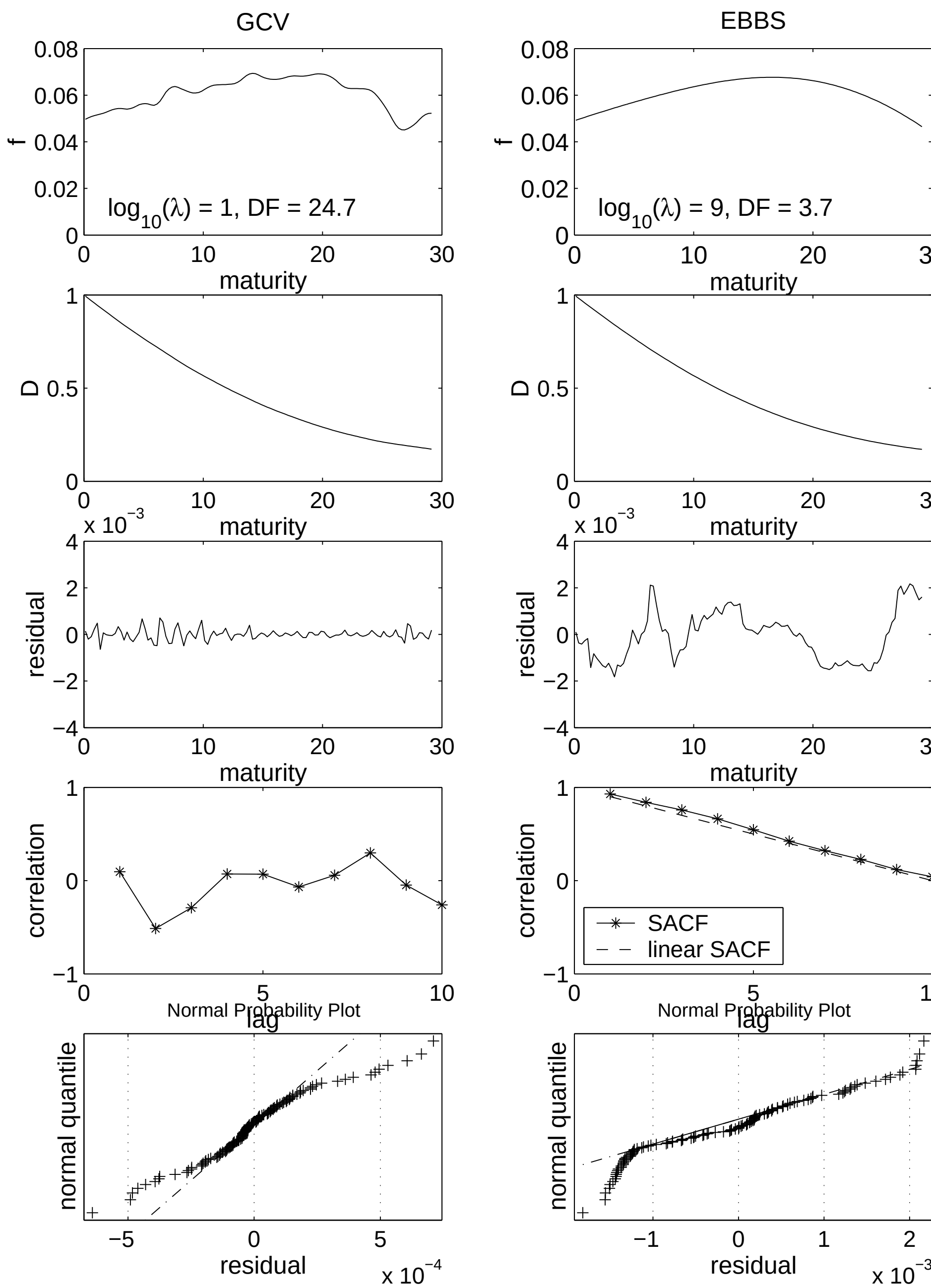


Figure 3: The fitted forward rate and discount function, residuals, and sample autocorrelation function of the residuals for λ chosen by GCV (left) and by EBBS (right). Normal plots of residuals also included.

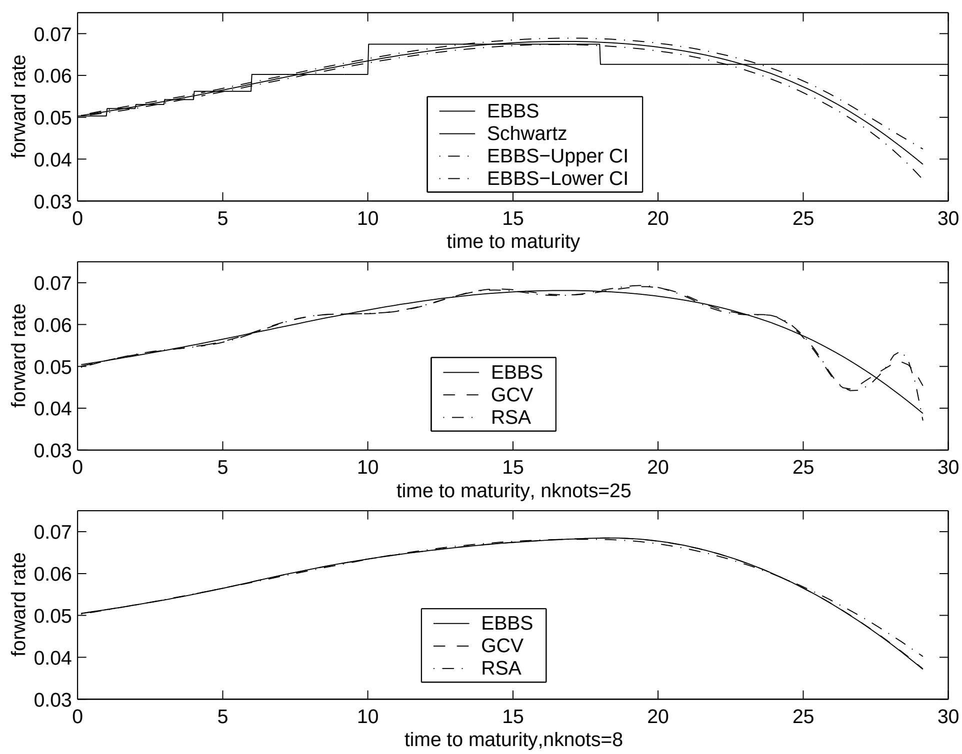


Figure 4: Fitted forward rate curves on US STRIPS on December 31, 1995. (a) Quadratic P-splines with λ chosen by EBBS compared to Schwartz's (1998) piecewise constant spline with no penalty. (b) Comparison of 25-knot quadratic P-spline fits by GCV, RSA, and EBBS. (c) Same as (b) but with only 8 knots.

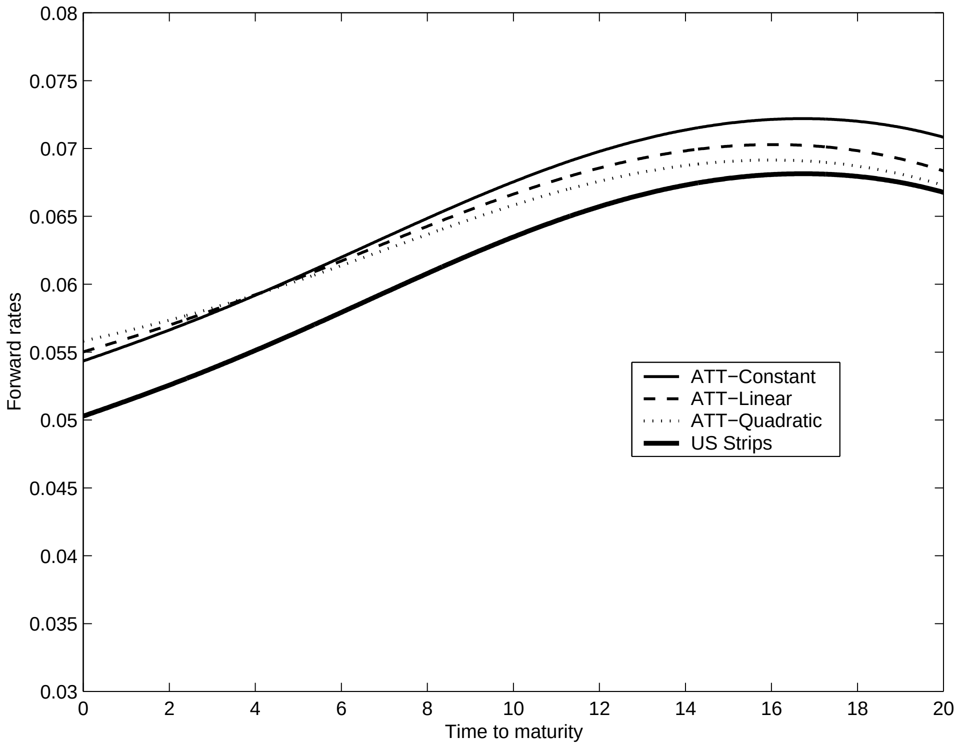


Figure 5: Fitted forward rate curves for AT&T bonds on December 31, 1995. Constant, linear, and quadratic spreads. Note: there are no AT&T bonds with maturities beyond 11.2 years, so the estimates beyond that maturity are extrapolations.

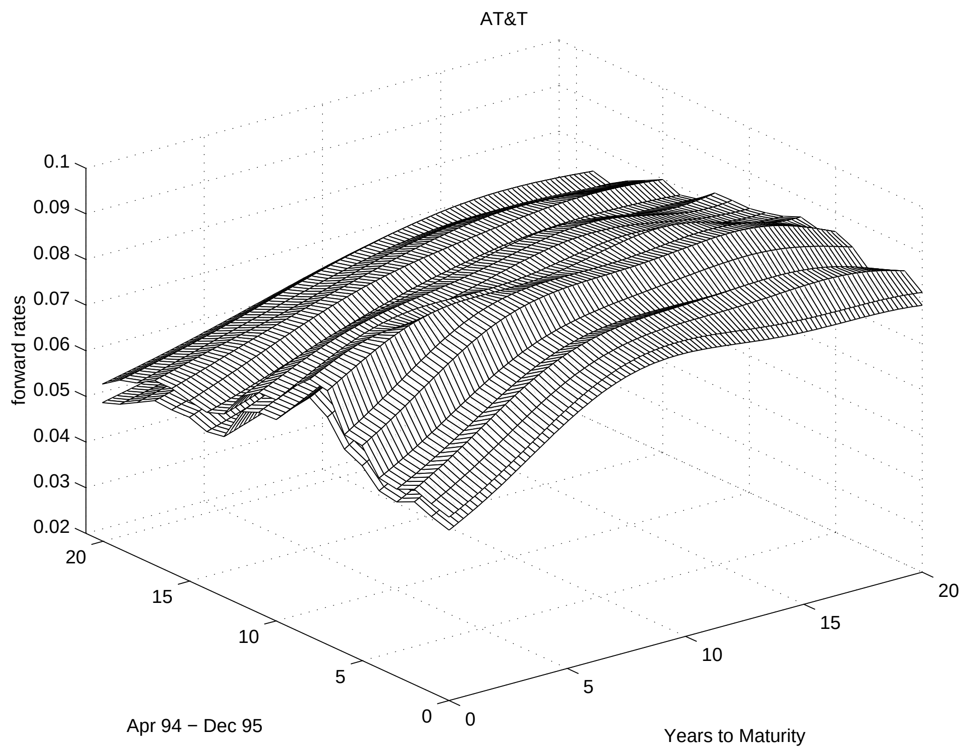


Figure 6: Fitted forward rate curves on AT&T and US STRIPS over the 21 month period of April 1994 to December 1995, $p = 2$ and smoothing parameter λ is chosen by GCV with constant spread.