Universität St.Gallen

A simple and general approach to fitting the discount curve under no-arbitrage constraints

Matthias R. Fengler, Lin-Yee Hin

August 2014 Discussion Paper no. 2014-23

School of Economics and Political Science, Department of Economics

University of St.Gallen

Editor: Martina Flockerzi
University of St.Gallen
School of Economics and Political Science
Department of Economics
Bodanstrasse 8
CH-9000 St. Gallen
Phone +41 71 224 23 25
Fax +41 71 224 31 35
Email seps@unisg.ch

Publisher: School of Economics and Political Science
Department of Economics
University of St.Gallen
Bodanstrasse 8
CH-9000 St. Gallen
Phone +41 71 224 23 25
Fax +41 71 224 31 35

Electronic Publication: http://www.seps.unisg.ch

A simple and general approach to fitting the discount curve under no-arbitrage constraints

Matthias R. Fengler, Lin-Yee Hin

Authors' address:

Matthias R. Fengler
School of Economics and Political Science
University of St.Gallen
Bodanstrasse 6
CH-9000 St. Gallen
Phone +41 (0)71 224-2457
Fax +41 (0)71 224-2894
Email matthias.fengler@unisg.ch

Lin-Yee Hin
Department of Mathematics and Statistics
Curtin University of Technology
Kent Street
Bentley WA 6102
Australia
Email linyeehin@gmail.com

Abstract

We suggest a simple and general approach to fitting the discount curve under no-arbitrage constraints based on a penalized shape-constrained B-spline. Our approach accommodates B-splines of any order and fitting both under the L_1 and the L_2 loss functions. Simulations and an empirical analysis of US STRIPS data from 2001-2009 suggest that an active knot search and splines of order three and four are mandatory to obtain reasonable fits. The loss function appears to be less relevant.

Keywords

B-splines; Discount curve; No-arbitrage constraints; Monotone estimation; Yield curve.

JEL Classification

C44; C58; C61.

1 Introduction

The discount curve is the key ingredient in the valuation and risk-management not only of fixed income securities, but of almost any financial instrument that promises future cash-flows, such as structured debt, equity, and foreign exchange products, life annuities, and others. Moreover, the discount curve defines two additional objects of interest: (*i*) the term structure of interest rates, also called spot rate curve or yield curve; and (*ii*) the forward rate curve, which represents the return on an investment over an infinitesimally small time horizon that is arranged today, but made in the future.

To fix ideas, denote by $d : [0, +\infty) \to (0, 1]$ the discount function that must be monotonically decreasing by static no-arbitrage arguments; see McCulloch (1971). Under continuous compounding, the spot rate function associated with $d(\cdot)$ is the mapping $r : (0, +\infty) \to [0, +\infty)$ given by

$$r(\tau) = -\frac{1}{\tau} \log d(\tau) , \tag{1}$$

while the forward rate curve is described by the mapping $f : [0, +\infty) \to [0, +\infty)$ defined by

$$f(\tau) = -\frac{d'(\tau)}{d(\tau)} , \tag{2}$$

where $d'(\cdot)$ denotes the first-order derivative of $d(\cdot)$.

Because the discount curve is not directly observable, it needs to be estimated from prices of fixed income instruments, such as government bonds or fixed income futures. Due to their flexibility, spline estimators, usually based on B-splines, have a long tradition of being used for this purpose. Simple spline approximations have been suggested, inter alia, in McCulloch (1971), McCulloch (1975), Vasicek and Fong (1982), Litzenberger and Rolfo (1984), and Coleman et al. (1992); see Marangio et al. (2002) for a review and Ioannides (2003) and Yallup (2012) for empirical comparisons of these unconstrained spline methods on UK treasury bills and gilt data.

More recently, in an effort to obtain statistically more efficient, and in particular, arbitrage-free estimates of the discount curve, the focus has shifted toward shape-constrained spline approximations. Although well researched over the past decades, the shape-constrained estimation of the discount curve still lacks a coherent and satisfying framework to date. Extant estimators are often involved, are strongly tied to a specific degree of the spline, or depend on arguable assumptions. For instance, Laurini and Moura (2010) suggest a penalized linear B-spline approximation to the discount curve that is obtained by minimizing the L_1 loss function. The definition of the forward rate curve in (2), however, suggests that the discount function should be continuously differentiable. A linear B-spline approximation may therefore not be satisfactory for certain applications. Chiu et al. (2008) use a cubic B-spline instead, but for monotonicity to hold, they impose a local convexity condition to hold piecewise between the spline segments. This assumption could be questioned, as there is no economic theory that would require the discount curve to be convex, either locally or globally. Moreover, an order of differentiability is lost. Similarly, the cubic forward rate interpolation scheme of Hagan and Graeme (2006) depends on a convexity condition. In building on the monotonized (cubic) B-spline basis owed to Ramsay (1988), Ramponi (2003) avoids the assumption of local convexity for achieving monotonicity, but the estimator does not include the additional property $1 \geq d(\tau) > 0$. Another cubic B-spline is proposed in Barzanti and Corradi (1999), in which the L_1 solution is obtained by means of a linear program under the assumption of a one-sided error distribution. In this context, a one-sided error distribution may not be a natural assumption, because it yields estimates that are biased.

In this paper, we aim at closing this gap by providing a simple and general approach to fitting the discount curve. The estimator is a penalized shape-constrained B-spline. Smoothness of any order is achieved without any additional assumptions, although it is likely that quadratic or cubic B-splines are sufficient for practice. As penalties, we consider the ridge and a first-order difference penalty. The penalty regularizes the estimate in the presence of sparse data, and at the same time, penalizes oscillatory behavior. Because it is often argued that L_1 estimators are preferable, due to their robustness against outliers and their more benign behavior (He and Shi, 1998; Lavery, 2000; Cheng et al., 2005), we discuss both L_1 and L_2 estimation. Both estimation methods are straightforward to implement, because one can take advantage of standard solvers for quadratic programs and quantile regression. Finally, we discuss knot placement based on an active knot search and a subsequent knot deletion or relocation.

The generality of our framework allows us to study a wide range of different, but mutually comparable estimators. We consider a total of twelve estimators. They are constructed using linear, quadratic, and cubic B-splines, penalized by the ridge and the Eilers and Marx (1996) first-order finite difference penalties, an solved by minimizing either the L_1 or the L_2 loss function. The estimators are applied to simulated data and nine years of US STRIPS data from 2001-2009. We find modest differences among the estimators, with the quadratic and cubic ones performing the best, regardless of the loss function used for estimation.

The paper is organized as follows. In Section 2, we introduce the shape-constrained B-spline estimator. The simulation results are provided in Section 3 and the empirical applications in Section 4. Section 5 concludes.

2 Shape-constrained discount curve smoothing with B-splines

The purpose is to approximate up to some error ε with mean zero and finite variance, the unknown discount curve by means of a B-spline basis of order q (degree $q-1$):

$$d(\tau) = \sum_{j=1}^{N+q} \theta_j B_j(\tau; q) + \varepsilon \, , \tag{3}$$

where for $j = 1, \ldots, N+q$, $B_j(\tau; q)$ denotes the B-spline basis function of order q (degree $q-1$) and θ_j its parameter weight. We define the B-splines over the strictly increasing knot sequence

$$\xi_1 = \cdots = \xi_q < \xi_{q+1} < \cdots < \xi_{q+N} < \xi_{q+N+1} = \cdots = \xi_{N+2q} \, ,$$

where N is the number of interior knots. Because the interior knots are not coinciding, the spline will be differentiable $q-2$ times; see de Boor (2001) for a precise definition of B-splines and their properties.

In the following, we wish to obtain estimates $\hat{\theta}_j, j = 1, \ldots, N+q$, such that the discount factor function $d(\tau) = \sum_{j=1}^{N+q} \hat{\theta}_j B_j(\tau; q)$ does not admit arbitrage, i.e., (1) it is positive, (2) monotonically decreasing in time-to-maturity, (3) and obeys the limit constraints $d(0) = 1$ and $\lim_{\tau \to \infty} d(\tau) = 0$.

2.1 Estimation frameworks

We suppose that we are given a sample of observed discount factors $\{\tau_i, d(\tau_i)\}_{i=1}^{n}$.[1] Denote by $\mathbf{B}(\tau_i; q)$ the $(N+q) \times 1$ vector of B-spline basis functions of order q, evaluated at τ_i, $i = 1, \ldots, n$, and collect the sequence of B-spline weights in the vector $\boldsymbol{\theta} = \{\theta_j\}$, $j = 1, \ldots, (N+q)$. We will seek solutions to $\boldsymbol{\theta}$, both in the L_2 and the L_1 loss functions.

For the L_2 loss function, we minimize

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \mathbb{R}^{N+q}}{\arg\min} \sum_{i=1}^{n} (d_i - \mathbf{B}(\tau_i; q)^\top \boldsymbol{\theta})^2 + \lambda P_2 \,, \tag{4}$$

$$\text{subject to} \qquad 1 \geq \theta_1 > \theta_2 > \ldots > \theta_{N+q} \geq 0 \,, \tag{5}$$

where λ is the penalty parameter and P_2 is the penalty term. Specifically, P_2 can take the form of a ridge penalty given by $P_{2,R} = \sum_{j=1}^{N+q} |\theta_j|^2$ and the form of the first-order difference penalty $P_{2,EM1} = \sum_{j=1}^{N+q-1} |\theta_{j+1} - \theta_j|^2$ as introduced by Eilers and Marx (1996). This latter penalty, for which one can define generalizations in terms of higher-order difference penalties, differs from the usual penalty for the smoothing spline in that it only provides an approximation to the first-order derivative. The theoretical properties of this class of splines have been studied only recently in Claeskens et al. (2009).

When we search for an L_1 solution, we minimize

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \mathbb{R}^{N+q}}{\arg\min} \sum_{i=1}^{n} \left| d_i - \mathbf{B}(\tau_i; q)^\top \boldsymbol{\theta} \right| + \lambda P_1 \,, \tag{6}$$

$$\text{subject to} \qquad 1 \geq \theta_1 > \theta_2 > \ldots > \theta_{N+q} \geq 0 \,, \tag{7}$$

where the penalties now take the forms $P_{1,R} = \sum_{j=1}^{N+q} |\theta_j|$ and $P_{1,EM1} = \sum_{j=1}^{N+q-1} |\theta_{j+1} - \theta_j|$, respectively.

The constraints (5) and (7) implement the desired constraints of the upper and lower bounds on the discount factor and monotonicity. The first are ensured by the leftmost and rightmost inequalities,

[1] For the sake of simplicity, we assume the existence of a sample of observed discount factors. While this holds true for the US STRIPS data we use in our empirical application, in many practical cases, this may not apply. Typically, the samples include observed bond prices and their coupon payments. It is, however, not difficult to rewrite the objective functions for this case.

because the B-spline basis functions are positive and form a partition of unity. The inner inequalities ensure monotonicity by the variation diminution property of the B-spline: the number of sign changes in d is, at most, as large as in the sequence of $\{\theta_j\}$; see de Boor (2001, pp. 138–142). For linear and quadratic splines, these constraints are both necessary and sufficient, but for splines with higher degrees, they are only sufficient. The sufficiency part is easily seen as follows. Suppose $q \geq 2$ and, for simplicity, that the knots are equidistant with width Δ. Differentiating the spline function $d(\tau)$, one obtains $d'(\tau) = \sum_j \theta_j B'_j(\tau; q) = \Delta^{-1} \sum_j (\theta_j - \theta_{j-1}) B_j(\tau; q-1)$ by the properties of B-splines; see de Boor (2001, p. 116). Thus, $\theta_j < \theta_{j-1}$ implies a monotonically decreasing spline function. In consequence, the constraints (5) and (7) implement the desired constraints for any degree of the polynomial spline space.

Remarkably, these constraints appear to be rarely used, although they are easy to implement and originate from basic properties of the B-spline. Kelly and Rice (1991) use them to obtain a monotonic B-spline for fitting a dose-response curve. They also appear in the discount factor model of Barzanti and Corradi (1999), and more recently, in Fengler and Hin (2013), to impose monotonicity along an axial direction of a call-option price surface modeled by a tensor-product B-spline. Yet most studies on monotonic B-splines suggest imposing monotonicity directly on the resulting regression function, which only works well for linear and quadratic splines; see, e.g., He and Shi (1998), He and Ng (1999), or Meyer (2012). It fails, however, for cubic and higher-degree polynomials, because it is very intricate to characterize monotonicity in terms of conditions on the parameters; see Fredenhagen et al. (1999) for an in-depth discussion. Therefore, for cubic and higher-degree splines, one needs to content oneself with spline approximations that achieve monotonicity only at a finite collection of points; see Tobler (1996) for such an approach to modeling discount factors. Alternatively, one requires more structural assumptions, such as local convexity, as in Chiu et al. (2008), or additional and more sophisticated optimization techniques, as proposed in Turlach (2005) and Papp and Alizadeh (2014).

2.2 Implementation of the estimators

To describe the implementation of the estimators, we introduce the following notation. Define $\mathbf{P}_R = \mathbf{I}_{N+q}$, which is the unit matrix of size $(N + q)$, and the $(N + q - 1) \times (N + q)$ matrix

$$\mathbf{P}_{EM1} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & -1 & 1 \end{pmatrix} . \tag{8}$$

Moreover, denote by

$$\mathfrak{B} = \begin{pmatrix} \mathbf{B}(\tau_1; q)^\top \\ \vdots \\ \mathbf{B}(\tau_n; q)^\top \end{pmatrix} \tag{9}$$

the $n \times (N + q)$ B-spline collocation matrix for all n observations and collect the observed discount rates into the vector $\mathbf{d} = (d_1, \dots, d_n)^\top$. To implement the no-arbitrage constraints, we introduce the banded $(N + q + 1) \times (N + q)$ matrix

$$\mathbf{M}_1 = \begin{pmatrix} -1 & 0 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \dots & & 1 & -1 \\ 0 & \dots & & 0 & 1 \end{pmatrix} \tag{10}$$

and the vector $\mathbf{m}_1 = (-1, \mathbf{0}_{1\times(N+q)})^\top$, where $\mathbf{0}_{(N+q)\times 1}$ is an $(N+q)\times 1$ column vector, with all elements being zero.

In the L_2 estimation framework, for $\mathbf{P} \in \{\mathbf{P}_R, \mathbf{P}_{EM1}\}$, we estimate the B-spline coefficients $\boldsymbol{\theta}$ by means of the quadratic program

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^{N+q}} \quad \boldsymbol{\theta}^\top \left(\mathfrak{B}^\top\mathfrak{B} + \lambda\mathbf{P}^\top\mathbf{P}\right) \boldsymbol{\theta} - 2\mathbf{d}^\top\mathfrak{B}\boldsymbol{\theta} \tag{11}$$

$$\text{s.t.} \qquad \mathbf{M}_1\boldsymbol{\theta} \geq \mathbf{m}_1 \ , \tag{12}$$

which can be solved using standard quadratic programming methods, such as the Goldfarb-Idnani algorithm. We will make use of the function `solve.QP()` from the `R` package `quadprog` that is owed to Turlach and Weingessel (2014).

If the estimated spline function is defined at $\tau = 0$, i.e., if the knot sequence starts at zero by choosing $\xi_1 = \dots \xi_q = 0$, it can be useful to add the more powerful constraint $d(0) = 1$ to the program. This goal is achieved by introducing the $(N+q) \times (N+q)$ matrix

$$\mathbf{M}_2 = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \dots & & 1 & -1 \\ 0 & \dots & & 0 & 1 \end{pmatrix} \tag{13}$$

and the vector $\mathbf{m}_2 = (1, \mathbf{0}_{1\times(N+q-1)})^\top$. Then, constraint (12) must be replaced by

$$\begin{aligned} & \mathbf{M}_2\boldsymbol{\theta} \geq \mathbf{0}_{(N+q)\times 1} \\ \text{and} \qquad & \mathbf{m}_2^\top \boldsymbol{\theta} = 1 \ , \end{aligned} \tag{14}$$

which can be solved by standard quadratic programming methods as well.

To estimate the L_1 solution, further notation is required. For $\mathbf{P} \in \{\mathbf{P}_R, \mathbf{P}_{EM1}\}$, define the $(n+p) \times (N+q)$ matrix

$$\widetilde{\mathbf{X}} = \begin{pmatrix} \widetilde{\mathbf{x}}_1 \\ \vdots \\ \widetilde{\mathbf{x}}_{n+p} \end{pmatrix} = \begin{pmatrix} \mathfrak{B} \\ \lambda \mathbf{P} \end{pmatrix} \tag{15}$$

and introduce the $(n+p)$-vector

$$\widetilde{\mathbf{y}} = \begin{pmatrix} \widetilde{y}_1 \\ \vdots \\ \widetilde{y}_{n+p} \end{pmatrix} = \begin{pmatrix} \mathbf{d} \\ \mathbf{0}_{p\times 1} \end{pmatrix} \ , \tag{16}$$

where $p = N + q$ for $\mathbf{P}_R$ and $p = (N + q - 1)$ for $\mathbf{P}_{EM1}$. As noted in Ng and Maechler (2007), this notation allows one to cast the objective function $\sum_{i=1}^{n} |d_i - \mathbf{B}(\tau_i; q)^\top \boldsymbol{\theta}| + \lambda P_{1,j}$, $j = \{R, EM1\}$, into the more compact form $\sum_{i=1}^{n+p} |\widetilde{y}_i - \widetilde{\mathbf{x}}_i^\top \boldsymbol{\theta}|$.

In summary, we express the constrained L_1 problem as

$$\begin{aligned} \min_{\boldsymbol{\theta} \in \mathbb{R}^{N+q}} \quad & \sum_{i=1}^{n+p} \left| \widetilde{y}_i - \widetilde{\mathbf{x}}_i^\top \boldsymbol{\theta} \right| \\ \text{s.t.} \quad & \text{constraints (12) or (14)} \ . \end{aligned} \tag{17}$$

This linear program is a constrained median regression framework, which can be solved, for instance, using the Frisch-Newton algorithm as described in Koenker and Ng (2005). Specifically, we use the low level function `rq.fit.fnc()` from the `R` package `quantreg` which is owed to Koenker (2013).

2.3 Knot search and determination of the penalty parameter

The estimator requires both the selection of knots and of the penalty parameter. We address these questions iteratively, i.e., we first determine the knot sequence, and conditionally on this, select the penalty parameter λ. Because a good knot placement is an essential parametrization problem for the B-spline, we do not fix knots on an equidistant grid, but make use of a guided knot search followed by relocation and deletion strategies. This follows suggestions in He and Shi (1996), Zhou and Shen (2001), and Fengler and Hin (2013).

As in He and Ng (1999), all model selections are guided by a Schwarz information criterion (SIC). We define it by

$$\mathrm{SIC} = \log(\mathrm{ASR}) + \frac{1}{2}\rho(\lambda)\frac{\log(n)}{n}\,, \tag{18}$$

where $\mathrm{ASR} = \frac{1}{n}\sum_{i=1}^{n}(d_i - \hat{d}_i)^2$ and $\rho(\lambda) = \mathrm{tr}\{\mathfrak{B}(\mathfrak{B}^\top\mathfrak{B} + \lambda\mathbf{P}^\top\mathbf{P})^{-1}\mathfrak{B}^\top\}$ is an approximation to the degrees of freedom in the sense of Hastie and Tibshirani (1990).[2] For knot insertion, we set the penalty parameter to $\lambda = 10^{-10}$ and employ a B-spline of order $q = 2$. With all constraints being active and given boundary knots, we search over each subinterval in the working knot sequence such as to minimize SIC. This search is run for a user-specified number of layers; for our simulations and empirical demonstrations, four layers have appeared to be sufficient.

Next, we provide a knot adjustment strategy. We employ the B-spline at the desired order. We then iterate over each subinterval in the knot sequence resulting from the knot insertion phase and check, for each knot, whether (*i*) the knot deletion and (*ii*) the knot relocation result in an improvement of the SIC. After determining the knot sequence in this way, we minimize the SIC over the penalty

[2]The definition of $\rho(\lambda)$ is borrowed from the unconstrained L_2 solution. Thus, in the constrained L_2 case, it is only an approximation to the true degrees of freedom, which depend on the random number of active constraints. For the L_1 loss function, the case is yet more difficult, because even for unconstrained regression problems, appropriate definitions of model selection criteria continue to be heavily researched; see Gao and Fang (2011) for further details. Our procedure is just a practical way of addressing the difficult problem of model selection in this context.

parameter λ. The value at the smallest SIC is taken as the optimal penalty parameter.

2.4 Relation to other constrained models of the discount curve

Laurini and Moura (2010) suggest using the linear COBS procedure owed to He and Ng (1999) to estimate the discount factor curve. In this approach, for a given knot sequence, one minimizes the objective function

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^{N+q}} \quad \sum_{i=1}^{n} \left| d_i - \mathbf{B}(\tau_i; q)^\top \boldsymbol{\theta} \right| + \alpha \sum_{i=1}^{N} \left| \mathbf{B}'(\xi_{i+q}; q)^\top \boldsymbol{\theta} - \mathbf{B}'(\xi_{i+q-1}; q)^\top \boldsymbol{\theta} \right| \tag{19}$$

$$\text{s.t.} \qquad \mathbf{H}\boldsymbol{\theta} \geq \mathbf{0}_{(N+q)\times 1} \ , \tag{20}$$

where α is a penalty parameter and

$$\mathbf{H} = \begin{pmatrix} B_1'(\xi_q; q) & \cdots & B_{N+q}'(\xi_q; q) \\ \vdots & & \vdots \\ B_1'(\xi_{N+q+1}; q) & \cdots & B_{N+q}'(\xi_{N+q+1}; q) \end{pmatrix} \tag{21}$$

and $q = 2$. $B'(\cdot; q)$ denotes the first-order derivative of the B-spline basis function. The constraint $\mathbf{H}\boldsymbol{\theta} \geq \mathbf{0}$ ensures the monotonicity of the estimate.[3] Despite the notational differences, however, the COBS penalty coincides up to a factor of proportionality with the ridge penalty $P_{1,R}$ of the penalized B-spline when $q = 2$. Because our monotonicity constraints are both sufficient and necessary, the monotonicity constraint is also equivalent to our formulation. Hence, under the L_1 loss function and for $q = 2$, the COBS estimator, as applied by Laurini and Moura (2010), is obtained as a special case of our set-up. As noted in the introduction, the monotonicity constraints, as given in (20), do not generalize in a linear way for orders $q \geq 4$; see He and Shi (1998, p. 644).

Our estimator is also related to that of Barzanti and Corradi (1999), who consider a linear program of an unpenalized cubic B-spline approximation to the discount curve using the same formulation of the constraints as in (14). Yet for $q = 4$ and $\lambda = 0$, their L_1 estimator differs, because they estimate under the assumption of a one-sided error distribution. As discussed in Preve and Medeiros (2011), this yields biased estimates.

[3]For the sake of space, we disregard here the additional constraints that maintain positivity and ensure $d(0) \leq 1$.

3 Monte Carlo simulations

3.1 Simulation set-up

As in Ramponi (2003), we simulate from the Nelson-Siegel specification of the discount factor; see Nelson and Siegel (1987). This functional form is a widely accepted to be sufficiently flexible to mimic the most important shapes of empirically observed yield curves. It is given by

$$d(\tau) = \exp\left[-\tau\left\{\beta_0 + (\beta_1 + \beta_2)\left(1 - \exp\left(-\frac{\tau}{\vartheta}\right)\right)\frac{\vartheta}{\tau} - \beta_2 \exp\left(-\frac{\tau}{\vartheta}\right)\right\}\right] , \tag{22}$$

where $\beta_0, \beta_1, \beta_2$, and ϑ are parameters. The spot rate function is given by

$$r(\tau) = \beta_0 + (\beta_1 + \beta_2)\left\{1 - \exp\left(-\frac{\tau}{\vartheta}\right)\right\}\frac{\vartheta}{\tau} - \beta_2 \exp\left(-\frac{\tau}{\vartheta}\right) \tag{23}$$

and the forward rate function by

$$f(\tau) = \beta_0 + \left(\beta_1 + \beta_2\frac{\tau}{\vartheta}\right)\exp\left(-\frac{\tau}{\vartheta}\right) . \tag{24}$$

For the simulation, we choose $M = 200$ regularly spaced tenors ranging from $\tau_1 = 0.15$ to $\tau_{200} = 30$ years. The Nelson-Siegel yields $r(\tau_j)$, $j = 1, \ldots, M$, are perturbed using normal mean-zero error with standard deviation σ_j which is allowed to vary in j. To make our modeling of heteroscedasticity empirically relevant, we estimate it as follows. From the bucket-wise medians of the high minus low yield spreads that we observe in the US zero-coupon STRIPS data[4], we compute the spread $s_j = \text{Med}\left(\{h_{i,j} - l_{i,j}\}_{i=1}^{n_j}\right)$, where h_j and l_j denote the high and the low yields, respectively, and n_j is the total number of observations within each bucket ranging from $(\tau_{j-1}, \tau_j]$, $j = 1, \ldots, M$, with $\tau_0 = 0$. Because these spreads give a very rough picture, the standard deviation σ_j, $j = 1, \ldots, M$, used for the simulations are obtained from applying a running median with window width five over these 200 buckets. The median is used for robustness reasons. Figure 1 reports the total number of observations (left panel) and the estimated standard deviation (right panel) as a function of the maturity buckets. Starting with the smallest tenors, the median spread gradually decreases and is

[4]See Section 4 for a more detailed description of these data.

lowest for about 20-24 years (buckets 130 to 160). For the largest tenors, the estimated volatility rises again.

For the simulations, we consider three sets of parameters to generate different shapes of the yield curve: (1) a flat curve parametrized with $\beta_0 = 0.04, \beta_1 = \beta_2 = 0, \vartheta = 1$; (2) an increasing curve with $\beta_0 = 0.05, \beta_1 = -0.03, \beta_2 = 0, \vartheta = 2$; and (3) a humped-shaped curve with $\beta_0 = 0.05, \beta_1 = 0, \beta_2 = -0.1359, \vartheta = 2$. The humped-shaped curve produces an almost flat discount factor curve for $\tau \approx 2$ years. A non-constrained estimator is therefore likely to produce inadmissible estimates in this case. For each of the three parameters' settings, 1,000 random samples are drawn and subjected to the different estimators.

In the literature, there is consensus about the usefulness of constrained estimators, but less is known about the relative importance of the degrees of the spline and the loss functions used for constrained estimation. This is because a single constrained estimator is usually compared against unconstrained fits; see Barzanti and Corradi (1999); Ramponi (2003); Chiu et al. (2008); Laurini and Moura (2010). The general estimation framework that we suggest allows us to obtain a more complete picture of the impact of the B-spline order, the penalty, and the loss function on estimation efficiency. We therefore consider three orders of the B-spline, $q \in \{2, 3, 4\}$, and two penalties $\mathbf{P}_R$ and $\mathbf{P}_{EM1}$ for both the L_1 and the L_2 estimators. We use the constraints spelled out in (14). Together with the original COBS estimator for the discount curve as suggested by Laurini and Moura (2010), this yields thirteen estimators for comparison purposes.

3.2 Simulation results

In Table 1, we display the knot sequences and the penalty parameters underlying the estimations. For our estimators, they are obtained from a knot search as described in Section 2.3 on one randomly chosen sample. The boundary knots are set to 0.00 and 60.00. We choose a right boundary knot this far to the right for two reasons. First, it allows us to better approximate the $\lim_{\tau\to\infty} d(\tau) = 0$ behavior; second, this choice stabilizes the estimations of the derivative that we need to compute the forward rate curve; see Fengler and Hin (2013) for a detailed discussion of this issue. To avoid additional randomizations, the knots are kept fixed for all 1,000 runs. The COBS implementation, as used in Laurini and Moura (2010), does not actively search for optimal knot sites, but sets knots equidistantly in percentiles and deletes redundant knots according to the SIC. As discussed in Section 2.4, we can therefore attribute any differences between our L_1 estimator for $q = 2$ with ridge penalty $\mathbf{P}_R$ and the COBS estimator to the different knot placement strategies.

As can be seen from Table 1, our knot search suggests richer parametrized models (more knots) for the L_1 estimator than for the L_2 estimator. As regards the two penalties, there are, at most, minimal differences, because the influence of the penalty term is masked in the knot search phase. We also do not find a lot of variation across the different orders of the B-spline, with the exception of Setting 1, which is the case with the flat yield curve. The selected penalty parameters tend to be larger for the L_1 estimator, which may be due to the larger number of knots selected for this estimator.

For each estimator, Tables 2, 3 and 4 report the integrated mean squared error (IMSE), the integrated squared bias (IBias^2), and the integrated variance (IVar) for each curve, i.e.,

$$\text{IMSE} \equiv \int \mathrm{E}[\{\hat{m}(\tau) - m_0(\tau)\}^2 \,|\tau] d\tau = \underbrace{\int \{\mathrm{E}[\hat{m}(\tau)|\tau] - m_0(\tau)\}^2 \, d\tau}_{\text{IBias}^2} + \underbrace{\int \mathrm{Var}[\hat{m}(\tau)|\tau] d\tau}_{\text{IVar}} , \qquad (25)$$

where $m(\cdot) \in \{d(\cdot), r(\cdot), f(\cdot)\}$ are the discount factor curve, the spot rate and the forward rate curve, respectively. We also compute the integrated mean absolute deviation about the true function (IMAD), which we decompose into a bias and a dispersion component, namely, the integrated absolute deviation of the median about the true function (IBias-med) and the integrated mean absolute deviation of the estimates about the median (IMAD-med). For these, it holds that

$$\text{IMAD} \equiv \int \mathrm{E}\left[\left|\hat{m}(\tau) - m_0(\tau)\right| |\tau\right] d\tau \qquad (26)$$

$$\leq \underbrace{\int \left|\text{Med}[\hat{m}(\tau)|\tau] - m_0(\tau)\right| d\tau}_{\text{IBias-med}} + \underbrace{\int \mathrm{E}\left[\left|\hat{m}(\tau) - \text{Med}[\hat{m}(\tau)|\tau]\right| |\tau\right] d\tau}_{\text{IMAD-med}} . \qquad (27)$$

Table 2 shows that in our simulation environment, where the error distributions are inferred from the daily high-low spreads, the largest contributions to both the IMSE and the IMAD come from the bias components. The contributions of dispersion components are small. For Setting 1, the flat yield curve, all estimators produce almost the same fitting quality, independently of the orders of the underlying B-splines. This is because the knot search successfully eliminates redundant knots, and thus, avoids excessive variation of the higher order B-splines. Moreover, the results do not exhibit any particular difference between COBS and the other estimators, which shows that exact knot placement is irrelevant for the flat yield curve. This is in contrast to Settings 2 and 3. For the increasing and the humped-shaped yield curve, good knot placement matters. This corroborates the findings in Ramponi (2003). Moreover, the order of the spline is of importance. The best fits are achieved for the splines of order three and four. It is worth noting that, for our simulations, we observe neither systematic differences between the L_1 and the L_2 estimators, nor between the different penalty versions. In Tables 3 and 4, we additionally evaluate the estimates on the spot and the forward rate curve. The nonlinear transformations tend to amplify the differences among the estimators, yet overall, the previous conclusions are confirmed.

4 Empirical applications

We work with the US zero-coupon STRIPS data of the Thomson Reuters Tick History as supplied by the Securities Industry Research Centre of Asia-Pacific. STRIPS disentangle the coupon and principal components of eligible US Treasury notes and bonds. They can be held as separate securities. Because a payment is only received at maturity, we can treat them as zero coupon bonds. The data set consists of the daily yields of both the ‘interest rate only’ and the ‘principal only’ components as observed between January 2, 2001 and December 31, 2009. For each record day, the day high, low, open and close yield in percentage terms, along with the corresponding date of maturity, are available.

Because the yields are reported on an ISMA Act/Act 6M YTM basis, we first convert them into annualized continuous spot rates by means of $r = \log\left[\{1 + (Y/100)/2\}^2\right]$, where Y denotes the reported day close yield. We then compute the discount factors respecting the Act/Act day count convention, and all estimators are applied using the constraints in (14). The selections of knots and penalty parameters are performed on each sample date.

The results are summarized in Table 5. As in the simulations, the differences between the different penalty versions, on the one hand, and between the L_1 and the L_2 estimators, on the other hand, are not dramatic. Yet there appears to be a clear preference for the more flexible splines of order four. Comparing the fits of the COBS model with the L_1 fit at order $q = 2$ with ridge penalty, one again discerns the benefits of an active knot search.

In Figure 2, we display selected fits from the sample. While the differences in the discount factor curves are hardly visible, the spot rate curves demonstrate the versatility of the higher order fits vis-à-vis the linear fit with fixed knots. This is particularly striking for the hump-shaped fits as in Figure 2. The log-transformation produces unacceptable kinks in the resulting spot rate curve, although no derivative is needed for its computation. In summary, cubic fits are not only superior in terms of their statistical fitting quality, but also more appealing on simple visual and practical considerations.

5 Concluding remarks

We describe a simple and general framework for fitting the discount curve based on penalized B-splines. The fits are obtained under no-arbitrage constraints, i.e., the discount curve obeys $d(0) = 1$, and is monotonically decreasing and positive. The framework is independent of the order of the selected B-splines. In our implementations, we use linear, quadratic and cubic splines on simulated data and on a large data set of historical US STRIPS, and estimate under an L_1 and an L_2 loss function. We also discuss an active knot search for optimal knot placement.

Our results suggest (*i*) that the quadratic and cubic splines are superior to concurrent linear shape-constrained fits; (*ii*) that a good knot placement is an important parametrization device to capture the salient features of the discount curve; (*iii*) that both the ridge and the first-order difference penalty deliver comparable results; and (*iv*) that differences between L_1 and L_2 estimates are not dramatic when applying the penalized B-spline to highly liquid interest rate data as we do. Clearly,

L_1 estimation could be more relevant when working with more volatile data, for instance, the interest rate data of emerging markets.

References

Barzanti, L. and Corradi, C. (1999). A note on direct term structure estimation using monotonic splines. *Rivista di Matematica per le Scienze Economiche e Sociali*, 22(10):101–108.

Cheng, H., Fang, S.-C., and Lavery, J. E. (2005). A geometric programming framework for univariate cubic L_1 smoothing splines. *Annals of Operations Research*, 133:229–248.

Chiu, N. C., Fang, S. C., Lavery, J. E., Lin, J. Y., and Wang, Y. (2008). Approximating term structure of interest rates using cubic L_1 splines. *European Journal of Operational Research*, 184:990–1004.

Claeskens, G., Krivobokova, T., and Opsomer, J. O. (2009). Asymptotic properties of penalized spline estimators. *Biometrika*, 96(3):529–544.

Coleman, T. S., Fisher, L., and G., I. R. (1992). Estimating the term structure of interest rates from data that include the prices of coupon bonds. *Journal of Fixed Income*, 2(2):85–116.

de Boor, C. (2001). *A practical guide to splines*. Springer-Verlag, New York, revised edition.

Eilers, P. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2):89–102.

Fengler, M. R. and Hin, L.-Y. (2013). Semi-nonparametric estimation of the call-option price surface under strike and time-to-expiry no-arbitrage constraints. Technical report.

Fredenhagen, S., Oberle, H. J., and Opfer, G. (1999). On the construction of optimal monotone cubic spline interpolations. *Journal of Approximation Theory*, 96(2):182–201.

Gao, X. and Fang, Y. (2011). A note on the generalized degrees of freedom under the L_1 loss function. *Journal of Statistical Planning and Inference*, 141:677–686.

Hagan, P. and Graeme, W. (2006). Interpolation methods for curve construction. *Applied Mathematical Finance*, 13(2):89–129.

Hastie, T. and Tibshirani, R. (1990). *Generalized additive models*. Chapman and Hall, London.

He, X. and Ng, P. (1999). COBS: Qualitatively constrained smoothing via linear programming. *Computational Statistics*, 14(3):315–337.

He, X. and Shi, P. (1996). Bivariate tensor-product B-splines in a partly linear model. *Journal of Multivariate Analysis*, 58(2):162–181.

He, X. and Shi, P. (1998). Monotone *B*-spline smoothing. *Journal of the American Statistical Association*, 93(442):643–650.

Ioannides, M. (2003). A comparison of yield curve estimation techniques using UK data. *Journal of Banking & Finance*, 27:1–26.

Kelly, C. and Rice, J. (1991). Monotone smoothing with application to dose-response curves and the assessment of synergism. *Biometrics*, 46:1071–1085.

Koenker, R. (2013). *quantreg: Quantile Regression*. R package version 5.05.

Koenker, R. and Ng, P. (2005). Inequality constrained quantile regression. *Sankhȳa: The Indian Journal of Statistics*, 67(2):418–440.

Laurini, M. and Moura, M. (2010). Constrained smoothing B-splines for the term structure of interest rates. *Insurance: Mathematics and Economics*, 46:339–350.

Lavery, J. E. (2000). Shape preserving, multiscale fitting of univariate data by cubic L_1 smoothing splines. *Computer Aided Geometric Design*, 17:715–727.

Litzenberger, R. and Rolfo, R. (1984). An international study of tax effects on government bonds. *Journal of Finance*, 39:1–22.

Marangio, L., Bernaschi, M., and Ramponi, A. (2002). A review of techniques for the estimation of the term structure. *International Journal of Theoretical and Applied Finance*, 5(2):189–221.

McCulloch, J. (1971). Measuring the term structure of interest rates. *Journal of Business*, 44:19–31.

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

Meyer, M. C. (2012). Constrained penalized splines. *Canadian Journal of Statistics*, 40(1):190–206.

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

Ng, P. and Maechler, M. (2007). A fast and efficient implementation of qualitatively constrained quantile smoothing splines. *Statistical Modelling*, 7(4):315–328.

Papp, D. and Alizadeh, F. (2014). Shape constrained estimation using nonnegative splines. *Journal of Computational and Graphical Statistics*, 23(1):211–231.

Preve, D. and Medeiros, M. C. (2011). Linear programming-based estimators in simple linear regression. *Journal of Econometrics*, 165:128–136.

Ramponi, A. (2003). Adaptive and monotone spline estimation of the cross-sectional term structure. *International Journal of Theoretical and Applied Finance*, 6(2):195–212.

Ramsay, J. O. (1988). Monotone regression splines in action. *Statistical Science*, 3(4):425–461.

Tobler, J. (1996). *Schätzung von Zinsstrukturen für den SFr.-Kapitalmarkt unter Berücksichtigung von Friktionen*. Verlag Paul Haupt, Bern, Stuttgart, Wien.

Turlach, B. A. (2005). Shape constrained smoothing using smoothing splines. *Computational Statistics*, 20(1):81–104.

Turlach, B. A. and Weingessel, A. (2014). *Package quadprog*. R package version 1.5-5.

Vasicek, O. A. and Fong, H. G. (1982). Term structure modeling using exponential splines. *Journal of Finance*, 37(2):339–348.

Yallup, P. (2012). Models of the yield curve and the curvature of the implied forward rate function. *Journal of Banking & Finance*, 36:121–135.

Zhou, S. and Shen, X. (2001). Spatially adaptive regression splines and accurate knot selection schemes. *Journal of the American Statistical Association*, 96(453):247–259.

Tables and Figures

Table 1: **Knot sequences and penalty parameters.**

Estimator	Knots	λ
	Setting 1: $\beta_0 = 0.04, \beta_1 = 0, \beta_2 = 0, \vartheta = 1.$	
COBS	0.15, 6.00, 12.00, 18.00, 24.00, 30.00	0.4257
$L_1, q = 2, \mathbf{P}_{EM1}$	0.00, 0.96, 4.28, 7.61, 9.99, 15.61, 20.65, 26.55, 29.45, 38.70, 60 .00	0.0592
$L_1, q = 2, \mathbf{P}_R$	0.00, 1.14, 4.29, 7.61, 10.01, 15.72, 21.30, 51.22, 60.00	0.0903
$L_1, q = 3, \mathbf{P}_{EM1}$	0.00, 1.91, 2.69, 6.13, 8.66, 9.62, 18.90, 21.29, 28.55, 31.16, 42.46, 60.00	0.0753
$L_1, q = 3, \mathbf{P}_R$	0.00, 1.91, 2.69, 6.14, 8.65, 9.63, 11.06, 21.52, 32.29, 60	0.0919
$L_1, q = 4, \mathbf{P}_{EM1}$	0.00, 10.04, 28.88, 29.60, 34.52, 60.00	0.0810
$L_1, q = 4, \mathbf{P}_R$	0.00, 9.50, 47.21, 60.00	0.0705
$L_2, q = 2, \mathbf{P}_{EM1}$	0.00, 1.27, 4.32, 9.18, 12.75, 15.6, 20.85, 60.00	0.0001
$L_2, q = 2, \mathbf{P}_R$	0.00, 1.27, 4.32, 9.18, 12.75, 15.6, 20.85, 60.00	0.0001
$L_2, q = 3, \mathbf{P}_{EM1}$	0.00, 1.30, 2.67, 5.97, 8.62, 9.49, 15.6, 24.48, 60.00	0.0090
$L_2, q = 3, \mathbf{P}_R$	0.00, 1.30, 2.67, 5.97, 8.62, 9.49, 15.6, 24.48, 60.00	0.0003
$L_2, q = 4, \mathbf{P}_{EM1}$	0.00, 7.03, 46.99, 60.00	0.0009
$L_2, q = 4, \mathbf{P}_R$	0.00, 7.03, 46.99, 60.00	0.0016
	Setting 2: $\beta_0 = 0.05, \beta_1 = -0.03, \beta_2 = 0, \vartheta = 2.$	
COBS	0.15, 6.00, 12.00, 18.00, 24.00, 30.00	0.6623
$L_1, q = 2, \mathbf{P}_{EM1}$	0.00, 1.00, 5.34, 7.84, 12.79, 17.16, 23.4, 35.03, 57.11, 60.00	0.0955
$L_1, q = 2, \mathbf{P}_R$	0.00, 1.00, 5.34, 7.84, 12.79, 17.16, 23.32, 51.95, 57.07, 60.00	0.0904
$L_1, q = 3, \mathbf{P}_{EM1}$	0.00, 2.23, 4.06, 7.8, 13.28, 15.11, 23.26, 34.92, 57.11, 60.00	0.0818
$L_1, q = 3, \mathbf{P}_R$	0.00, 2.23, 4.06, 7.8, 13.28, 15.1, 23.28, 36.6, 57.07, 60.00	0.0803
$L_1, q = 4, \mathbf{P}_{EM1}$	0.00, 3.28, 6.57, 7.8, 12.85, 17.16, 22.75, 36.59, 57.11, 60.00	0.0852
$L_1, q = 4, \mathbf{P}_R$	0.00, 3.28, 6.57, 7.8, 12.83, 17.16, 22.76, 36.6, 57.07, 60.00	0.0510
$L_2, q = 2, \mathbf{P}_{EM1}$	0.00, 0.79, 5.75, 14.65, 59.58, 60.00	0.0001
$L_2, q = 2, \mathbf{P}_R$	0.00, 0.79, 5.75, 14.65, 60.00	0.0010
$L_2, q = 3, \mathbf{P}_{EM1}$	0.00, 1.24, 6.58, 17.98, 50.81, 60.00	0.0100
$L_2, q = 3, \mathbf{P}_R$	0.00, 1.24, 6.58, 17.97, 50.79, 60.00	0.0008
$L_2, q = 4, \mathbf{P}_{EM1}$	0.00, 7.24, 15.05, 57.02, 60.00	0.0034
$L_2, q = 4, \mathbf{P}_R$	0.00, 7.22, 15.06, 57.00, 60.00	0.0009
	Setting 3: $\beta_0 = 0.05, \beta_1 = 0, \beta_2 = -0.1359, \vartheta = 2.$	
COBS	0.15, 6.00, 12.00, 18.00, 24.00, 30.00	0.1203
$L_1, q = 2, \mathbf{P}_{EM1}$	0.00, 0.69, 3.85, 5.18, 14.36, 20.27, 23.65, 33.69, 60.00	0.0978
$L_1, q = 2, \mathbf{P}_R$	0.00, 0.69, 3.85, 5.18, 14.36, 20.27, 23.7, 38.32, 45.29, 60.00	0.0932
$L_1, q = 3, \mathbf{P}_{EM1}$	0.00, 0.78, 3.65, 7.23, 20.35, 22.3, 24.03, 37.37, 45.62, 60.00	0.0890
$L_1, q = 3, \mathbf{P}_R$	0.00, 0.78, 3.65, 7.23, 20.44, 22.3, 24.03, 40.11, 45.27, 60.00	0.0752
$L_1, q = 4, \mathbf{P}_{EM1}$	0.00, 2.80, 4.65, 7.21, 12.79, 17.11, 23.79, 37.42, 45.62, 60.00	0.0867
$L_1, q = 4, \mathbf{P}_R$	0.00, 2.60, 4.80, 7.17, 13.01, 17.33, 24.29, 35.37, 45.29, 60.00	0.0487
$L_2, q = 2, \mathbf{P}_{EM1}$	0.00, 0.65, 3.47, 4.87, 14.29, 22.39, 60.00	0.0001
$L_2, q = 2, \mathbf{P}_R$	0.00, 0.65, 3.47, 4.87, 14.29, 22.39, 60.00	0.0003
$L_2, q = 3, \mathbf{P}_{EM1}$	0.00, 1.05, 3.86, 6.97, 23.85, 60.00	0.0082
$L_2, q = 3, \mathbf{P}_R$	0.00, 1.05, 3.86, 6.97, 23.85, 60.00	0.0006
$L_2, q = 4, \mathbf{P}_{EM1}$	0.00, 3.47, 3.86, 6.97, 12.47, 45.65, 60.00	0.0002
$L_2, q = 4, \mathbf{P}_R$	0.00, 3.47, 3.86, 6.97, 12.47, 45.65, 60.00	0.0010

Table 2: **Summary statistics for discount factors.** IMSE: integrated mean squared error. IBias^2: integrated squared bias. IVar: integrated variance. IMAD: integrated mean absolute deviation of the sample estimates about the true value. IBias-med: integrated absolute deviation of the median about the true value. IMAD-med: integrated mean absolute deviation of the sample estimates about the median.

Estimator	IMSE ($\times 10^{-1}$)	IBias^2 ($\times 10^{-1}$)	IVar ($\times 10^{-5}$)	IMAD	IBias-med	IMAD-med ($\times 10^{-2}$)
Setting 1: $\beta_0 = 0.04, \beta_1 = 0, \beta_2 = 0, \vartheta = 1.$						
COBS	1.1993	1.1992	0.8463	1.1277	1.1278	0.8563
$L_1, q = 2, \mathbf{P}_{EM1}$	1.1987	1.1986	1.0524	1.1206	1.1208	0.9360
$L_1, q = 2, \mathbf{P}_R$	1.1997	1.1997	0.8019	1.1207	1.1208	0.8330
$L_1, q = 3, \mathbf{P}_{EM1}$	1.1986	1.1985	1.0388	1.1204	1.1205	0.9370
$L_1, q = 3, \mathbf{P}_R$	1.1986	1.1986	0.9086	1.1203	1.1205	0.8759
$L_1, q = 4, \mathbf{P}_{EM1}$	1.1990	1.1989	0.5049	1.1205	1.1207	0.6482
$L_1, q = 4, \mathbf{P}_R$	1.1990	1.1989	0.4233	1.1206	1.1206	0.5955
$L_2, q = 2, \mathbf{P}_{EM1}$	1.2002	1.2001	0.5395	1.1210	1.1209	0.6818
$L_2, q = 2, \mathbf{P}_R$	1.2001	1.2001	0.5395	1.1210	1.1209	0.6819
$L_2, q = 3, \mathbf{P}_{EM1}$	1.2001	1.2000	0.5252	1.1209	1.1209	0.6775
$L_2, q = 3, \mathbf{P}_R$	1.1996	1.1996	0.5732	1.1208	1.1208	0.7076
$L_2, q = 4, \mathbf{P}_{EM1}$	1.1994	1.1994	0.2627	1.1204	1.1204	0.4741
$L_2, q = 4, \mathbf{P}_R$	1.1988	1.1987	0.2867	1.1202	1.1201	0.4894
Setting 2: $\beta_0 = 0.05, \beta_1 = -0.03, \beta_2 = 0, \vartheta = 2.$						
COBS	1.6052	1.6052	0.6380	1.2983	1.2983	0.7415
$L_1, q = 2, \mathbf{P}_{EM1}$	1.6038	1.6038	0.6773	1.2921	1.2921	0.7694
$L_1, q = 2, \mathbf{P}_R$	1.6035	1.6035	0.6719	1.2920	1.2920	0.7664
$L_1, q = 3, \mathbf{P}_{EM1}$	1.6033	1.6032	0.9175	1.2920	1.2920	0.8905
$L_1, q = 3, \mathbf{P}_R$	1.6032	1.6031	0.6900	1.2918	1.2920	0.7844
$L_1, q = 4, \mathbf{P}_{EM1}$	1.6030	1.6030	0.7064	1.2917	1.2917	0.7923
$L_1, q = 4, \mathbf{P}_R$	1.6032	1.6031	0.7192	1.2917	1.2917	0.8000
$L_2, q = 2, \mathbf{P}_{EM1}$	1.6109	1.6108	0.2246	1.2925	1.2924	0.4495
$L_2, q = 2, \mathbf{P}_R$	1.6107	1.6107	0.2245	1.2923	1.2923	0.4494
$L_2, q = 3, \mathbf{P}_{EM1}$	1.6050	1.6050	0.2169	1.2923	1.2922	0.4461
$L_2, q = 3, \mathbf{P}_R$	1.6039	1.6039	0.2154	1.2921	1.2920	0.4447
$L_2, q = 4, \mathbf{P}_{EM1}$	1.6041	1.6041	0.2644	1.2945	1.2944	0.4877
$L_2, q = 4, \mathbf{P}_R$	1.6036	1.6035	0.2457	1.2943	1.2942	0.4658
Setting 3: $\beta_0 = 0.05, \beta_1 = 0, \beta_2 = -0.1359, \vartheta = 2.$						
COBS	1.0309	1.0308	1.1179	0.9698	0.9698	0.9867
$L_1, q = 2, \mathbf{P}_{EM1}$	1.0299	1.0298	1.0242	0.9509	0.9507	0.9491
$L_1, q = 2, \mathbf{P}_R$	1.0296	1.0295	1.0166	0.9507	0.9505	0.9462
$L_1, q = 3, \mathbf{P}_{EM1}$	1.0286	1.0285	1.0565	0.9524	0.9523	0.9644
$L_1, q = 3, \mathbf{P}_R$	1.0290	1.0289	1.0619	0.9525	0.9524	0.9640
$L_1, q = 4, \mathbf{P}_{EM1}$	1.0289	1.0288	1.0983	0.9535	0.9535	0.9972
$L_1, q = 4, \mathbf{P}_R$	1.0290	1.0289	1.1275	0.9530	0.9529	1.0101
$L_2, q = 2, \mathbf{P}_{EM1}$	1.0306	1.0305	0.4859	0.9506	0.9505	0.6548
$L_2, q = 2, \mathbf{P}_R$	1.0305	1.0305	0.4858	0.9506	0.9504	0.6549
$L_2, q = 3, \mathbf{P}_{EM1}$	1.0305	1.0305	0.4054	0.9517	0.9516	0.5985
$L_2, q = 3, \mathbf{P}_R$	1.0296	1.0296	0.4693	0.9516	0.9514	0.6419
$L_2, q = 4, \mathbf{P}_{EM1}$	1.0297	1.0296	0.5141	0.9554	0.9553	0.6737
$L_2, q = 4, \mathbf{P}_R$	1.0294	1.0294	0.4741	0.9553	0.9552	0.6501

Table 3: **Summary statistics for spot rates.** IMSE: integrated mean squared error. IBias^2: integrated squared bias. IVar: integrated variance. IMAD: integrated mean absolute deviation of the sample estimates about the true value. IBias-med: integrated absolute deviation of the median about the true value. IMAD-med: integrated mean absolute deviation of the sample estimates about the median.

Estimator	IMSE ($\times 10^{-3}$)	IBias^2 ($\times 10^{-3}$)	IVar ($\times 10^{-7}$)	IMAD ($\times 10^{-1}$)	IBias-med ($\times 10^{-1}$)	IMAD-med ($\times 10^{-3}$)
			Setting 1: $\beta_0 = 0.04, \beta_1 = 0, \beta_2 = 0, \vartheta = 1$			
COBS	1.2493	1.2491	1.8275	1.2775	1.2777	1.2001
$L_1, q = 2, \mathbf{P}_{EM1}$	1.2062	1.2059	2.5932	1.2146	1.2143	1.3426
$L_1, q = 2, \mathbf{P}_R$	1.2064	1.2062	1.8810	1.2130	1.2128	1.1868
$L_1, q = 3, \mathbf{P}_{EM1}$	1.2054	1.2052	2.0128	1.2129	1.2130	1.2737
$L_1, q = 3, \mathbf{P}_R$	1.2053	1.2051	2.0446	1.2127	1.2129	1.2421
$L_1, q = 4, \mathbf{P}_{EM1}$	1.2055	1.2055	0.8976	1.2136	1.2138	0.8634
$L_1, q = 4, \mathbf{P}_R$	1.2056	1.2056	0.7859	1.2134	1.2134	0.8036
$L_2, q = 2, \mathbf{P}_{EM1}$	1.2071	1.2070	1.2525	1.2134	1.2132	0.9681
$L_2, q = 2, \mathbf{P}_R$	1.2070	1.2069	1.2525	1.2133	1.2131	0.9681
$L_2, q = 3, \mathbf{P}_{EM1}$	1.2064	1.2062	1.6039	1.2132	1.2130	1.0283
$L_2, q = 3, \mathbf{P}_R$	1.2061	1.2060	1.7245	1.2131	1.2128	1.0643
$L_2, q = 4, \mathbf{P}_{EM1}$	1.2062	1.2061	0.7838	1.2076	1.2074	0.7113
$L_2, q = 4, \mathbf{P}_R$	1.2058	1.2057	0.9155	1.2072	1.2070	0.7456
			Setting 2: $\beta_0 = 0.05, \beta_1 = -0.03, \beta_2 = 0, \vartheta = 2.$			
COBS	2.0719	2.0717	1.2777	1.5878	1.5878	1.0406
$L_1, q = 2, \mathbf{P}_{EM1}$	2.0191	2.0189	2.0404	1.5334	1.5334	1.1892
$L_1, q = 2, \mathbf{P}_R$	2.0188	2.0186	2.0345	1.5332	1.5332	1.1869
$L_1, q = 3, \mathbf{P}_{EM1}$	2.0184	2.0175	8.9916	1.5326	1.5316	1.5970
$L_1, q = 3, \mathbf{P}_R$	2.0178	2.0176	1.9820	1.5323	1.5323	1.2025
$L_1, q = 4, \mathbf{P}_{EM1}$	2.0171	2.0169	1.8894	1.5293	1.5291	1.1924
$L_1, q = 4, \mathbf{P}_R$	2.0171	2.0169	1.9043	1.5293	1.5291	1.1997
$L_2, q = 2, \mathbf{P}_{EM1}$	2.0329	2.0328	1.2207	1.5406	1.5406	0.7697
$L_2, q = 2, \mathbf{P}_R$	2.0323	2.0321	1.2187	1.5397	1.5396	0.7695
$L_2, q = 3, \mathbf{P}_{EM1}$	2.0320	2.0319	1.4165	1.5452	1.5450	0.7856
$L_2, q = 3, \mathbf{P}_R$	2.0318	2.0316	1.4761	1.5456	1.5455	0.7908
$L_2, q = 4, \mathbf{P}_{EM1}$	2.0319	2.0318	0.9358	1.5556	1.5555	0.7822
$L_2, q = 4, \mathbf{P}_R$	2.0315	2.0315	0.8767	1.5556	1.5554	0.7513
			Setting 3: $\beta_0 = 0.05, \beta_1 = 0, \beta_2 = -0.1359, \vartheta = 2.$			
COBS	1.1679	1.1676	2.3740	1.0738	1.0736	1.3303
$L_1, q = 2, \mathbf{P}_{EM1}$	0.9064	0.9060	3.9779	0.9320	0.9314	1.4308
$L_1, q = 2, \mathbf{P}_R$	0.9062	0.9058	3.9173	0.9318	0.9312	1.4233
$L_1, q = 3, \mathbf{P}_{EM1}$	0.9113	0.9108	5.5606	0.9455	0.9453	1.4982
$L_1, q = 3, \mathbf{P}_R$	0.9117	0.9112	5.2582	0.9456	0.9455	1.4759
$L_1, q = 4, \mathbf{P}_{EM1}$	0.9504	0.9500	4.2224	0.9702	0.9699	1.5175
$L_1, q = 4, \mathbf{P}_R$	0.9382	0.9378	4.0074	0.9626	0.9623	1.5046
$L_2, q = 2, \mathbf{P}_{EM1}$	0.9048	0.9046	1.5172	0.9300	0.9295	0.9435
$L_2, q = 2, \mathbf{P}_R$	0.9047	0.9045	1.5311	0.9298	0.9293	0.9459
$L_2, q = 3, \mathbf{P}_{EM1}$	0.8998	0.8997	1.6184	0.9261	0.9255	0.8994
$L_2, q = 3, \mathbf{P}_R$	0.8992	0.8990	1.7338	0.9260	0.9253	0.9440
$L_2, q = 4, \mathbf{P}_{EM1}$	0.9709	0.9708	1.5033	0.9827	0.9826	0.9768
$L_2, q = 4, \mathbf{P}_R$	0.9695	0.9694	1.4527	0.9821	0.9819	0.9527

Table 4: **Summary statistics for forward rates.** IMSE: integrated mean squared error. IBias^2: integrated squared bias. IVar: integrated variance. IMAD: integrated mean absolute deviation of the sample estimates about the true value. IBias-med: integrated absolute deviation of the median about the true value. IMAD-med: integrated mean absolute deviation of the sample estimates about the median.

Estimator	IMSE ($\times 10^{-2}$)	IBias^2 ($\times 10^{-3}$)	IVar ($\times 10^{-5}$)	IMAD ($\times 10^{-1}$)	IBias-med ($\times 10^{-1}$)	IMAD-med ($\times 10^{-2}$)
Setting 1: $\beta_0 = 0.04, \beta_1 = 0, \beta_2 = 0, \vartheta = 1$						
COBS	0.3460	3.4566	0.3239	2.0649	2.0651	0.5194
$L_1, q = 2, \mathbf{P}_{EM1}$	0.3529	3.4681	6.1392	2.0758	2.0784	1.1022
$L_1, q = 2, \mathbf{P}_R$	0.3332	3.3282	0.3825	2.0400	2.0404	0.5717
$L_1, q = 3, \mathbf{P}_{EM1}$	0.3435	3.4086	2.6816	2.0656	2.0825	0.9244
$L_1, q = 3, \mathbf{P}_R$	0.3360	3.3517	0.8182	2.0527	2.0524	0.6968
$L_1, q = 4, \mathbf{P}_{EM1}$	0.3363	3.3538	0.8820	2.0534	2.0595	0.3287
$L_1, q = 4, \mathbf{P}_R$	0.3318	3.3167	0.0953	2.0446	2.0447	0.2271
$L_2, q = 2, \mathbf{P}_{EM1}$	0.3295	3.2918	0.2928	2.0271	2.0270	0.4891
$L_2, q = 2, \mathbf{P}_R$	0.3294	3.2911	0.2928	2.0270	2.0269	0.4891
$L_2, q = 3, \mathbf{P}_{EM1}$	0.3373	3.3703	0.3095	2.0575	2.0586	0.4570
$L_2, q = 3, \mathbf{P}_R$	0.3336	3.3315	0.4277	2.0483	2.0517	0.5318
$L_2, q = 4, \mathbf{P}_{EM1}$	0.3382	3.3818	0.0393	2.0619	2.0615	0.1534
$L_2, q = 4, \mathbf{P}_R$	0.3366	3.3654	0.0477	2.0583	2.0584	0.1629
Setting 2: $\beta_0 = 0.05, \beta_1 = -0.03, \beta_2 = 0, \vartheta = 2.$						
COBS	6.2838	6.2810	0.2815	2.8044	2.8053	4.7481
$L_1, q = 2, \mathbf{P}_{EM1}$	6.1182	6.1144	0.3816	2.7337	2.7342	5.8257
$L_1, q = 2, \mathbf{P}_R$	6.1159	6.1121	0.3767	2.7336	2.7343	5.7835
$L_1, q = 3, \mathbf{P}_{EM1}$	6.1674	6.1531	1.4265	2.7460	2.7442	9.1344
$L_1, q = 3, \mathbf{P}_R$	6.1094	6.1044	0.4992	2.7368	2.7370	6.0393
$L_1, q = 4, \mathbf{P}_{EM1}$	6.1646	6.1599	0.4694	2.7440	2.7451	6.0305
$L_1, q = 4, \mathbf{P}_R$	6.1049	6.0993	0.5647	2.7335	2.7324	6.5505
$L_2, q = 2, \mathbf{P}_{EM1}$	5.8250	5.8247	0.0362	2.6434	2.6432	1.7602
$L_2, q = 2, \mathbf{P}_R$	5.8244	5.8241	0.0361	2.6431	2.6429	1.7597
$L_2, q = 3, \mathbf{P}_{EM1}$	6.0953	6.0950	0.0331	2.7555	2.7557	1.5973
$L_2, q = 3, \mathbf{P}_R$	6.0871	6.0867	0.0316	2.7544	2.7543	1.5581
$L_2, q = 4, \mathbf{P}_{EM1}$	6.2288	6.2282	0.0614	2.7782	2.7784	2.0687
$L_2, q = 4, \mathbf{P}_R$	6.1677	6.1670	0.0722	2.7682	2.7677	2.0070
Setting 3: $\beta_0 = 0.05, \beta_1 = 0, \beta_2 = -0.1359, \vartheta = 2.$						
COBS	4.7101	4.7065	0.3600	2.3436	2.3448	5.5033
$L_1, q = 2, \mathbf{P}_{EM1}$	4.4404	4.4347	0.5657	2.2219	2.2219	6.4729
$L_1, q = 2, \mathbf{P}_R$	4.4219	4.4163	0.5622	2.2183	2.2185	6.4594
$L_1, q = 3, \mathbf{P}_{EM1}$	4.5062	4.4964	0.9828	2.2179	2.2197	7.1716
$L_1, q = 3, \mathbf{P}_R$	4.4089	4.3984	1.0520	2.2000	2.2005	7.2991
$L_1, q = 4, \mathbf{P}_{EM1}$	4.4585	4.4529	0.5595	2.2048	2.2061	6.3470
$L_1, q = 4, \mathbf{P}_R$	4.3794	4.3713	0.8048	2.1828	2.1818	7.2223
$L_2, q = 2, \mathbf{P}_{EM1}$	4.5038	4.5023	0.1538	2.2346	2.2340	3.4021
$L_2, q = 2, \mathbf{P}_R$	4.5026	4.5011	0.1540	2.2345	2.2339	3.4040
$L_2, q = 3, \mathbf{P}_{EM1}$	4.4062	4.4050	0.1175	2.1827	2.1851	2.3812
$L_2, q = 3, \mathbf{P}_R$	4.3351	4.3329	0.2252	2.1687	2.1652	3.3669
$L_2, q = 4, \mathbf{P}_{EM1}$	4.4501	4.4484	0.1627	2.2227	2.2214	3.2598
$L_2, q = 4, \mathbf{P}_R$	4.4323	4.4311	0.1223	2.2200	2.2194	2.9336

Table 5: **Goodness-of-fit on US STRIPS data.** Reported are the time series means of the mean squared error (MSE), the median absolute deviation (MAD) and the median absolute deviation (MedAD) of the fits on US STRIPS data from January 2, 2001 to December 31, 2009.

	Discount factor			Spot rate		
	MSE (10^{-5})	MAD (10^{-3})	MedAD (10^{-3})	MSE (10^{-6})	MAD (10^{-4})	MedAD (10^{-4})
COBS	4.330	3.557	2.944	11.607	14.743	12.748
$L_1, q = 2, \mathbf{P}_R$	3.099	2.359	1.655	4.402	7.350	5.224
$L_1, q = 2, \mathbf{P}_{EM1}$	3.074	2.353	1.650	4.330	7.332	5.245
$L_1, q = 3, \mathbf{P}_R$	2.705	2.045	1.394	3.370	6.510	4.977
$L_1, q = 3, \mathbf{P}_{EM1}$	2.797	2.066	1.394	3.541	6.590	4.975
$L_1, q = 4, \mathbf{P}_R$	2.260	1.847	1.293	2.104	5.329	3.866
$L_1, q = 4, \mathbf{P}_{EM1}$	2.261	1.853	1.292	2.118	5.344	3.865
$L_2, q = 2, \mathbf{P}_R$	2.580	2.586	2.187	2.580	7.038	5.884
$L_2, q = 2, \mathbf{P}_{EM1}$	2.582	2.589	2.191	2.576	7.038	5.888
$L_2, q = 3, \mathbf{P}_R$	2.176	2.064	1.549	2.174	6.158	5.299
$L_2, q = 3, \mathbf{P}_{EM1}$	2.178	2.068	1.552	2.179	6.174	5.318
$L_2, q = 4, \mathbf{P}_R$	2.195	2.000	1.454	2.019	5.705	4.232
$L_2, q = 4, \mathbf{P}_{EM1}$	2.199	2.008	1.456	1.987	5.689	4.231

Figure 1: **Patterns of heteroscedasticity for Monte Carlo simulations**. Panel A shows the number of spread observations pooled in each bucket. Panel B exhibits the pattern of heteroscedasticity as estimated from bucket-specific yield spreads, which are computed from daily high-low yield differences. Source: US STRIPS data 2001-2009.

Panel A: **Panel B**:

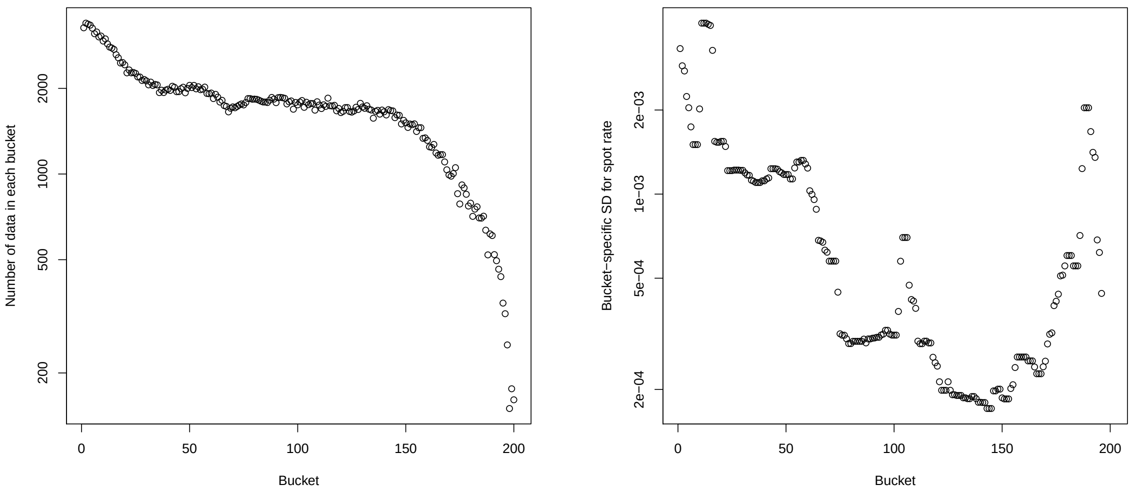
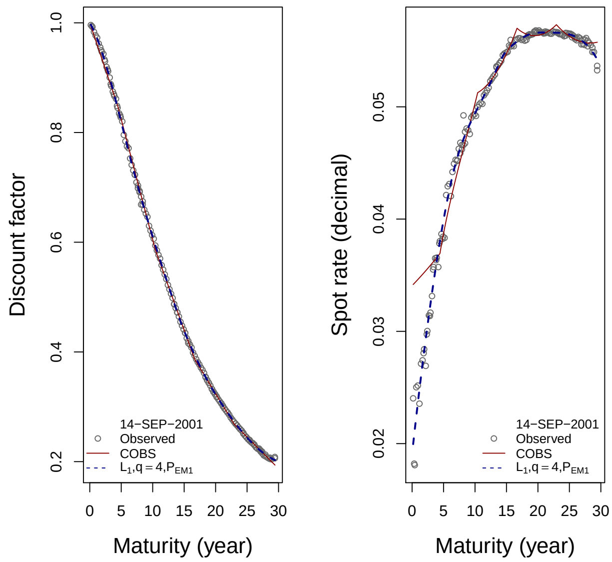
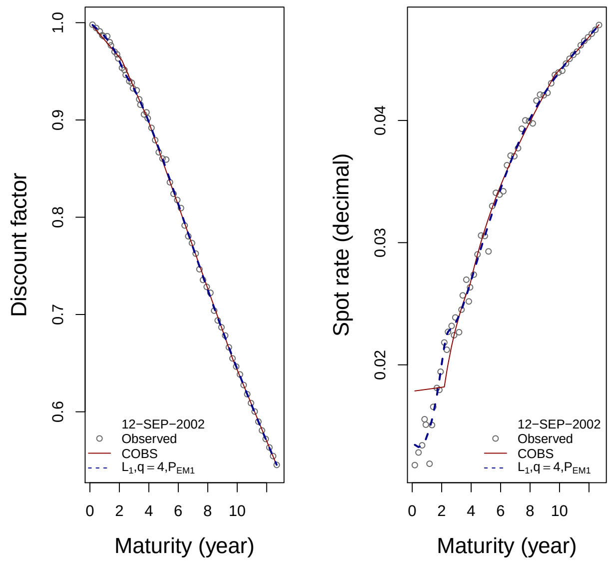


Figure 2: **Selected Fits**. The fitted discount and spot rate factor curves are plotted for selected sample days. Source: US STRIPS data 2001-2009.

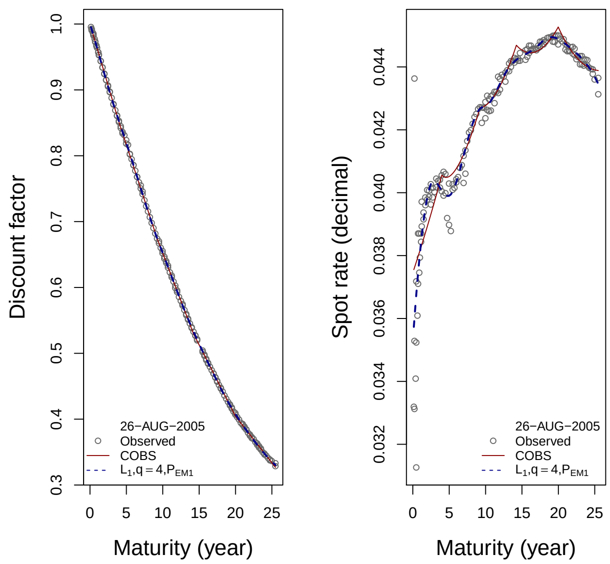
Panel A:



Panel B:



Panel C:



Panel D: