COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1205

ESTIMATING YIELD CURVES BY KERNEL SMOOTHING METHODS

O. Linton, E. Mammen, J. Nielsen and C. Tanggaard

December 1998

ESTIMATING YIELD CURVES BY KERNEL SMOOTHING METHODS

O. Linton,[1] E. Mammen,[2] J. Nielsen,[3] and C. Tanggaard[4]

January 5, 1999

Abstract

We introduce a new method for the estimation of discount functions, yield curves and forward curves from government issued coupon bonds. Our approach is nonparametric and does not assume a particular functional form for the discount function although we do show how to impose various restrictions in the estimation. Our method is based on kernel smoothing and is defined as the minimum of some localized population moment condition. The solution to the sample problem is not explicit and our estimation procedure is iterative, rather like the backfitting method of estimating additive nonparametric models. We establish the asymptotic normality of our methods using the asymptotic representation of our estimator as an infinite series with declining coefficients. The rate of convergence is standard for one dimensional nonparametric regression.

Key words: Coupon bonds; Forward curve; Hilbert Space; Local linear; Nonparametric regression; Yield curve.

[0 1]Cowles Foundation for Research in Economics, Yale University, 30 Hillhouse Avenue, New Haven, CT 06520–8281, USA. Phone: (203) 432–3699. Fax: (203) 432–6167. Email: linton@econ.yale.edu. Research was supported by the National Science Foundation and the North Atlantic Treaty Organization. Thanks to Roger Farmer, Andrew Jeffrey, Thong Nguyen, Peter Phillips, and John Rust for helpful comments. A computer program in Gauss is available for doing the computations.

1 Introduction

The term structure of interest rates is a central concept in monetary and financial economics. Prices of fixed income securities like bonds, swaps, and mortgage backed bonds (MBB's) are functions of the yield curve, and pricing of derivatives also depends on the yield curve. The spread between long and short term interest rates carries information about the level of future interest rates, see for example Campbell and Shiller (1991) and Engsted and Tanggaard (1995). The slope of the yield curve has frequently been used in empirical studies as a predictor of future inflation and national incomes, see Frankel and Lown (1994) and Estrella and Mishkin (1998) for example. Therefore, estimation of yield curves has had a long tradition among financial researchers and practitioners. See Campbell, Lo, and Mackinlay (1997) for further discussion.

A fundamental problem is that the yield to maturity on coupon bonds are not directly comparable between bonds with different maturities or coupons. Thus, there is a need for a standardized way of measuring the term structure of interest rates. One such standard is the yield curve of zero-coupon bonds issued by sovereign lenders.

The construction of this yield curve poses several problems for applied research. First, many governments do not issue longer term (i.e., greater than 1-2 years) zero-coupon bonds. Hence the yield curve must be inferred from other instruments. A simple solution can be derived from the law of one price by assuming absence of arbitrage. Arbitrage in the bond market will cause the price p of any bond (coupon or zero) with payments $b(\tau_j)$ at time τ_j to be equal to the discounted value of the future cash flow $\pi = \sum_{j=1}^{m} b(\tau_j)d(\tau_j)$, where the discount factor is $d(\tau_j)$ at time τ_j. The future income stream, $b(\tau_1), \ldots, b(\tau_m)$, is assumed known and non-random. The second problem is that, in practice, small pricing errors perhaps due to non-synchronous trading, taxation, illiquidity,

[2]Institute für Angewandte Mathematik, Ruprecht–Karls–Universität Heidelberg, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany. Supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 373 "Quantifikation und Simulation Ökonomischer Prozesse", Humboldt-Universität zu Berlin and Project MA1026/6-1.

[3]Codanhus, 60 Gammel Kongevej, DK-1790 Copenhagen V, Denmark. Email: npj@codan.dk

[4] Department of Finance, The Aarhus School of Business, Fuglesangs Alle 4, DK-8210 Aarhus V, Denmark. E-mail: cat@hha.dk. Research is supported by the Danish Social Science Research Council through the Center for Analytical Finance, CAF. A computer program in Pascal is available for doing the computations.

and bid-ask spreads necessitates adding an error term to π. The error term should be sufficiently small to ensure that they do not represent (gross) violations of the law of one price (no-arbitrage condition). For example, Amihud and Mendelson (1991) find that close to maturity notes and T-bills with identical payment streams can be priced differently, although when taking account of the various transactions costs, there were not, on average, arbitrage opportunities.

The statistical problem we address is one of finding the function $d(\cdot)$ that minimizes the deviation between observed prices in the market and the observed present value. The estimation is done using a carefully selected sample of bonds or any instrument (swaps) with known cash flow. Note that, based on a continuous time approximation, we have $d(t) = \exp(-ty(t))$, where $y(t)$ is the yield curve, and $y(t) = \exp(-\int_0^t f(s)ds)$, where $f(t)$ is the forward curve, see Anderson, Breedon, Deacon, Derry, and Murphy (1996, pp 12-13). Both of these relationships are invertible, so that knowing d is equivalent to knowing y or f. We shall suppose that the discount function $d(\cdot)$ is a continuous and indeed smooth function of time to maturity. Although this is not guaranteed by purely arbitrage arguments, it does seem plausible.

Following the seminal work of McCulloch (1971), the standard approach to estimation here is to assume a parametric specification for $d(t)$ or $y(t)$ or $f(t)$ and to use linear or nonlinear least squares to estimate the unknown parameters. For example, McCulloch (1971, 1975), Shea (1984) use regression splines for $d(t)$, Chambers, Carleton, and Waldman (1984) uses polynomials for $y(t)$, Vasicek and Fong (1982) uses exponential splines for $d(t)$, while Nelson and Siegel (1988) proposes powers of exponentials for $y(t)$. An approach based on linear programming methods has been suggested by Schaefer (1981). If the specification is considered parametric, i.e., to be a complete and correct representation of the functions of interest, i.e., the mean, then standard asymptotic theory can be used to derive the limiting distribution of the estimator and to justify confidence intervals obtained from this. However, a number of these authors are arguing against adherence to any fixed model and really are viewing the problem as being nonparametric. Some recent studies by Fisher, Nychka, and Zervos (1995) and Tanggaard (1997) have also taken this line. When we view the estimation problem as nonparametric, there is little existing theory regarding the distribution of the estimators; for example, no-one has established the asymptotic distribution of the spline estimates discussed above.

We adopt a nonparametric approach in which we do not a priori specify the functional form of the discount function or forward curve. In fact, we base our procedures on perhaps the most central of all smoothing methods, the kernel method. The flexibility of our method is very important in practical applications because parametric estimates are often flawed by specification biases. It is not immediately obvious how to estimate the function $d(\cdot)$ by kernel methods, since this function affects the mean function indirectly through a convolution with the payment function. We solve the estimation problem by defining the function d as the solution of some population mean squared error criterion, and then localizing the sample version of this to obtain an empirical criterion function. Several versions of the localization are possible including local constant and local linear [which has some well known advantages in other contexts, see Tsybakov (1986) and Fan (1992)]. It turns out that our methods do not generally have explicit solutions, i.e., our estimator is defined as the solution of a linear integral equation. In practice, our solutions are defined through the method of successive approximations, but see Rust (1997) for discussion of some alternative methods. We also give a 'backfitting' interpretation to our procedure, as in Opsomer and Ruppert (1997) and Linton, Mammen and Nielsen (1998). We establish the convergence of our iterative scheme and establish the asymptotic properties of the estimator. We use the method of successive approximations to write our estimator as an infinite series with declining coefficients and thereby obtain its asymptotic distribution $-$ it is asymptotically normal at the standard rate of convergence for one-dimensional kernel regression. The asymptotic distribution of the implied estimators of $y(t)$ and $f(t)$ can be easily obtained by the delta method. Our regularity conditions are high level, but we show how they are satisfied in some leading cases. We also exploit the relationships with the yield curve and forward curve to suggest alternative methods, thus we write $d(\cdot) = \psi(\theta(\cdot))$ for some known function ψ, making θ now the object of estimation. The purpose of this is to give some added flexibility and/or to enforce consistency with theory. For example, by taking $d(t) = \exp(-ty(t))$ we can directly impose the restrictions that $d(0) = 1$ and $d(t) > 0$ for all t, at the same time we are directly estimating the yield curve itself. We apply our methods to a sample of US Treasury bonds.

We point out that the estimation problem is similar to that considered in Engle, Granger, Rice, and Weiss (1986) in which electricity demand over a billing period is modelled as a sum of individual daily demands each determined by temperature on the day concerned. They used splines, which effectively parameterizes the function d and makes the estimation problem standard nonlinear regression. They did not provide any asymptotic theory to justify their approach, at least not for the pointwise distribution of the nonparametric part. A similar estimation problem occurs quite widely with grouped data. For example, Chesher (1997) estimates the individual nutrient intake-age relationship from household level intake and individual characteristics like age. Again, he used splines but did not provide any justification for the validity of his method. A related problem arises in nonparametric simultaneous equations [Newey and Powell (1988)] and in estimating solutions of integral equations [Wahba (1979), Nychka, Wahba, Goldfarb, and Pugh (1984), and O'Sullivan (1986)]. See also Hausman and Newey (1995) for a related problem involving differential equations. Finally, see Horowitz (1998) for a discussion of other deconvolution problems.

In section 2 we discuss [for reasons of completeness] smoothing of pure discount bounds. In section 3 we present our new methods for smoothing the yield curve. In subsection 3.1 we present the local constant version of our estimate with ψ the identity, i.e., the object of interest is the discount function, while subsection 3.2 gives the local linear extension. Section 3.3 describes how to estimate y and f from the estimate of d. We then present in subsections 3.4-3.6 the exponential version of these methods. We present the asymptotic properties of our methods in section 4 and apply the method to data in section 5. In section 6 we describe extensions to testing parametric hypotheses and to estimating semiparametric models. Proofs are given in the appendix.

2 Smoothing of Pure Discount Bond Yields

In an efficient bond market, \$1 delivered at some future date τ has one, and only one, price, d. The function $d(\tau)$, giving the discount factor as a function of τ, is called the discount function. In practice small pricing errors due to rounding-off, tax effects, and inefficiencies distort the pricing mechanism. Therefore, the discount function must be derived from a sample of noisy zero coupon bond prices, $p_1, \ldots, p_n$ with times to maturity $\tau_1 \leq \ldots \leq \tau_n$. We shall not require that the times be equally spaced. The statistical model is based on random perturbations of the present value pricing relationship, i.e.,

$$p_i = b_i(\tau_i)d(\tau_i) + \varepsilon_i,\ \ i = 1, \ldots, n \tag{1}$$

where ε_i constitute an independent random sequence with $E[\varepsilon_i] = 0$, $i = 1, \ldots, n$, where $b_i(\tau_i)$ is the payment, i.e., the principal, returned to the bond investor when bond i matures at date τ_i. Note that $b_i(\tau_i)$ and τ_i are known and fixed constants, and that the statistical problem is one of estimating the discount function $d(\tau)$. The discount function can be defined as that smooth function $d(\cdot)$ which minimizes the population criterion function

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E[\{p_i - b_i(\tau_i)\theta(\tau_i)\}^2]$$

with respect to $\theta(\cdot)$. The limiting criterion function exists and has a unique minimum provided that $\{\tau_i\}$ is dense in a compact interval and that the first two moments of p_i are finite. This suggests the following localized sample based criterion function

$$Q_n(\theta) = \sum_{i=1}^{n} W_i(\tau) \{p_i - b_i(\tau_i)\theta\}^2\,, \tag{2}$$

where $\{W_i\}$ is a set of smoothing weights, depending only on the $\tau' s$, which are small for τ_i distant from τ, i.e., the criterion is weighting preferentially bonds with similar maturities to τ. Let $\widehat{d}(\cdot)$ minimize $Q_n(\theta)$ with respect to $\theta(\cdot)$. The solution to the first order condition is given by

$$\widehat{d}(\tau) = \frac{\sum_{i=1}^{n} W_i(\tau) b_i(\tau_i) p_i}{\sum_{i=1}^{n} W_i(\tau) b_i^2(\tau_i)}. \tag{3}$$

In kernel smoothing, the weight sequence is derived from a kernel function. A kernel function is a symmetric, continuously bounded real function, K, that integrates to one. The weights are then determined by the formula $W_i(\tau) = K_h(\tau - \tau_i)$, where $K_h(u) = K(u/h)/h$, where $K(\cdot)$ is a density function and $h = h(n)$ is a sequence of positive numbers. The bandwidth parameter, h, determines the degree of smoothing. Small values make the estimated curve $\widehat{d}(\tau)$ very approximate and irregular, while large values of h make the estimate close to the sample average. This estimation method was suggested in Tanggaard (1992). The asymptotic distribution of $\widehat{d}(\tau)$ is covered by Theorems 1 and 2 of Gozalo and Linton (1998).

There are two aspects of the estimation procedure. First, there is the smoothing aspect. Smoothing is necessary because prices are noisy. The noise can originate from minor market imperfections, non-synchronous trading, price-discreteness (tick size), temporary imbalances in demand and supply. Furthermore, tax-effects (clienteles), and illiquidity premia may also affect the present value relationship. The need for smoothing can be reduced by a careful sample selection procedure. However, in general it is not possible to completely avoid the need for smoothing. The second aspect is interpolation. Interpolation becomes necessary when markets are not complete. In the US market, for example, T-notes are issued with original time to maturity 1,2,3,5, 10, and 20 years. Thus, if we want to avoid using off-the-run issues in estimation, there is a clear need for interpolation of the yield curve between 5 and 10 years. And in general we do not want to plot the yields as a scatter plot, which further necessitates a smooth graphical picture of the yield curve.

However, a proper interpolation of yields between maturities where no payments are due, requires a formal model of the dynamics of the term structure of interest rates. This is not the objective of this paper, and we therefore take a purely statistical approach and smooth the yield curve using kernel smoothing.

3 Smoothing of Coupon Bonds

As discussed in section 2, most bond markets do not posses zero coupon bonds for useful spans of maturities. The discount function must therefore be extracted from quotations of coupon bond data instead. Coupon bonds generate several payments at future dates, and in an efficient bond market, the present value of these future payments should, apart from a small error, be equal to the trading price, p_i. As in section 2, the sample consists of n bonds with quoted prices $p_1, \ldots, p_n$. Furthermore, $b_i(\tau_{ij}) \neq 0$ denotes the payment returned to the owner of bond i at date τ_{ij}, where $\tau_{i1} < \ldots < \tau_{im_i}$ are the possible payment dates. For United States treasury issued notes and bonds, the payments are usually equal semiannual coupons c_i, $j = 1, \ldots, m_{i-1}$, while in the final period one receives the redemption value R_i [usually this is normalized to be 100] plus the final coupon c_i.[1] The model considered in section 2 corresponds to the special case where $m_i = 1$. In general, m_i will be larger than one and will vary from bond to bond. Frequently, however, some maturities will coincide, so that the total number, say m, of distinct times $\{\tau_{ij}\}$ will lie somewhere between $\max_{1\leq i\leq n} m_i$ and $\sum_{i=1}^{n} m_i$. The statistical model we adopt is

$$p_i = \sum_{j=1}^{m_i} b_i(\tau_{ij})d(\tau_{ij}) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{4}$$

where ε_i is an independent random sequence satisfying $E[\varepsilon_i] = 0$, $i = 1, \ldots, n$. The statistical problem is to extract estimates of the unknown discount function $d(\cdot)$ based on a sample of observed bond prices, coupon payments, and times.[2] We can write this relation as $p = Bd + \varepsilon$, where p, ε are the $n \times 1$ vectors of prices and errors respectively, while B is an $n \times m$ matrix containing the payments, and d is the $m \times 1$ vector of discount factors. This would suggest estimating the unknown vector d by regression techniques such as least squares. However, the number m can be very large [$m >> n$] in a typical sample of bonds, because there is little overlap in payment times. Furthermore, for each bond the payments are usually the same until the last period, i.e., $b_i(\tau_{ij}) = b_i(\tau_{ik})$ for all j, k with $1 \leq j, k < m_i$. This tends to make the matrix B rank deficient and preclude the direct use of ordinary least squares to estimate d. The problem is really that the finite sample least squares criterion function does not impose smoothness on the function d [this problem arises generically in standard nonparametric regression and, indeed, in the no-coupon bond case]. We can interpret the function $d(\cdot)$ in (4) as being the continuous minimizer [with respect to $\theta(\cdot)$] of the limiting criterion function

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E\left\{p_i - \sum_{j=1}^{m_i} b_i(\tau_{ij})\theta(\tau_{ij})\right\}^2, \tag{5}$$

[1] Sometimes there are short or long first and last coupons, meaning that these coupons may be larger or smaller than the other coupons. Other bond markets have similar payment schemes with annual and semi-annual payments being the norm.

[2] Note that the Engle, Granger, Rice and Weiss (1986) model allowed the regression function to be observed only through a linear functional but also included some parametric effects.

which exists provided the sequence $\{\tau_{ij}\}$ becomes dense in a compact set. We choose an alternative finite sample criterion which imposes smoothness throughout its approach to infinity. One simple method is to group the data into bins which have similar τ_{ij}, and then to do least squares on this grouped data. This amounts to a histogram approach to estimation. We shall use a kernel version of this procedure, which improves on the poor bias of this method.

3.1 Local Constant Smoothing

Our trick is to generalize an interpretation of a smoother in terms of a global projection. We first outline this interpretation in the pure discount bond setting and then see how to apply it to the general case (4). Consider the following global criterion function

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\int K_h(t-\tau_i)\left\{p_i - b_i(\tau_i)\theta(t)\right\}^2 dt, \tag{6}$$

and let $\widehat{\theta}(\cdot)$ minimize this criterion with respect to functions $\theta(\cdot)$. The solution to this calculus of variations problem can be obtained from the necessary condition for $\widehat{\theta}(\cdot)$ to be a minimizer, see Weinstock (1951, pp 20-22), which is that the Gateaux derivative of Q_n with respect to θ must be zero in all directions. To find the solution, however, it is convenient just to look in the direction of point masses. Let δ_τ be the Dirac [generalized] function at τ, that is, $\int \delta_\tau(u)f(u)du = f(\tau)$ for any function f which is continuous at τ, see Lighthill (1958). Now replace $\theta(\cdot)$ by $\widehat{\theta}(\cdot) + \epsilon\delta_\tau(\cdot)$, where ϵ is a real number, and differentiate the criterion function with respect to ϵ at $\epsilon = 0$. We obtain

$$\sum_{i=1}^{n}\int K_h(t-\tau_i)b_i(\tau_i)\delta_\tau(t)\{p_i - b_i(\tau_i)\widehat{\theta}(t)\}dt = 0.$$

Now use the fact that $\int \delta_\tau(t)K_h(t-\tau_i)dt = K_h(\tau-\tau_i)$ and $\int \delta_\tau(t)\widehat{\theta}(t)K_h(t-\tau_i)dt = \widehat{\theta}(\tau)K_h(\tau - X_i)$ to see that $\widehat{\theta}(\tau) = \sum_{i=1}^{n} p_i b_i(\tau_i)K_h(\tau-\tau_i)/\sum_{i=1}^{n} b_i^2(\tau_i)K_h(\tau-\tau_i)$, which is in fact the kernel version of (3). In this case, we already arrived at $\widehat{\theta}(\tau)$ as the minimizer of a local criterion function and the global criterion function (6) just provides a nice interpretation.[3] In our case, however, it is essential

[3] See Nielsen and Linton (1998) for some discussion of a similar example where global criterion functions provide interpretation.

to start with the global criterion function.

We define the following localized sample criterion function

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\int\left\{p_i - \sum_{j=1}^{m_i} b_i(\tau_{ij})\theta(s_{ij})\right\}^2 \prod_{l=1}^{m_i}\left\{K_h(s_{il}-\tau_{il})\ ds_{il}\right\}, \tag{7}$$

where for each $\{s_{il}\}$, the local criterion function

$$\sum_{i=1}^{n}\left\{p_i - \sum_{j=1}^{m_i} b_i(\tau_{ij})\theta(s_{ij})\right\}^2 \prod_{l=1}^{m_i} K_h(s_{il}-\tau_{il})$$

is preferentially weighting bonds whose maturities $\tau_{i1},\ldots,\tau_{im_i}$ are close to the sequence $\{s_{il}\}$. Although the local criterion does impose the requirement that $\theta(\cdot)$ is constant in a small neighborhood of s_{ij}, unfortunately it does not provide sensible estimates because we have to allow the evaluation points of θ to vary with both i and j, so that there is not enough restriction imposed at a single point. The averaging in (7) imposes the necessary structure.[4] We define our estimator $\widehat{d}_{LC}(\cdot)$ as the minimizer of $Q_n(\theta)$ with respect to $\theta(\cdot)$. We discuss existence of a solution below, but we next show how to solve the minimization in (7) using the method introduced above. Replace $\theta(\cdot)$ in (7) by $\widehat{d}_{LC}(\cdot)+\epsilon\delta_s(\cdot)$, where ϵ is a real number, and differentiate the right hand side of (7) with respect to ϵ at $\epsilon=0$. This gives the following first order condition

$$\begin{aligned}
0 &= \sum_{i=1}^{n}\sum_{r=1}^{m_i}\int\left\{p_i - \sum_{j=1}^{m_i}\mathbf{1}[j\neq r]b_i(\tau_{ij})\widehat{d}_{LC}(s_{ij}) \ - b_i(\tau_{ir})\widehat{d}_{LC}(s)\right\} \\
&\quad b_i(\tau_{ir})\prod_{j=1}^{m_i}\mathbf{1}[j\neq r]\left\{K_h(s_{ij}-\tau_{ij})\ ds_{ij}\right\}K_h(s-\tau_{ir}) \\
&= \sum_{i=1}^{n}\sum_{r=1}^{m_i} p_i b_i(\tau_{ir})K_h(s-\tau_{ir}) - \sum_{i=1}^{n}\sum_{r=1}^{m_i}\sum_{\substack{j=1\\ j\neq r}}^{m_i} b_i(\tau_{ij})b_i(\tau_{ir})K_h(s-\tau_{ir})\int \widehat{d}_{LC}(t)K_h(t-\tau_{ij})\ dt
\end{aligned} \tag{8}$$

[4]Note also that by a change of variable

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\int\left\{p_i - \sum_{j=1}^{m_i} b_i(\tau_{ij})\theta(\tau_{ij}+hu_{ij})\right\}^2 \prod_{l=1}^{m_i}\left\{K(u_{il})du_{il}\right\},$$

which is asymptotically equivalent to (5).

$$-\sum_{i=1}^{n}\sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(s-\tau_{ir})\widehat{d}_{LC}(s),$$

which has used the fact that $\int K_h(s_{ij}-\tau_{ij})ds_{ij}=1$ to eliminate many of the integrals. This can be written as

$$\widehat{d}_{LC}(s)=\overline{d}(s)+\int \widehat{H}_{LC}(s,t)\widehat{d}_{LC}(t)dt, \tag{9}$$

where

$$\overline{d}(s)=\frac{\sum_{i=1}^{n}\sum_{r=1}^{m_i} p_i b_i(\tau_{ir})K_h(s-\tau_{ir})}{\sum_{i=1}^{n}\sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(s-\tau_{ir})}, \tag{10}$$

$$\widehat{H}_{LC}(s,t)=-\frac{\sum_{i=1}^{n}\sum_{r=1}^{m_i}\sum_{j\neq r}^{m_i} b_i(\tau_{ir})b_i(\tau_{ij})K_h(s-\tau_{ir})K_h(t-\tau_{ij})}{\sum_{i=1}^{n}\sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(s-\tau_{ir})}, \tag{11}$$

Relation (9) suggests the following iteration for the calculation of $\widehat{d}_{LC}$:

$$\widehat{d}_{LC}^{[a+1]}(s)=\overline{d}(s)+\int \widehat{H}_{LC}(s,t)\widehat{d}_{LC}^{[a]}(t)dt, \quad a=0,1,\ldots, \tag{12}$$

where the starting value $\widehat{d}_{LC}^{[0]}$ is equal to $\overline{d}$, say. The integration in (12) can be computed numerically. We discuss this further in the application section below, see also Linton, Mammen, and Nielsen (1997) for a related computation. The quantity $\overline{d}(s)$ can be thought of as the minimizer with respect to the scalar θ of the criterion function

$$\sum_{i=1}^{n}\sum_{r=1}^{m_i}\{p_i-b_i(\tau_{ir})\theta\}^2 K_h(s-\tau_{ir}). \tag{13}$$

This function corresponds to a sum of zero-coupon bond criteria. Therefore, an alternative interpretation of our algorithm is that at each stage we are applying the smoother defined by (13) with p_i replaced by the partial residuals

$$\widehat{p}_{ir}^{[a]}=p_i-\sum_{\substack{j=1\\ j\neq r}}^{m_i} b_i(\tau_{ij})\int \widehat{d}_{LC}^{[a]}(t)K_h(t-\tau_{ij})dt.$$

This gives our algorithm a backfitting interpretation [see Hastie and Tibshirani (1990)] in which the basic smoothing operation is given by (13).

A necessary condition that the iterative calculation $\widehat{d}_{LC}$ converges can be based on a check of the operator norm of the operator $\widehat{\mathcal{H}}_{LC} g(\cdot) = \int \widehat{H}_{LC}(\cdot, t) g(t)\, dt$. For a norm $\|\cdot\|$ on the functions g the corresponding norm of the operator $\widehat{\mathcal{H}}_{LC}$ is defined as $\|\widehat{\mathcal{H}}_{LC}\| = \sup_{\|g\|=1} \|\widehat{\mathcal{H}}_{LC} g\|$. If $\|g\|$ is defined as the supremum norm $\|g\|_\infty = \sup_x |g(x)|$, we get that $\|\widehat{\mathcal{H}}_{LC}\|_\infty = \sup_s \int \widehat{H}_{LC}(s,t)\, dt$. If $\|g\|$ is defined as the L_2 norm $\|g\|_2^2 = \int |g(x)|^2\, dx$, then $\|\widehat{\mathcal{H}}_{LC}\|_2$ can be calculated by an iterative algorithm $\|\widehat{\mathcal{H}}_{LC}\|_2^2 = \lim_{k\to\infty} \|\widehat{\mathcal{H}}_{LC} e_k\|_2^2$, where the function e_k is defined as $e_k = \widehat{\mathcal{H}}_{LC} e_{k-1} / \|\widehat{\mathcal{H}}_{LC} e_{k-1}\|_2$. Under reasonable conditions, the function $\overline{d}$ has a bounded sup norm and a bounded L_2 norm. Furthermore, it follows from a Neumann expansion that the linear transformation $I - \widehat{\mathcal{H}}_{LC}$ has the inverse $(I - \widehat{\mathcal{H}}_{LC})^{-1} = \sum_{k=0}^{\infty} \widehat{\mathcal{H}}_{LC}^k$, provided $\|\widehat{\mathcal{H}}_{LC}\| < 1$ with respect to either the sup norm or to the L_2 norm (see e.g., Riesz and Sz.-Nagy (1990), p. 152). When $\|\widehat{\mathcal{H}}_{LC}\|_2 < 1$ or $\|\widehat{\mathcal{H}}_{LC}\|_\infty < 1$, it therefore follows that the iterative calculation of $\widehat{d}_{LC}$ converges and that the solution is

$$\widehat{d} = (I - \widehat{\mathcal{H}}_{LC})^{-1} \overline{d} = \sum_{k=0}^{\infty} \widehat{\mathcal{H}}_{LC}^k \overline{d}.$$

The resulting estimate $\widehat{d}_{LC}$ can be interpreted as least squares projection of the data in an appropriate function space. For this purpose we consider the following norm on n tuples of functions $g = (g_1, \ldots, g_n)$, namely

$$\|g\|^2 = \sum_{i=1}^{n} \int \left\{ \sum_{j=1}^{m_i} b_i(\tau_{ij}) g_i(s_{ij}) \right\}^2 \prod_{j=1}^{m_i} \left\{ K_h(s_{ij} - \tau_{ij})\, ds_{ij} \right\}. \tag{14}$$

If one puts $\eta_i(\cdot)$ as equal to $p_i[\sum_{j=1}^{m} b_i(\tau_{ij})]^{-1}$ [which is a constant function], then the function tuple $(\widehat{d}_{LC}, \ldots, \widehat{d}_{LC})$ is the projection of $\eta = (\eta_1, \ldots, \eta_n)$ onto the linear subspace $\{g = (g_1, \ldots, g_n) : g_1 = \ldots = g_n\}$ with respect to the norm $\|\cdot\|$. This relation helps to understand that $\widehat{d}_{LC}$ is always well defined (up to differences that have norm 0 with respect to $\|\cdot\|$). In particular, for the unique existence of $\widehat{d}_{LC}$ it is not required that $\|\widehat{\mathcal{H}}_{LC}\| < 1$. Furthermore, our approach allows us to incorporate restrictions on the shape of $\overline{d}$. For example, a natural constraint is to suppose that d is monotone. A natural estimate based on kernel smoothing is defined as minimizer of (7) where the minimization runs now over the constrained class of functions [e.g., over all monotone functions $\overline{d}$.] Let us denote the resulting estimate by $\widehat{\overline{d}}_{constr}$. Then it is easy to see that $\widehat{\overline{d}}_{constr}$ is the projection of

$\widehat{\overline{d}}$ onto the constrained class with respect to the norm $\|\cdot\|$ defined in (14). This means that

$$\widehat{d}_{constr} = \arg\min_{\overline{d}} \sum_{i=1}^{n} \int \left\{ \sum_{j=1}^{m_i} b_i(\tau_{ij})[\overline{d}(s_{ij}) - \widehat{d}(s_{ij})] \right\}^2 \prod_{j=1}^{m_i} \{K_h(s_{ij} - \tau_{ij})\ ds_{ij}\},$$

where the minimization runs over the constrained class of functions d. See Mammen (1991) and Matzkin (1994) for discussion of nonparametric estimation of monotone and other constrained functions.

Finally, it can be expected that ε_i will be heteroskedastic, since bonds with a long time to maturity can be affected by many sources of errors. Consequently, we expect the variance of ε_i, denoted σ_i^2, will vary proportionately with τ_{im_i}. In this case a weighted criterion function taking account of the different accuracies of each bond would perhaps lead to more efficient estimates, at least in terms of variance. Suppose that $\sigma_i = \gamma f(\tau_{im_i})$ for some known function f and unknown parameter γ. Then, the weighted criterion function is given by (5) where we redefine $p_i \to p_i/f(\tau_{im_i})$ and $b_i(\tau_{ij}) \to b_i(\tau_{ij})/f(\tau_{im_i})$. Therefore, estimation proceeds as above with the transformed data.

3.2 Local linear smoothing

Kernel smoothing leads to estimates with design dependent bias and with poor accuracy at boundaries, see for example Tsybakov (1986) and Fan (1992). An approach that is known to overcome these disadvantages in regression is local linear smoothing. For the definition of the local linear smooth in our model we just replace $\theta(\cdot)$ in (7) by a linear function. Thus, consider the following tuple $(\widehat{d}_{LL}, \widehat{d}_{LL,1})$ that minimizes

$$\sum_{i=1}^{n} \int \left\{ p_i - \sum_{j=1}^{m_i} b_i(\tau_{ij})[d(s_{ij}) + (\tau_{ij} - s_{ij})d_1(s_{ij})] \right\}^2 \prod_{j=1}^{m_i} \{K_h(s_{ij} - \tau_{ij})\ ds_{ij}\} \tag{15}$$

with respect to (d, d_1). The local linear smooth is now defined as $\widehat{d}_{LL}$. Again, the minimizer of (15) can be easily calculated by an iterative algorithm. To solve the minimization in (15) we proceed as in the last section and put $d = \widehat{d}_{LL} + \epsilon\delta_s$ where ϵ is a real number and where δ_s is the Dirac function in s. Differentiation of the right hand side with respect to ϵ gives (at $\epsilon = 0$):

$$
\begin{aligned}
0 &= \sum_{i=1}^{n}\sum_{r=1}^{m_i}\int\Bigg\{p_i - \sum_{j=1}^{m_i}\mathbf{1}[j\neq r]b_i(\tau_{ij})\widehat{d}_{LL}(s_{ij}) \;- b_i(\tau_{ir})\widehat{d}_{LL}(s) \;- b_i(\tau_{ir})\widehat{d}_{LL,1}(s)(\tau_{ir}-s) \\
&\quad - \sum_{j=1}^{m_i}\mathbf{1}[j\neq r]b_i(\tau_{ij})\widehat{d}_{LL,1}(s_{ij})(\tau_{ij}-s_{ij})\Bigg\} \\
&\quad \times b_i(\tau_{ir})\prod_{j=1}^{m_i}\mathbf{1}[j\neq r]\left\{K_h(s_{ij}-\tau_{ij})\;ds_j\right\}K_h(s-\tau_{ir}) \\
&= \sum_{i=1}^{n}\sum_{r=1}^{m_i}p_i b_i(\tau_{ir})K_h(s-\tau_{ir}) - \sum_{i=1}^{n}\sum_{r=1}^{m_i}b_i(\tau_{ir})^2K_h(s-\tau_{ir})\widehat{d}_{LL}(s) \\
&\quad - \sum_{i=1}^{n}\sum_{r=1}^{m_i}\sum_{\substack{j=1\\ j\neq r}}^{m_i}b_i(\tau_{ij})b_i(\tau_{ir})K_h(s-\tau_{ir})\int\widehat{d}_{LL}(t)K_h(t-\tau_{ij})\;dt \\
&\quad - \sum_{i=1}^{n}\sum_{r=1}^{m_i}\sum_{\substack{j=1\\ j\neq r}}^{m_i}b_i(\tau_{ij})b_i(\tau_{ir})K_h(s-\tau_{ir})\int\widehat{d}_{LL,1}(t)(\tau_{ij}-t)K_h(t-\tau_{ij})\;dt \\
&\quad - \sum_{i=1}^{n}\sum_{r=1}^{m_i}b_i(\tau_r)^2K_h(s-\tau_{ir})\widehat{d}_{LL,1}(s)(\tau_{ir}-s).
\end{aligned}
$$

Doing the same for $d_1 = \widehat{d}_{LL,1} + \epsilon\delta_s$ gives another equation:

$$
\begin{aligned}
0 &= \sum_{i=1}^{n}\sum_{r=1}^{m_i}\int\Bigg\{p_i - \sum_{j=1}^{m_i}\mathbf{1}[j\neq r]b_i(\tau_{ij})\widehat{d}_{LL}(s_j) \;- b_i(\tau_{ir})\widehat{d}_{LL}(s) \;- b_i(\tau_{ir})\widehat{d}_{LL,1}(s)(\tau_{ir}-s) \\
&\quad - \sum_{j=1}^{m_i}\mathbf{1}[j\neq r]b_i(\tau_{ij})\widehat{d}_{LL,1}(s_{ij})(\tau_{ij}-s_{ij})\Bigg\}b_i(\tau_{ir})(\tau_{ir}-s) \\
&\qquad \prod_{j=1}^{m_i}\mathbf{1}[j\neq r]\left\{K_h(s_{ij}-\tau_{ij})\;ds_{ij}\right\}K_h(s-\tau_{ir}) \\
&= \sum_{i=1}^{n}\sum_{r=1}^{m_i}p_i b_i(\tau_{ir})(\tau_{ir}-s)K_h(s-\tau_{ir}) - \sum_{i=1}^{n}\sum_{r=1}^{m_i}b_i(\tau_{ir})^2(\tau_{ir}-s)^2K_h(s-\tau_{ir})\widehat{d}_{LL,1}(s) \\
&\quad - \sum_{i=1}^{n}\sum_{r=1}^{m_i}b_i(\tau_{ir})^2(\tau_{ir}-s)K_h(s-\tau_{ir})\widehat{d}_{LL}(s)
\end{aligned}
$$

$$-\sum_{i=1}^{n}\sum_{r=1}^{m_i}\sum_{\substack{j=1\\j\neq r}}^{m_i} b_i(\tau_{ij})b_i(\tau_{ir})(\tau_{ir}-s)K_h(s-\tau_{ir})\int \widehat{d}_{LL}(t)K_h(t-\tau_{ij})\,dt$$
$$-\sum_{i=1}^{n}\sum_{r=1}^{m_i}\sum_{\substack{j=1\\j\neq r}}^{m_i} b_i(\tau_{ij})b_i(\tau_{ir})(\tau_{ir}-s)K_h(s-\tau_{ir})\int \widehat{d}_{LL,1}(t)(\tau_{ij}-t)K_h(t-\tau_{ij})\,dt.$$

This can be written as:

$$\begin{pmatrix} \widehat{d}_{LL}(s) \\ h\widehat{d}_{LL,1}(s) \end{pmatrix} = \begin{pmatrix} \bar{d}(s) \\ \bar{d}_1(s) \end{pmatrix} + \int \widehat{H}_{LL}(s,t) \begin{pmatrix} \widehat{d}_{LL}(t) \\ h\widehat{d}_{LL,1}(t) \end{pmatrix} dt, \tag{16}$$

where

$$\begin{pmatrix} \bar{d}(s) \\ \bar{d}_1(s) \end{pmatrix} = M(s)^{-1} \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n}\sum_{r=1}^{m_i} p_i b_i(\tau_{ir})K_h(s-\tau_{ir}) \\ \frac{1}{n}\sum_{i=1}^{n}\sum_{r=1}^{m_i} p_i b_i(\tau_{ir})K_h(s-\tau_{ir})\frac{\tau_{ir}-s}{h} \end{pmatrix}$$

and

$$M(s) = \frac{1}{n}\sum_{i=1}^{n}\sum_{r=1}^{m_i} \begin{pmatrix} b_i(\tau_{ir})^2K_h(s-\tau_{ir}) & b_i(\tau_{ir})^2\frac{\tau_{ir}-s}{h}K_h(s-\tau_{ir}) \\ b_i(\tau_{ir})^2\frac{\tau_{ir}-s}{h}K_h(s-\tau_{ir}) & b_i(\tau_{ir})^2\frac{(\tau_{ir}-s)^2}{h^2}K_h(s-\tau_{ir}) \end{pmatrix} \tag{17}$$

Furthermore $\widehat{H}_{LL}(s,t)$ is equal to $M(s)^{-1}\bar{H}_{LL}(s,t)$, where $\bar{H}_{LL}(s,t)$ is a two times two matrix with elements:

$$\bar{H}_{11}(s,t) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{r=1}^{m_i}\sum_{\substack{j=1\\j\neq r}}^{m_i} b_i(\tau_{ir})b_i(\tau_{ij})K_h(s-\tau_{ir})K_h(t-\tau_{ij}),$$

$$\bar{H}_{12}(s,t) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{r=1}^{m_i}\sum_{\substack{j=1\\j\neq r}}^{m_i} b_i(\tau_{ir})b_i(\tau_{ij})\frac{\tau_{ij}-t}{h}K_h(s-\tau_{ir})K_h(t-\tau_{ij}),$$

$$\bar{H}_{21}(s,t) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{r=1}^{m_i}\sum_{\substack{j=1\\j\neq r}}^{m_i} b_i(\tau_{ir})b_i(\tau_{ij})\frac{\tau_{ir}-s}{h}K_h(s-\tau_{ir})K_h(t-\tau_{ij}),$$

$$\bar{H}_{22}(s,t) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{r=1}^{m_i}\sum_{\substack{j=1\\j\neq r}}^{m_i} b_i(\tau_{ir})b_i(\tau_{ij})\frac{\tau_{ir}-s}{h}\frac{\tau_{ij}-t}{h}K_h(s-\tau_{ir})K_h(t-\tau_{ij}).$$

Equation (16) suggests the following iteration for the calculation of $\widehat{d}_{LL}$ and $\widehat{d}_{LL,1}$:

$$\begin{pmatrix} \widehat{d}_{LL}^{[a+1]}(s) \\ h\widehat{d}_{LL,1}^{[a+1]}(s) \end{pmatrix} = \begin{pmatrix} \overline{d}(s) \\ \overline{d}_1(s) \end{pmatrix} + \int \widehat{H}_{LL}(s,t) \begin{pmatrix} \widehat{d}_{LL}^{[a]}(t) \\ h\widehat{d}_{LL,1}^{[a]}(t) \end{pmatrix} dt. \tag{18}$$

As starting value $(\widehat{d}_{LL}^{[0]}, h\widehat{d}_{LL,1}^{[0]})$ one can choose $(\overline{d}, \overline{d}_1)$ for example.

A sufficient condition that the iterative calculation $\widehat{d}_{LL}$ and $\widehat{d}_{1,LL}$ converges can be based on a check of the operator norm of the operator $\widehat{\mathcal{H}}_{LL}g(\cdot) = \int \widehat{H}_{LL}(\cdot,t)g(t)\ dt$. It is easy to see that the iterative calculation of $\widehat{d}_{LL}$ and $\widehat{d}_{1,LL}$ converges if $\|\widehat{\mathcal{H}}_{LL}\| < 1$, where $\|\widehat{\mathcal{H}}_{LL}\|$ is defined with respect to the sup norm or to the L_2 norm. In this case the limit is given as

$$\begin{pmatrix} \widehat{d}_{LL} \\ h\widehat{d}_{LL,1} \end{pmatrix} = \sum_{k=0}^{\infty} \widehat{\mathcal{H}}_{LL}^{k} \begin{pmatrix} \overline{d} \\ \overline{d}_1 \end{pmatrix}.$$

3.3 Estimating the yield and forward curve

Practitioners often prefer to have visual presentation (and an estimate) of both the yield curve and the forward curve. By definition the two representations are equivalent. However, in discussions of, for example, the effects of monetary policy it is customary to relate the discussion to forward curves [see Svensson (1994) and Dahlquist and Svensson (1996)]. Estimates of the yield curve and forward curve can be obtained from estimates of d by the relations $y(t) = -\log(d(t))/t$ and $f(t) = -d'(t)/d(t)$. Specifically, $\widehat{y}_j(t) = -\log(\widehat{d}_j(t))/t$, where $j = LC,LL$. In the forward curve case we can take the following estimator based on the local constant,

$$\widehat{f}_{LC}(t) = -\frac{\int K_h'(t-s)\widehat{d}_{LC}(s)ds}{\int K_h(t-s)\widehat{d}_{LC}(s)ds},$$

or we can make use of the local linear estimation to define

$$\widehat{f}_{LL}(t) = -\frac{\widehat{d}_{LL,1}(t)}{\widehat{d}_{LL}(t)}.$$

3.4 Local Constant Exponential Smoothing

We now consider an important modification to the basic smoothing methods we outlined above. Instead of fitting locally linear functions we fit now locally exponential functions. This approach is motivated by the fact that we can expect the discount function to be closer to log linear than linear, as was suggested by Vasicek and Fong (1982). We shall fit the function $d(s) = \exp(-y(s)s)$, in other words we will estimate the yield curve directly. This specification imposes the natural restriction that $d(\cdot)$ be positive and that $d(0) = 1$.[5] Alternative specifications here include $d(s) = 1 + sg(s)$ for some function g, which was used in Jordan (1994), and $d(s) = \exp(-su(s)/(1+s))$, which imposes the additional restriction that $d(s) \to 0$ as $s \to \infty$ and was used in Tanggaard (1997). See Gozalo and Linton (1998) for discussion of local nonlinear smoothing methods in nonparametric regression.

We define now the function $\widehat{y}_{LCE}(\cdot)$ which minimizes

$$\sum_{i=1}^{n} \int \left\{ p_i - \sum_{j=1}^{m_i} b_i(\tau_{ij}) \exp\{-\tau_{ij} y(s_{ij})\} \right\}^2 \prod_{j=1}^{m_i} \{K_h(s_{ij} - \tau_{ij})\ ds_{ij}\}, \tag{19}$$

with respect to $y(\cdot)$. To solve the minimization we put $y = \widehat{y}_{LCE} + \epsilon\delta_s$ where ϵ is a real number and δ_s is the Dirac [generalized] function in s. Differentiation of the right hand side with respect to ϵ gives (at $\epsilon = 0$):

$$\begin{aligned} 0 &= \sum_{i=1}^{n} \sum_{k=1}^{m_i} \int \left\{ p_i - \sum_{j=1}^{m_i} \mathbf{1}[j \neq k] b_i(\tau_{ij}) \exp\{-\tau_{ij}\widehat{y}_{LCE}(s_{ij})\} \ - b_i(\tau_{ik}) \exp\{-\tau_{ik}\widehat{y}_{LCE}(s)\} \right\} \\ &\quad \times b_i(\tau_{ik})\tau_{ik} \exp\{-\tau_{ik}\widehat{y}_{LCE}(s)\} \prod_{j=1}^{m_i} \mathbf{1}[j \neq k] \{K_h(s_{il} - \tau_{il})\ ds_{il}\} K_h(s - \tau_{il}) \\ &= \sum_{i=1}^{n} \sum_{k=1}^{m_i} p_i b_i(\tau_{ik})\tau_{ik} \exp\{-\tau_{ik}\widehat{y}_{LCE}(s)\} K_h(s - \tau_{ik}) \\ &\quad - \sum_{i=1}^{n} \sum_{k=1}^{m_i} b_i(\tau_{ik})^2 \tau_{ik} \exp\{-2\tau_{ik}\widehat{y}_{LCE}(s)\} K_h(s - \tau_{ik}) \\ &\quad - \sum_{i=1}^{n} \sum_{k=1}^{m_i} \sum_{\substack{j=1 \\ j \neq k}}^{m_i} b_i(\tau_{ij}) b_i(\tau_{ik}) \tau_{ik} \exp\{-\tau_{ik}\widehat{y}_{LCE}(s)\} K_h(s - \tau_{ik}) \int \exp\{-\tau_{ij}\widehat{y}_{LCE}(t)\} K_h(t - \tau_{ij})\ dt. \end{aligned} \tag{20}$$

[5] It also allows estimation of $y(s)$ for s close to zero, which is generally not possible when the discount function is estimated directly because the direct estimate does not impose that $d(s) = 1 + O(s)$ as $s \to 0$ [i.e., $\lim_{s\to 0} -\log \widehat{d}(s)/s$ may not exist, while $\lim_{s\to 0} \widehat{y}(s)$ does].

There are a number of solution methods applicable here. We suggest the following approach based on the local nonlinear least squares criterion function in θ

$$\sum_{i=1}^{n}\sum_{k=1}^{m_i}\{p_i - b_i(\tau_{ik})\exp(-\tau_{ik}\theta)\}^2 K_h(s-\tau_{ik}). \tag{21}$$

The first order condition for the minimizer of this function is precisely the first two terms in (20). Now let $\widehat{y}_{LCE}(t)$ be given and define $\widehat{y}_{LCE}^{[a]}(t)$ as the minimizer of (21) with p_i replaced by

$$p_{ik}^{[a]} = p_i - \sum_{\substack{j=1\\ j\neq k}}^{m_i} b_i(\tau_{ij})\int \exp\{-\tau_{ij}\widehat{y}_{LCE}(t)\}K_h(t-\tau_{ij})dt.$$

Taking some initial value $\widehat{y}_{LCE}^{[0]}(t)$, we proceed $a = 1, 2, \ldots$. The nonlinear optimization problem (21) can itself be approximately solved using Newton's method or Fisher scoring. Note that a consistent initial value is provided by $\widehat{y}_j(t) = -\log\widehat{d}_j(t)/t$, where $j = LC, LL$. We can also define 'one-step' estimators, denoted $\widehat{y}_{LCE}^{j}(t)$, which use $\widehat{y}_j(t)$ as initial condition and terminate at $a = 1$.

3.5 Local Linear Exponential Smoothing

The natural extension of the local constant procedure to the local linear is also important for the same reasons outlined above. Define the following criterion function.

$$\int \sum_{i=1}^{n}[p_i - \sum_{j=1}^{m_i} b_i(\tau_{ij})\exp\{-\tau_{ij}y(s_{ij}) - \tau_{ij}(\tau_{ij}-s_{ij})y_1(s_{ij})\}]^2 \prod_{l=1}^{m_i}\{K_h(s_{il}-\tau_{il})ds_{il}\},$$

and let $(\widehat{y}_{LLE}, \widehat{y}_{LLE,1})$ minimize this with respect to (y, y_1). As before, let $y = \widehat{y}_{LLE} + \epsilon\delta_s$ and

$$\begin{aligned}
0 \;=\; & \sum_{i=1}^{n}\sum_{k=1}^{m_i} p_i b_i(\tau_{ik})K_h(s-\tau_{ik})\exp\{-\tau_{ik}\widehat{y}_{LLE}(s) - \tau_{ik}\widehat{y}_{LLE,1}(s)(\tau_{ik}-s)\}\tau_{ik} \\
& -\sum_{i=1}^{n}\sum_{k=1}^{m_i}\sum_{\substack{j=1\\ j\neq k}}^{m_i} b_i(\tau_{ij})b_i(\tau_{ik})K_h(s-\tau_{ik})\exp\{-\tau_{ik}\widehat{y}_{LLE}(s) - \tau_{ik}\widehat{y}_{LLE,1}(s)(\tau_{ik}-s)\}\tau_{ik} \\
& \times \int K_h(t-\tau_{ij})\exp\{-\tau_{ij}\widehat{y}_{LLE}(t) - \tau_{ij}\widehat{y}_{LLE,1}(t)(\tau_{ij}-t)\}dt \\
- \; & \sum_{i=1}^{n}\sum_{k=1}^{m_i} b_i(\tau_{ik})^2\exp\{-2\tau_{ik}\widehat{y}_{LLE}(s) - 2\tau_{ik}\widehat{y}_{LLE,1}(s)(\tau_{ik}-s)\}\tau_{ik}.
\end{aligned}$$

Furthermore, let $y_1 = \widehat{y}_{LLE,1} + \epsilon\delta_s$ and

$$\begin{aligned} 0 &= \sum_{i=1}^{n}\sum_{k=1}^{m_i} p_i b_i(\tau_{ik}) K_h(s-\tau_{ik}) \exp\{-\tau_{ik}\widehat{y}_{LLE}(s) - \tau_{ik}\widehat{y}_1(s)(\tau_{ik}-s)\}\tau_{ik}(\tau_{ik}-s) \\ &\quad - \sum_{i=1}^{n}\sum_{k=1}^{m_i}\sum_{\substack{j=1\\ j\neq k}}^{m_i} b_i(\tau_{ij}) b_i(\tau_{ik}) K_h(s-\tau_{ik}) \exp\{-\tau_{ik}\widehat{y}_{LCE}(s) - \tau_{ik}\widehat{y}_{LLE,1}(s)(\tau_{ik}-s)\}\tau_{ik}(\tau_{ik}-s) \\ &\quad \times \int K_h(t-\tau_{ij}) \exp\{-\tau_{ij}\widehat{y}_{LLE}(t) - t\widehat{y}_{LLE,1}(t)(\tau_{ij}-t)\}dt \\ &- \sum_{i=1}^{n}\sum_{k=1}^{m_i} b_i(\tau_{ij})^2 \exp\{-2\tau_{ik}\widehat{y}_{LLE}(s) - 2\tau_{ik}\widehat{y}_{LLE,1}(s)(\tau_{ik}-s)\}\tau_{ik}(\tau_{ik}-s). \end{aligned}$$

We can obtain iterative equations from this as in the previous section to define estimates of $\widehat{y}_{LLE}(s)$ and $\widehat{y}_{LLE,1}(s)$.

3.6 Estimating the discount function and forward curve

The discount function can of course be estimated by $\widehat{d}_j(t) = \exp(-t\widehat{y}_j(t))$, where $j = LCE, LLE$. The implicit forward curve, $f(t)$, is defined through the relation $y(t) = \int_0^t f(u)du/t$. Note that differentiation of the last equality gives $f(t) = y(t) + ty'(t)$. Given an estimate, $\widehat{y}$, of the yield curve y and an estimate $\widehat{y}'$ of its derivative, we can estimate f by

$$\hat{f}(t) = \widehat{y}(t) + t\widehat{y}'(t).$$

An estimate $\widehat{y}'$ of y' is given by the estimate $\widehat{y}_{LLE,1}$ (see the last section) or can be calculated by smoothed differentiation of an estimate $\widehat{y}$ of y, $\widehat{y}'_{LCE}(t) = \int K'_h(t-s)\widehat{y}_{LCE}(s)ds$.

4 Asymptotic Properties

Our main results are for the local constant and local linear discount function smoothing methods. We also provide a result for the one-step local exponential methods and indicate the expected result for the 'iterate to convergence' versions. As we show, the asymptotic variances of the four procedures described above are identical [when estimating the same quantity], although the bias is certainly influenced by the chosen functional form. For comparison, Gozalo and Linton (1998) show that the asymptotic variance of local nonlinear least squares estimators does not depend on the specific functional form used in the estimation, although the bias does.

4.1 Local Constant Smoothing

Our asymptotics will be stated for the case of nonrandom τ_{ij} and m_i. We concentrate on the case of a nonrandom "design" because a stochastic model for τ_{ij} and m_i that would be adequate for our application is rather complicated. It must take into consideration some very special properties of these data (e.g., $\tau_{ij+1} - \tau_{ij}$ is constant and a multiple of half years). Furthermore our model (4) arises in other applications that would require another stochastic model for the design (for another application see e.g., Engle et al. (1986)). So we prefer to state the theorem under conditions that may work also in these applications. Our conditions are 'high level' and we discuss further below how they can be verified in some special cases.

Assumptions for Theorem 1

(A1) Model (4) holds. The error variables $\varepsilon_1, \ldots, \varepsilon_n$ are independent with mean zero and with $E\varepsilon_i^{2+\delta} \leq C^{2+\delta}$ for n large enough with constants $C, \delta > 0$. [This implies that $\sigma_i^2 = \mathrm{var}[\varepsilon_i] \leq C^2$ for n large enough.] The values τ_{ij} and m_i are deterministic and they, the distributions of ε_i and the functions $b_i(\cdot)$ may depend on the sample size n. For two constants $\gamma_1, \gamma_2 > 0$ the bandwidth h fulfills $n^{-1+\gamma_1} \leq h \leq n^{-\gamma_2}$ for n large enough.

(A2) There exist constants $\rho < 1$, $C > 0$ and $\kappa \geq 1$ such that for n large enough $\|\widehat{\mathcal{H}}_{LC}\|_p < C$ and $\|\widehat{\mathcal{H}}_{LC}^{\kappa}\|_p < \rho$ for $p = 2$.

For the asymptotic treatment of our estimate at a fixed point s we make the following condition.

(A3) There exists a constant C such that for n large enough

$$\int \widehat{H}_{LC}(s,t)^2 \, dt \quad \leq \quad C^2, \tag{22}$$

$$\frac{1}{n}\sum_{i=1}^{n}\int\left|\int \widehat{H}_{LC}(t,u)R_i(u)\ du\right|^2\ dt \quad = \quad o(h^{-1}), \tag{23}$$

$$|\int \widehat{H}_{LC}(s,u)R_i(u)\ du| \quad = \quad o((n/h)^{1/2}\log(n)^{-1}), \tag{24}$$

$$\max_{1\le i\le n}|R_i(s)| \quad = \quad o((n/h)^{1/2}), \tag{25}$$

where

$$\begin{aligned} R_i(t) &= \frac{\sum_{l=1}^{m_i} b_i(\tau_{il})K_h(t-\tau_{il})}{\frac{1}{n}\sum_{i=1}^{n}\sum_{r=1}^{m_i} b_i(\tau_{ir})^2K_h(t-\tau_{ir})}, \\ R(t) &= \frac{1}{n}\sum_{i=1}^{n} R_i(t). \end{aligned}$$

Furthermore, it holds that

$$s_n^{LC}(s)^{-2} = O(1). \tag{26}$$

The quantity $s_n^{LC}(s)$ is defined in the statement of Theorem 1 below.

For the uniform asymptotic expansion of our estimate on an interval S we need the following additional condition.

(A4) Condition (22) holds for all $s \in S$ with a constant C that does not depend on s. The supremum of the left hand side of (23) and of (24) over $s \in S$ is of order $o(h^{-1})$ or $o((n/h)^{1/2}\log(n)^{-1})$, respectively. The error variables have a finite Laplace transform

$$\sup_{1\le i\le n} E\exp(t|\varepsilon_i|) < C \tag{27}$$

for a constant C, for $t > 0$ small enough, and for n large enough. Furthermore there exists a constant C such that for n large enough for all $s \in S$

$$\int\left|\frac{\partial}{\partial s}\widehat{H}_{LC}(s,t)\right| R(t)\ dt \le n^C. \tag{28}$$

Our results can be used to treat random τ_{ij} and m_i. In this case, one has to assume that our conditions hold conditionally given τ_{ij} and m_i with probability tending to one. In particular, if one assumes that the variables m_i are i.i.d. from some distribution with bounded support and that the distribution of $(\tau_{i1}, \ldots, \tau_{im_i})$, conditional on m_i, is absolutely continuous with respect to Lebesgue measure, it follows [under some additional regularity conditions] that the kernel $\widehat{H}_{LC}(s,t)$ converges to a function $H_{LC}(s,t)$ with probability one as $n \to \infty$. In this case for the verification of (A2) it suffices to show that the operator $\mathcal{H}_{LC}$ defined by $\mathcal{H}_{LC}g(s) = \int H_{LC}(s,t)g(t)dt$ satisfies $\|\mathcal{H}_{LC}^{\kappa}\| < 1$ for a $\kappa \geq 1$. Similarly one can check (22) and (28). The operator $\mathcal{H}_{LC}$ depends on the payment functions and on the time distributions. Suppose that $m_i = m$ for all i and that $b_i(\tau_{ij}) = \beta_j$ for all i, j. Then, if the time distributions are completely homogenous, i.e., $f_{ij} = f_{lk}$ for all i, j, l, k, one can show that

$$[\|\mathcal{H}_{LC}^{\kappa}\|_2]^{1/\kappa} \to \frac{\sum \sum_{j \neq r}^{m} \beta_j \beta_r}{\sum_{j=1}^{m} \beta_j^2} \tag{29}$$

for $\kappa \to \infty$. The limit will be less than one provided the β_j's are sufficiently heterogeneous. This condition is likely to met in practice because the redemption value is usually considerably larger than the coupons. For an analytical example, suppose that $\beta_j = c$ for $j = 1, \ldots, m-1$ and that $\beta_m = 100 + c$, then $\|\mathcal{H}_{LC}^{\kappa}\|_2 < 1$ for κ large enough, provided $c(m-1)/(100+c) < (\sqrt{5}-1)/2$. This amounts to the requirement that slightly more than half of the total payment be received at the end point. This condition can be satisfied provided the dataset contains many short maturity bonds/small coupon bonds relative to long maturity/large coupons.

We give now a rough discussion why the other conditions hold under appropriate regularity assumptions on the distribution of τ_{ij} and m_i. We will do this for the case that m_i has bounded support. For a motivation of conditions (23) and (24) consider for simplicity the case $m_i = 1$ (with probability one), $b_i \equiv 1$. Then we get $\int \widehat{H}_{LC}(s,u)R_i(u)\ du \approx \int H_{LC}(s,u)R_i(u)\ du \approx H_{LC}(s,\tau_{i1})/f(\tau_{i1})$, where f is the density of τ_{i1}. So under boundedness conditions of H/f one expects that the left hand side of (23) and (24) is of order $O(1)$. For more general conditions on m_i and b_i a similar discussion applies with another definition of f. For a check of (25) note first that under regularity conditions the denominator in the definition of $R_i(s)$ converges to a smooth function g. So we get that the left hand side of (25) is of order $O(h^{-1})$ if $\max_i b_i(s)$ and the kernel K are bounded and if $g(s) > 0$. So (25) follows because $nh \to \infty$. The assumption (27) can be considerably weakened when more specific assumptions are made on the data generating process and kernel.

In our first theorem we will state asymptotic properties of the local constant estimate $\widehat{d}_{LC}$.

THEOREM 1 [ASYMPTOTIC NORMALITY OF LOCAL CONSTANT ESTIMATE]. *Suppose that Conditions A1-A2 hold and that A3 holds for a fixed s. Then*

$$\sqrt{nh}\frac{\widehat{d}_{LC}(s) - d(s) - \beta_n^{LC}(s)}{s_n^{LC}(s)} \Longrightarrow N(0,1), \tag{30}$$

where

$$\begin{aligned} s_n^{LC}(s)^2 &= \frac{h\frac{1}{n}\sum_{i=1}^n \left[\sum_{r=1}^{m_i} b_i(\tau_{ir})K_h(s-\tau_{ir})\right]^2 \sigma_i^2}{\left[\frac{1}{n}\sum_{i=1}^n \sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(s-\tau_{ir})\right]^2}, \\ \beta_n^{LC}(s) &= \sum_{k=0}^{\infty} \widehat{\mathcal{H}}_{LC}^k \beta_n^{*,LC}(s), \\ \beta_n^{*,LC}(s) &= \frac{\sum_{i=1}^n \sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(s-\tau_{ir})[d(\tau_{ir}) - d(s)]}{\sum_{i=1}^n \sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(s-\tau_{ir})} \\ &\quad + \frac{\sum_{i=1}^n \sum_{r=1}^{m_i} \sum_{j=1,r\neq j}^{m_i} b_i(\tau_{ir}) b_i(\tau_{ij}) K_h(s-\tau_{ir}) \int K_h(t-\tau_{ij})[d(\tau_{ij}) - d(t)]\, dt}{\sum_{i=1}^n \sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(s-\tau_{ir})}. \end{aligned}$$

Under the assumption that conditions A1-A2 hold and that A4 holds for a finite interval S the following uniform expansion holds

$$\sup_{s\in S}\left| \widehat{d}_{LC}(s) - d(s) - \beta_n^{LC}(s) - \frac{\sum_{i=1}^n \sum_{r=1}^{m_i} b_i(\tau_{ir})K_h(s-\tau_{ir})\varepsilon_i}{\sum_{i=1}^n \sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(s-\tau_{ir})} \right| = o_P((nh)^{-1/2}). \tag{31}$$

A version of Theorem 1 can be proved for the case that the operator norm $\|\widehat{\mathcal{H}}_{LC}^{\kappa}\|_\infty < 1$ for a $\kappa \geq 1$, i.e. that Condition A2 holds for $p = \infty$.

The expansion (31) shows that asymptotically the estimate $\widehat{d}_{LC}(s)$ behaves like a statistic that is linear in the error variables and that has a similar structure to a univariate kernel regression smoother. Proceeding as in standard theory of kernel smoothing one can use expansion (31) for the construction of pointwise confidence intervals and of uniform confidence bands and for the discussion of optimal bandwidths, see e.g. Härdle (1990). We will not pursue these points here, although we give formulae for standard errors in section 5 below. Under additional regularity conditions it can be shown that the bias β_n^{LC} is of order h^2. This follows by showing that $\beta_n^{*,LC}$ is of this order. For this reason we will get the same asymptotic optimal rates in our model as in standard nonparametric regression.

Theorem 1 can be used to determine the asymptotic distribution of the corresponding estimates of $y(t)$ and $f(t)$. By the delta method,

$$\widehat{y}_{LC}(s) - y(s) = -\frac{\widehat{d}_{LC}(s) - d(s)}{sd(s)} + o_p\left(\left|\widehat{d}_{LC}(s) - d(s)\right|\right), \tag{32}$$

provided $s > 0$, from which we see that the asymptotic bias and standard deviation of $\widehat{y}_{LC}(s)$ are those of $\widehat{d}_{LC}(s)$ multiplied by $1/sd(s)$. The estimated forward curve $\widehat{f}_{LC}(s)$ can be handled similarly. Suppose that K has a bounded support $[-a, a]$ and that it has a bounded derivative. Then if A1-A4 hold for $S = [s - ah, s + ah]$ we get from (31) that

$$\widehat{f}_{LC}(s) = f(s) + \Delta(s) + \sum_{i=1}^{n} w_{n,i}(s)\varepsilon_i + o_P((nh^3)^{-1/2}),$$

where $\Delta(s) = d^{-1}(s) \int K_h(s-t)[d'(t) - d'(s)]dt + d^{-1}(s) \int K_h'(s-t)\beta_n^{LC}(t)dt - d^{-2}(s)d'(s) \int K_h(s-t)\beta_n^{LC}(t)dt$, and

$$w_{n,i}(s) = d^{-1}(s) \int K_h'(s-t) \frac{\sum_{r=1}^{m_i} b_i(\tau_{ir})K_h(t-\tau_{ir})}{\sum_{l=1}^{n}\sum_{j=1}^{m_l} b_l(\tau_{lj})^2 K_h(t-\tau_{lj})}\, dt.$$

One can check that under regularity conditions this implies that $\widehat{f}_{LC}(s)$ has an asymptotic bias of order h^2 and variance of order $(nh^3)^{-1}$. This coincides with rates of convergence of derivative estimates based on (smoothed) derivatives of local constant smoothers in standard nonparametric regression.

4.2 Local Linear Smoothing

For a 2×2 matrix M we define $\|M\|^2 = \sup_{e^T e = 1} e^T M e$.

Assumptions for Theorem 2

(B1) Assumption A1 holds and there exist constants $\rho < 1$, $C > 0$ and $\kappa \geq 1$ such that for n large enough $\|\widehat{\mathcal{H}}_{LL}\|_2 = \sup_{\|g\|_2=1} \|\widehat{\mathcal{H}}_{LL} g\|_2 < C$ and $\|\widehat{\mathcal{H}}_{LL}^{\kappa}\|_2 = \sup_{\|g\|_2=1} \|\widehat{\mathcal{H}}_{LL}^{k} g\|_2 < \rho$, where for a function $g : \mathbb{R}^2 \to \mathbb{R}^2$ we define $\|g\|_2^2 = \int_{\mathbb{R}^2} \|g(x)\|^2 \, dx$.

For the asymptotic treatment of our estimate at a fixed point s we make the following condition.

(B2) There exists a constant C such that for n large enough

$$\int \|\widehat{H}_{LL}(s,t)\|^2 \, dt \leq C^2, \tag{33}$$

$$\frac{1}{n}\sum_{i=1}^{n} \int \left\| \int \widehat{H}_{LL}(s,u) R_i(u) \, du \right\|^2 \, dt = o(h^{-1}), \tag{34}$$

$$\| \int \widehat{H}_{LL}(s,u) R_i(u) \, du\| = o((n/h)^{1/2} \log(n)^{-1}), \tag{35}$$

$$\max_{1 \leq i \leq n} \|R_i(s)\| = o((nh)^{1/2}),$$

where now

$$\begin{aligned} R_i(t) &= M(s)^{-1} \sum_{l=1}^{m_i} b_i(\tau_{il}) K_h(t - \tau_{il}) \begin{pmatrix} 1 \\ (\tau_{il} - t)/h \end{pmatrix}, \\ R(t) &= \frac{1}{n} \sum_{i=1}^{n} R_i(t). \end{aligned}$$

Furthermore, it holds that

$$\|S_n^{LL}(s)^{-1}\| = O(1),$$

where

$$S_n^{LL}(s) = h M(s)^{-1} \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 \begin{pmatrix} v_{i,0}^2(s) & v_{i,0}(s) v_{i,1}(s) \\ v_{i,0}(s) v_{i,1}(s) & v_{i,1}^2(s) \end{pmatrix} M(s)^{-1},$$

$$v_{i,j}(s) = \sum_{r=1}^{m_i} b_i(\tau_{ir}) K_h(s - \tau_{ir}) \left[\frac{\tau_{ir} - s}{h} \right]^j \qquad \textit{for } i = 1, \ldots, n \textit{ and } j = 0, 1.$$

(B3) It holds that

$$\frac{\sum_{i=1}^n \sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(s-\tau_{ir}) \frac{\tau_{ir}-s}{h}}{\sum_{i=1}^n \sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(s-\tau_{ir}) \left[\frac{\tau_{ir}-s}{h}\right]^2} \to 0,$$
$$\frac{\sum_{i=1}^n \sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(s-\tau_{ir}) \frac{\tau_{ir}-s}{h}}{\sum_{i=1}^n \sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(s-\tau_{ir})} \to 0.$$

For the uniform asymptotic expansion of our estimate on an interval S we need the following additional condition.

(B4) Condition (33) holds for all $s \in S$ with a constant C that does not depend on s. The supremum of the left hand side of (34) and of (35) over $s \in S$ is of order $o(h^{-1}\log(n)^{-2})$ or $o((n/h)^{1/2}\log(n)^{-1})$, respectively. The error variables have a finite Laplace transform

$$\sup_{1\le i\le n} E\exp(t|\varepsilon_i|) < C$$

for a constant C, for $t > 0$ small enough, and for n large enough. Furthermore there exists a constant C such that for n large enough for all $s \in S$

$$\int \left\| \frac{\partial}{\partial s} \widehat{H}_{LC}(s,t) \right\| R(t)\ dt \le n^C. \tag{36}$$

Theorem 2 [Asymptotic Normality of Local Linear Estimate]. *Suppose that Condition B1 holds and that B2 holds for a fixed s. Then*

$$\sqrt{nh} S_n^{LL}(s)^{-1/2} \left[\begin{pmatrix} \widehat{d}_{LL}(s) - d(s) \\ h\left(\widehat{d}_{LL,1}(s) - d'(s)\right) \end{pmatrix} - \beta_n^{LL}(s) \right] \Longrightarrow N(0, I_2), \tag{37}$$

where I_2 is a 2×2 identity matrix and where

$$\begin{aligned} \beta_n^{LL}(s) &= \sum_{k=0}^{\infty} \widehat{\mathcal{H}}_{LL}^k \beta_n^{*,LL}(s), \\ \beta_n^{*,LL}(s) &= M(s)^{-1} \sum_{i=1}^n \sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(s-\tau_{ir})[d(\tau_{ir}) - d(s) - (\tau_{ir}-s)d'(s)] \begin{pmatrix} 1 \\ (\tau_{ir}-s)/h \end{pmatrix} \end{aligned}$$

$$+M(s)^{-1}\sum_{i=1}^{n}\sum_{r=1}^{m_i}\sum_{j=1,r\neq j}^{m_i} b_i(\tau_{ir})b_i(\tau_{ij})K_h(s-\tau_{ir})\int K_h(t-\tau_{ij})[d(\tau_{ij})-d(t)$$
$$-(\tau_{ij}-t)d'(t)]\ dt\begin{pmatrix}1\\(\tau_{ir}-s)/h\end{pmatrix}.$$

Here, again σ_i^2 denotes the variance of p_i. Under the additional assumption of B3 we get that $s_n^{LC}(s)^{-1}\widehat{d}_{LL}(s)$ has asymptotic variance one, i.e., $\widehat{d}_{LL}(s)$ has the same asymptotic variance as $\widehat{d}_{LC}(s)$. Under the assumption that Condition B1 holds and that B4 holds for a finite interval S the following uniform expansion holds

$$\sup_{s\in S}\left\|\begin{pmatrix}\widehat{d}_{LL}(s)\\h\widehat{d}_{LL,1}(s)\end{pmatrix}-\begin{pmatrix}d(s)\\hd'(s)\end{pmatrix}-\beta_n^{LL}(s)\right.$$
$$\left.-M(s)^{-1}\frac{1}{n}\sum_{i=1}^{n}\sum_{r=1}^{m_i}b_i(\tau_{ir})K_h(s-\tau_{ir})\varepsilon_i\begin{pmatrix}1\\(\tau_{ir}-s)/h\end{pmatrix}\right\|=o_P((nh)^{-1/2}). \qquad (38)$$

The asymptotic properties of $\widehat{y}_{LL}(s)$ and $\widehat{f}_{LL}(s)$ follow as above.

4.3 Local Exponential Smoothing

We just outline the result for the local constant exponential method. Consider the one-step problem in which $\widehat{y}^j_{LCE}(s)$ minimizes

$$Q_n(\theta)=\sum_{i=1}^{n}\sum_{k=1}^{m_i}\left\{\widehat{p}_{ik}-b_i(\tau_{ik})\exp(-\tau_{ik}\theta)\right\}^2K_h(s-\tau_{ik})$$

with respect to θ, where $\widehat{p}_{ik}=p_i-\sum_{j=1, j\neq k}^{m_i}b_i(\tau_{ij})\int\exp\{-\tau_{ij}\widehat{y}_j(t)\}K_h(t-\tau_{ij})dt$ and $\widehat{y}_j(t)=-\log\widehat{d}_j(t)/t$ with $j=LC,LL$. We shall approximate $\widehat{y}^j_{LCE}(s)$ by an infeasible procedure that is easier to work with. Let $Q_n^*(\theta)$ be the same criterion function with $\widehat{p}_{ik}$ replaced by $p_{ik}=p_i-\sum_{j=1, j\neq k}^{m_i}b_i(\tau_{ij})\exp\{-\tau_{ij}y(\tau_{ij})\}$, and let $\widehat{y}^*_{LCE}(s)$ be the minimizer of $Q_n^*(\theta)$. It is a straightforward application of Gozalo and Linton (1998, Theorem 2) to deduce the asymptotic distribution of $\widehat{y}^*_{LCE}(s)$; specifically, we obtain $\widehat{y}^*_{LCE}(s)\sim y(s)+\beta_n^*(s)+\alpha_n^*(s)$, where

$$\beta_n^*(s)=\frac{\sum_{i=1}^n\sum_{r=1}^{m_i}K_h(s-\tau_{ir})b_i(\tau_{ir})^2\tau_{ir}\exp(-2\tau_{ir}y(s))[\exp(-\tau_{ir}(y(\tau_{ir})-y(s))-1]}{\sum_{i=1}^n\sum_{r=1}^{m_i}K_h(s-\tau_{ir})b_i(\tau_{ir})^2\tau_{ir}^2\exp(-2\tau_{ir}y(s))}$$

$$\alpha_n^*(s) = \frac{\sum_{i=1}^n \sum_{r=1}^{m_i} K_h(s-\tau_{ir}) b_i(\tau_{ir}) \tau_{ir} \exp(-\tau_{ir} y(s)) \varepsilon_i}{\sum_{i=1}^n \sum_{r=1}^{m_i} K_h(s-\tau_{ir}) b_i(\tau_{ir})^2 \tau_{ir}^2 \exp(-2\tau_{ir} y(s))}.$$

We now turn to the feasible one-step estimator. In the previous sections we established the uniform asymptotic expansion $\widehat{y}_j(s) \sim y(s) + \beta_n^j(s) + \alpha_n^j(s)$, where $\alpha_n^j(s)$ is a sum of independent mean zero random variables, while $\beta_n^j(s)$ is a deterministic bias function $[j = LC, LL]$. Combining the expansions for $\widehat{y}_{LCE}^*(s)$ and $\widehat{y}_j(s)$, we obtain after some manipulations that $\widehat{y}_{LCE}^j(s) \sim \widehat{y}_{LCE}^*(s) + \beta_n^{**}(s)$ uniformly in s, where $\beta_n^{**}(s) =$

$$\frac{\sum_{i=1}^n \sum_{r=1}^{m_i} \sum_{j=1, r\neq j}^{m_i} K_h(s-\tau_{ir}) b_i(\tau_{ir}) b_i(\tau_{ij}) \tau_{ir} e^{-\tau_{ir} y(s) - \tau_{ij} y(\tau_{ij})} \int K_h(t-\tau_{ij})[e^{-\tau_{ij}(y(t)-y(\tau_{ij}))} - 1]dt}{\sum_{i=1}^n \sum_{r=1}^{m_i} K_h(s-\tau_{ir}) b_i(\tau_{ir})^2 \tau_{ir}^2 \exp(-2\tau_{ir} y(s))}$$

$$+\frac{\sum_{i=1}^n \sum_{r=1}^{m_i} \sum_{j=1, r\neq j}^{m_i} K_h(s-\tau_{ir}) b_i(\tau_{ir}) b_i(\tau_{ij}) \tau_{ir} e^{-\tau_{ir} y(s) - \tau_{ij} y(\tau_{ij})} \tau_{ij} \int K_h(t-\tau_{ij}) \beta_n^j(t) dt}{\sum_{i=1}^n \sum_{r=1}^{m_i} K_h(s-\tau_{ir}) b_i(\tau_{ir})^2 \tau_{ir}^2 \exp(-2\tau_{ir} y(s))}.$$

Note that the stochastic terms from $\widehat{y}_j(s)$ do not appear in the asymptotics for $\widehat{y}_{LCE}^j(s)$, i.e., the asymptotic variance of $\widehat{y}_{LCE}^j(s)$ is the same as that of $\widehat{y}_{LCE}^*(s)$; this follows by the same arguments used in the proof of Theorem 1. The bias is changed though. If we iterate once more the variance is again unaffected but the bias formula replaces $\beta_n^j(s)$ by $\beta_n^*(s) + \beta_n^{**}(s)$.

In conclusion, $\widehat{y}_{LCE}^j(s)$ is asymptotically normal with variance $\mathrm{var}(\alpha_n^*(s))$.

5 Numerical Results

5.1 Methodology

5.1.1 Bandwidth Choice

The first problem when implementing the algorithm is to choose a suitable bandwidth. The bandwidth must be large enough to produce a smooth curve, while on the other being small enough not to produce too much bias. In the calculations presented here we have used the ocular method. For an automatic bandwidth choice we recommend the Generalized Cross-Validation (GCV) method discussed in Hastie and Tibshirani (1990, p. 49). Denote by S_h the $n \times n$ smoother matrix that maps the vector of prices into their corresponding present values. This matrix has elements $s_{\ell i}$, where for the linear smoother the weights are, approximately,

$$s_{\ell i} = \sum_{j=1}^{m_\ell} b_\ell(\tau_{\ell j}) \frac{\sum_{r=1}^{m_i} b_i(\tau_{ir}) K_h(\tau_{\ell j} - \tau_{ir})}{\sum_{i=1}^{n} \sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(\tau_{\ell j} - \tau_{ir})},$$

while for the nonlinear smoother

$$s_{\ell i} = \sum_{j=1}^{m_\ell} b_\ell(\tau_{\ell j}) \frac{\sum_{r=1}^{m_i} K_h(\tau_{\ell j} - \tau_{ir}) b_i(\tau_{ir}) \tau_{ir}^2 \exp(-2\tau_{ir}\widehat{y}(\tau_{\ell j}))}{\sum_{i=1}^{n} \sum_{r=1}^{m_i} K_h(\tau_{\ell j} - \tau_{ir}) b_i(\tau_{ir})^2 \tau_{ir}^2 \exp(-2\tau_{ir}\widehat{y}(\tau_{\ell j}))}.$$

We can now define the GCV-criterion as

$$GCV = \frac{\sum_{i=1}^{n} \widehat{\varepsilon}_i^2}{\left[1 - \mathrm{tr}\,(S_h)\,/n\right]^2}, \tag{39}$$

where $\widehat{\varepsilon}_i = p_i - \sum_{r=1}^{m_i} b_i(\tau_{ir})\widehat{d}(\tau_{ir})$ are the corresponding residuals and $\widehat{d}(s)$ is any of our estimates of $d(s)$. Minimizing GCV over h produces an estimate of the bandwidth which is optimal according to mean squared error, under some conditions.

5.1.2 Confidence Intervals

Let

$$\begin{aligned}
\widehat{v}_n(s)^2 &= \frac{\sum_{i=1}^{n} \left[\sum_{r=1}^{m_i} b_i(\tau_{ir}) K_h(s - \tau_{ir})\right]^2 \widehat{\varepsilon}_i^2}{\left[\sum_{i=1}^{n} \sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(s - \tau_{ir})\right]^2} \\
\widehat{v}_{yn}(s)^2 &= \frac{\sum_{i=1}^{n} \left[\sum_{r=1}^{m_i} K_h(s - \tau_{ir}) b_i(\tau_{ir}) \tau_{ir} \exp(-\tau_{ir}\widehat{y}(s))\right]^2 \widehat{\varepsilon}_i^2}{\left[\sum_{i=1}^{n} \sum_{r=1}^{m_i} K_h(s - \tau_{ir}) b_i(\tau_{ir})^2 \tau_{ir}^2 \exp(-2\tau_{ir}\widehat{y}(s))\right]^2},
\end{aligned}$$

where $\widehat{\varepsilon}_i = p_i - \sum_{r=1}^{m_i} b_i(\tau_{ir})\widehat{d}(\tau_{ir})$ are the corresponding residuals and $\widehat{d}(s)$ is any of our estimates of $d(s)$. Then

$$\widehat{d}(s) \pm z_{\alpha/2}\widehat{v}_n(s)$$

is an asymptotic $1 - \alpha$ confidence interval for $d(s)$, where z_α is the $1 - \alpha$ quantile of a standard normal distribution. As in usual nonparametric regression this property holds if the bias of $\widehat{d}(s)$ is of smaller order than $(nh)^{-1/2}$. Because under appropriate regularity conditions the bias is of order h^2, this requires a choice of h that is of order $o(n^{-1/5})$ (i.e., undersmoothing). For other choices of h the confidence intervals need bias corrections, see the related discussions in usual nonparametric regression in Härdle (1990). Furthermore,

$$\widehat{y}(s) \pm z_{\alpha/2}\widehat{v}_{yn}(s)$$

is an asymptotic $1-\alpha$ confidence interval for $y(s)$ subject to the same provisos.

5.1.3 Implementing the algorithm

We use the local linear exponential method throughout because of its superior bias performance, which has been established in simulations available from the authors upon request. Furthermore, to take account of the pronounced heteroskedasticity we use a weighting scheme with the weight on bond i being proportional to the inverse of the bonds time to maturity [as suggested by Chambers et al. (1984)]. A computer program in Pascal is available from Tanggaard and in Gauss from Linton for carrying out the calculations.

5.2 Application

The smoothing procedure is now illustrated using actual market data. The data consists of dealer quotes of US Treasury bonds in the 6 year period November 1990 to October 1996. The data set was obtained from the Federal Reserve Bank of New York based on quotations at 3:30 PM from a sample of market participants.[6] According to the Federal Reserve Bank of New York, the prices are believed to be reliable, thus, representing actual trading possibilities.

The number of outstanding issues at a given time is more than 200, and there are several considerations to take into account when reducing the sample. First of all, callable bonds with interest rate dependent cashflow must be removed. Today all US Treasury bonds are issued as non-callable, and as we consider only recently issued bonds, this is really not a problem. Because we use quotes

[6] The data were posted on the ftp-server ftp://ftp.ny.frb.org/qsheet.

instead of traded prices, non-synchronous trading is not a problem. However, tax–clientele effects and illiquidity may affect the estimation.

US Treasury notes are issued with original redemption dates after 2,3,5,10, and 20 years and with a coupon which is close to the market rate. Thus, in a short period after the auction the bonds may trade close to par. This reduces the effects of different taxation of interest income and capital gains. We therefore, consider the present value equation on a pre-tax basis. T-Bills are zero-coupon bonds with remaining time to maturity less than or equal to one year. As there is no interest income, we can also consider T-Bills as priced on a before-tax basis.

The market for US Treasuries is the most liquid security market in the world. For example, the bid-ask spread on a T-Bill is about 1/16 the bid-ask spread on the highly liquid IBM stock, and the T-Bills are an order of magnitude more liquid than the Notes of comparable maturity [see Amihud and Mendelsohn (1991)]. We do not know the depth behind the quotes used here, so we abstain from a detailed selection according to liquidity. But, for each day in our sample we group the Notes according to their original maturity and select the most recently issued Notes in each group. This is because the most recent issued (on-the-run) Notes are generally considered more liquid than older (off-the-run) issues.

In order to reduce the effects of having to interpolate the yield curve between 5 and 10 years we also select 2 bonds with remaining time to maturity close to but less than 7.5 years. Furthermore, in order to avoid interpolating the yield curve between 10 and 20 years we exclude the longest Notes (original maturity 20 years). Finally, for each day we include 12 T-Bills. Thus, for each day we have a sample of 24 bonds in the maturity interval from 0 to 10 years. Accordingly, we smooth the yield curve on the interval $[0, 10]$ only.

The next problem is the selection of an appropriate bandwidth. We have an automatic procedure based on minimizing the linearized GCV statistic. In principle we can choose a new value of h every day. However, for many applications it is better to use a suitable fixed bandwidth every day. One explanation is that choosing a new value of h every day adds some noise. For example, popular models of the term structure of interest rates [Hull-White (1990) and Heath-Jarrow-Morton (1992)], used in pricing interest-rate derivatives, require estimates of the volatility of forward interest rates. Needless to say that such volatility estimates can be severely affected by noise if forward rates are estimated from coupon bond data. At the risk of introducing some extra bias, the problem can be reduced by using the same fixed bandwidth for an extended period. Clearly, this may in certain periods give biased estimates, and the procedure requires a careful validation of the results. Based on visual inspection and experiments with automatic bandwidth selection we decided to $h = 3$ as a bandwidth for the full period.

The smoothing procedure also requires a discretization of the interval $[0, 10]$. This discretization is used among other things for evaluation of the integral in the smoothing equations (see sections 3.4 and 3.5). The density of this grid may affect the estimated yield and forward curves. Experiments with different densities indicate that there is no visible difference between the results when using 100 and 1000 grid points. In order to save computing time we use a discretization of $[0, 10]$ into 100 intervals.

Insert figures 1-8 about here

Figure 1 shows a time series plot of the estimated 0, 5, and 10 year yields. The estimation day was the 15th of each month (or the nearest trading day to the 15th). In order to assess the possible effects of using a fixed bandwidth we also show in figure 2 the evolution of (empirical) root mean squared errors (RMSE) [i.e., $\sqrt{\sum_{i=1}^{n} \widehat{\varepsilon}_i^2 / n}$] for each day in the period. Except for a peek in 1991 there is no systematic time series pattern in the errors. A more careful investigation revealed that the peak is due to an increased spread in the long end of the term structure. This is illustrated in figure 3, which shows a similar peak in the spread between the 10 and the 7.5 year yield. Whether this is due to a genuine increase in the spread or the result of an inappropriately chosen bandwidth will not be pursued here.

Figure 4 shows the estimated yield curve for October 15, 1996 together with associated 95% pointwise confidence intervals. For smaller maturities the confidence interval is about 20 basis points, while in the long end the error is considerably larger. The explanation, of course, is that all bonds in the sample convey information about the short end yields, while the long end is determined by a few noisy issues.

Figure 4 together with the plot of the forward curve in figure 5, indicates that there is a need for a variable that varies across maturities. This problem will, however, be left for future research.

Figure 6 shows the discount function for October 15, 1996 together with 95% confidence pointwise intervals. The discount function is much more smooth also for longer maturities than the yield and forward curves.

Figure 7 shows a plot of prices against fitted values for October 15, 1996. Except for a few outliers the model clearly fits data very well. The outliers may be removed by a more careful sample selection procedure that might include only the most recent issues. This will, however, not be pursued further.

Finally, figure 8 shows a plot of residuals against time to maturity. The graph clearly demonstrates how the variability in prices increases with time to maturity, thus, justifying the weighting procedure.

6 Extensions and Conclusions

We can use the nonparametric methods described above to provide a test of a parametric specification for d, y, or f. This type of analysis is the subject of Chapter 3 of Anderson et al. (1996). Define the following test statistic

$$T_n = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left\{ \widehat{d}(\tau_{ij}) - d_{\widehat{\theta}}(\tau_{ij}) \right\}^2 \pi(\tau_{ij}),$$

where $\widehat{d}$ is any of the above nonparametric estimators, $d_{\widehat{\theta}}$ is a parametric estimator, and π is a weighting function used to downplay boundary observations. Under reasonable conditions, $(T_n - \mu_n)/V_n^{1/2}$ is asymptotically standard normal under the null hypothesis that $d = d_{\theta_0}$ for some θ_0, where μ_n, V_n are the mean and variance of T_n. See Härdle and Mammen (1993) for an exposition of the theoretical issues involved.

Consideration of tax effects, coupon effects, and other anomalies have lead various authors to consider what amounts to a semiparametric modification of the arbitrage relationship (4). For example, McCullogh (1975) specified that

$$p \quad = \quad (1-t)c \sum_{j=1}^{m} d(\tau_j) + \{R - t_{CG}(R-p)\} d(\tau_m)$$

$$= \frac{(1-t)c\sum_{j=1}^{m} d(\tau_j)}{(1-t_{CG})d(\tau_m)} + R,$$

where c is the coupon payed for $j = 1, \ldots, m$ and R is the redemption value. He assumed that the capital gains rate t_{CG} is known but the income tax for the marginal investor t, is not known and must be estimated from the data along with the discount function. See also , see Jordan (1984) and Eom, Subrahmanyam, and Uno (1998). Motivated by this, consider the following model

$$p_i = \sum_{j=1}^{m_i} b_i(\tau_{ij}, Z_{ij}, \theta)d(\tau_{ij}) + \alpha^T X_i + \varepsilon_i, \tag{40}$$

where Z_{ij} and X_i are of covariates thought to affect the bond price. The payment function modification is often undertaken to account for taxes, where not all relevant tax rates are known with certainty and hence should be estimated from the data. Liquidity effects can also be modelled in this way. The additive effect of X has a more informal justification in terms of looking for systematically different pricing of some group of bonds like low coupons. A general estimation method that works here is to apply our nonparametric estimations for each parameter value θ, α and then optimize some profiled likelihood function, as for example in Klein and Spady (1994). We just outline the theory for the case where $b_i(\tau_{ij}, Z_{ij}, \theta) = b_i(\tau_{ij})$, which includes the Engle et al. (1986) model.

For any sequence u_i let $d_u(\cdot)$ denote the [sequence of] functions which minimize $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\{u_i - \sum_{j=1}^{m_i} b_i(\tau_{ij})\theta(\tau_{ij})\}^2$ with respect to functions $\theta(\cdot)$ [and let $\widehat{d}_u(\cdot)$ be the empirical version].[7] We take $u_i = p_i$ and $u_i = X_{li}$, respectively. This operation is a projection, i.e., is linear and idempotent, as discussed above, so that

$$\overline{p}_i = \sum_{j=1}^{m_i} b_i(\tau_{ij})d(\tau_{ij}) + \alpha^T \overline{X}_i, \tag{41}$$

where $\overline{p}_i = \sum_{j=1}^{m_i} b_i(\tau_{ij})d_p(\tau_{ij})$ and $\overline{X}_{li} = \sum_{j=1}^{m_i} b_i(\tau_{ij})d_{X_l}(\tau_{ij})$, while $\overline{X}_i$ is the vector of all $\overline{X}_{li}$. Subtracting (41) from (40) gives the regression equation $p_i - \overline{p}_i = \alpha^T(X_i - \overline{X}_i) + \varepsilon_i$. We now replace $\overline{p}_i$ and $\overline{X}_{li}$ by the estimates $\widehat{\overline{p}}_i = \sum_{j=1}^{m_i} b_i(\tau_{ij})\widehat{d}_p(\tau_{ij})$ and $\widehat{\overline{X}}_{li} = \sum_{j=1}^{m_i} b_i(\tau_{ij})\widehat{d}_{X_l}(\tau_{ij})$ constructed above and let

$$\widehat{\alpha} = \left[\sum_{i=1}^{n}(X_i - \widehat{\overline{X}}_i)(X_i - \widehat{\overline{X}}_i)^T\right]^{-1} \sum_{i=1}^{n}(X_i - \widehat{\overline{X}}_i)(p_i - \widehat{\overline{p}}_i).$$

[7] If u_i is random then the limit should be interpreted as the almost sure limit, which is assumed to exist.

This estimator is similar to the Robinson (1988) estimator of the slopes in partially linear regression except that we have replaced conditional expectation by a different type of projection. Having estimated α, we let $\widehat{d}_{\widehat{\alpha}}(\cdot)$ be $\widehat{d}(\cdot)$ applied to $p_i - \widehat{\alpha}^T X_i$. It is possible to establish the asymptotic properties of $\widehat{\alpha}$ under slightly stronger regularity conditions. Specifically, we must have that X_i does not itself lie in the linear space $\{\sum_{j=1}^{m_i} b_i(\tau_{ij})\theta(\tau_{ij}) \ : \ \theta \in L_2\}$. Furthermore, we must use undersmoothing to control the bias terms. We then have:

$$\sqrt{n}(\widehat{\alpha} - \alpha) \Longrightarrow N(0, \Omega^{-1}\Phi\Omega^{-1}), \tag{42}$$

where $\Phi = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \sigma_i^2 E[(X_i - \overline{X}_i)(X_i - \overline{X}_i)^T]$ and $\Omega = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n E((X_i - \overline{X}_i)(X_i - \overline{X}_i)^T]$. Furthermore, $\widehat{d}_{\widehat{\alpha}}(\cdot)$ has the same asymptotic distribution as $\widehat{d}_{\alpha_0}(\cdot)$. Finally, the asymptotic covariance matrix can be consistently estimated by $\widehat{\Omega}^{-1}\widehat{\Phi}\widehat{\Omega}^{-1}$ where $\widehat{\Phi} = \frac{1}{n}\sum_{i=1}^n \widehat{\varepsilon}_i^2 (X_i - \widehat{\overline{X}}_i)(X_i - \widehat{\overline{X}}_i)^T$ and $\widehat{\Omega} = \frac{1}{n}\sum_{i=1}^n (X_i - \widehat{\overline{X}}_i)(X_i - \widehat{\overline{X}}_i)^T$, where $\widehat{\varepsilon}_i = p_i - \sum_{j=1}^{m_i} b_i(\tau_{ij})\widehat{d}(\tau_{ij}) - \widehat{\alpha}^T X_i$ are residuals. The covariates can be discrete or continuous. This method can easily be extended to incorporate parametric or nonparametric heteroskedasticity in the errors.

7 Appendix: Proofs of Theorems

PROOF OF THEOREM 1. For simplicity we suppose that Condition A2 holds with $\kappa = 1$. The general case can be treated by slightly more complicated arguments. We start with the proof of (31). Suppose that A1 and A2 hold and that A4 holds for a finite interval S. We note first that for functions g with $\|g\|_2 < \infty$ one gets by application of the Cauchy Schwarz inequality from (22), see also (A4), that

$$\sup_{s\in S} |\int \widehat{H}_{LC}(s,t)g(t)\ dt| \le \sup_{s\in S} |\int \widehat{H}_{LC}(s,t)^2\ dt|^{1/2}\|g\|_2 \le C\|g\|_2. \tag{43}$$

Note now that

$$E\widehat{d}_{LC}(s) = \arg\min_u \sum_{i=1}^n \int \left\{\sum_{j=1}^{m_i} b_i(\tau_{ij})[d(\tau_{ij}) - u(s_{ij})]\right\}^2 \prod_{j=1}^{m_i}\{K_h(s_{ij} - \tau_{ij})\ ds_{ij}\} \tag{44}$$

and

$$\widehat{d}_{LC}(s) - E\widehat{d}_{LC}(s) = \arg\min_u \sum_{i=1}^{n} \int \left\{ \varepsilon_i - \sum_{j=1}^{m_i} b_i(\tau_{ij}) u(s_{ij}) \right\}^2 \prod_{j=1}^{m_i} \{K_h(s_{ij} - \tau_{ij})\ ds_{ij}\} . \tag{45}$$

For the proof of (31) we show first that

$$E\widehat{d}_{LC}(s) - d(s) = \beta_n^{LC}(s). \tag{46}$$

For the proof of (46) note first that equation (44) implies that

$$E\widehat{d}_{LC}(s) - d(s) = \arg\min_u \sum_{i=1}^{n} \int \left\{ \sum_{j=1}^{m_i} b_i(\tau_{ij})[d(\tau_{ij}) - u(s_{ij}) - d(s_{ij})] \right\}^2 \prod_{j=1}^{m_i} \{K_h(s_{ij} - \tau_{ij})\ ds_{ij}\} .$$

This minimization can be solved like in Section 3.1. This gives

$$\begin{aligned} 0 \;=\; & \sum_{i=1}^{n} \sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(s - \tau_{ir})[d(\tau_{ir}) - d(s)] \\ & + \sum_{i=1}^{n} \sum_{r=1}^{m_i} \sum_{\substack{j=1 \\ r \neq j}}^{m_i} b_i(\tau_{ir}) b_i(\tau_{ij}) K_h(s - \tau_{ir}) \int K_h(t - \tau_{ij})[d(\tau_{ij}) - d(t)]\ dt \\ & - \sum_{i=1}^{n} \sum_{r=1}^{m_i} \sum_{\substack{j=1 \\ r \neq j}}^{m_i} b_i(\tau_{ir}) b_i(\tau_{ij}) K_h(s - \tau_{ir}) \int K_h(t - \tau_{ij}) \left[E\widehat{d}_{LC}(t) - d(t) \right]\ dt \\ & - \left[E\widehat{d}_{LC}(s) - d(s) \right] \sum_{i=1}^{n} \sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(s - \tau_{ir}). \end{aligned}$$

This equation can be written as

$$E\widehat{d}_{LC}(s) - d(s) = \beta_n^{*,LC}(s) + \int \widehat{H}(s,t) \left[E\widehat{d}_{LC}(t) - d(t) \right]\ dt.$$

Iterative application of this equation gives (46) because of (A2).

For the study of the stochastic component $\widetilde{d}_{LC}(s) = \widehat{d}_{LC}(s) - E\widehat{d}_{LC}(s)$ we get from (45) that

$$\widetilde{d}_{LC}(s) = \sum_{k=0}^{\infty} \widehat{\mathcal{H}}_{LC}^k \widetilde{d}_{LC}^*(s),$$

where

$$\widetilde{d}^*_{LC}(s) = \frac{\sum_{i=1}^n \sum_{r=1}^{m_i} \varepsilon_i b_i(\tau_{ir}) K_h(s-\tau_{ir})}{\sum_{i=1}^n \sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(s-\tau_{ir})}.$$

We show now that

$$\sup_{s\in S} \left|\sum_{k=1}^{\infty} \widehat{\mathcal{H}}^k_{LC} \widetilde{d}^*_{LC}(s)\right| = o_P((nh)^{-1/2}) \tag{47}$$

holds. This shows

$$\sup_{s\in S} |\widetilde{d}_{LC}(s) - \widetilde{d}^*_{LC}(s)| = o_P((nh)^{-1/2}). \tag{48}$$

Therefore, (46) and (47) imply (31).

For the proof of claim (47) we will show

$$\sup_{s\in S} |\widehat{\mathcal{H}}^k_{LC} \widetilde{d}^*_{LC}(s)| \le \rho^k R_n, \tag{49}$$

where $\rho < 1$ is introduced in Condition (A2) and where R_n is a random variable with $R_n = o_P((nh)^{-1/2})$. This implies (47) because of

$$\sup_{s\in S} \left|\sum_{k=1}^{\infty} \widehat{\mathcal{H}}^k_{LC} \widetilde{d}^*_{LC}(s)\right| \le \sum_{k=1}^{\infty} \rho^k R_n = R_n/(1-\rho) = o_P((nh)^{-1/2}).$$

For the proof of (49) we will show that

$$\begin{aligned} \|\widehat{\mathcal{H}}_{LC} \widetilde{d}^*_{LC}\|_2 &= o_P((nh)^{-1/2}), \qquad (50)\\ \sup_{s\in S} |\widehat{\mathcal{H}}_{LC} \widetilde{d}^*_{LC}(s)| &= o_P((nh)^{-1/2}). \qquad (51) \end{aligned}$$

We argue now that (50) and (51) imply (49). Note that (A2) implies

$$\|\widehat{\mathcal{H}}^k_{LC} \widetilde{d}^*_{LC}\|_2 \le \rho^{k-1} \|\widehat{\mathcal{H}}_{LC} \widetilde{d}^*_{LC}\|_2$$

for $k \ge 1$. Because of (43) this implies

$$\sup_{s\in S} |\widehat{\mathcal{H}}^k_{LC} \widetilde{d}^*_{LC}(s)| \le C \|\widehat{\mathcal{H}}^{k-1}_{LC} \widetilde{d}^*_{LC}\|_2 \le \rho^k R_n$$

for $k \geq 2$ with $R_n = \|\widehat{\mathcal{H}}_{LC}\widetilde{d}^*_{LC}\|_2/\rho^2$. With (50) and (51) we get (49). So for the proof of (31) it remains to show (50) and (51). For the proof of (50) we get using (A1) and (23)

$$\begin{aligned}
E\|\widehat{\mathcal{H}}_{LC}\widetilde{d}^*_{LC}\|_2^2 &= E\,\frac{1}{n^2}\sum_{i,j=1}^{n}\int \widehat{H}_{LC}(s,t)\widehat{H}_{LC}(s,u)R_i(t)R_j(u)\;du\;dt\;ds\;\varepsilon_i\varepsilon_j \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\int \widehat{H}_{LC}(s,t)\widehat{H}_{LC}(s,u)R_i(t)R_i(u)\;du\;dt\;ds\;\sigma_i^2 \\
&\leq \frac{1}{n^2}\sum_{i=1}^{n}\int\left|\int \widehat{H}_{LC}(t,u)R_i(u)\;du\right|^2\;dt\;C \\
&= o((nh)^{-1}).
\end{aligned}$$

This shows (50). For the proof of (51) we write

$$\widehat{\mathcal{H}}_{LC}\widetilde{d}^*_{LC}(s) = \frac{1}{n}\sum_{i=1}^{n} r_i(s)\varepsilon_i,$$

where

$$r_i(s) = \int \widehat{H}_{LC}(s,u)R_i(u)\;du.$$

Note first that for $s \in S$ and $a, c > 0$ we have

$$\begin{aligned}
&\Pr(\;(nh)^{1/2}[\widehat{\mathcal{H}}_{LC}\widetilde{d}^*_{LC}(s)] > c) \\
&\quad\leq\; E\;\exp\left(\frac{a\log(n)}{c}\left[(nh)^{1/2}\widehat{\mathcal{H}}_{LC}\widetilde{d}^*_{LC}(s) - c\right]\right) \\
&\quad\leq\; n^{-a}\prod_{i=1}^{n} E\;\exp\left((a/c)\log(n)(h/n)^{1/2}r_i(s)\varepsilon_i\right),
\end{aligned}$$

where the independence of the error variables ε_i has been used. We use now $\exp(x) \leq 1 + x + (1/2)x^2[1+\exp(x)]$. With $d_i(s) = (a/c)\log(n)(h/n)^{1/2}r_i(s)\varepsilon_i$ this gives the following upper bound for the last term

$$\begin{aligned}
&\leq\; n^{-a}\prod_{i=1}^{n}\left\{1 + E\;\left(d_i(s)\right) + E\;\left(d_i(s)\right)^2\left[1+\exp\left(d_i(s)\right)\right]\right\} \\
&=\; n^{-a}\prod_{i=1}^{n}\left\{1 + E\;\left(d_i(s)\right)^2\left[1+\exp\left(d_i(s)\right)\right]\right\}.
\end{aligned}$$

We use now that for a constant D the following inequality holds

$$E\ \varepsilon_i^2\left[1+\exp\left(d_i(s)\right)\right] \leq D.$$

This follows from the uniform bound for the Laplace transform of ε_i, see (27), and from

$$\sup_{s\in S}(a/c)\log(n)(h/n)^{1/2}|r_i(s)| \to 0,$$

see (24). In a next step we apply $1+x \leq \exp(x)$. This gives

$$\begin{aligned}
&\Pr(\ (nh)^{1/2}[\widehat{\mathcal{H}}_{LC}\widetilde{d}^*_{LC}(s)] > c) \\
&\leq\ n^{-a}\ \prod_{i=1}^{n}\left[1+D\left((a/c)\log(n)(h/n)^{1/2}r_i(s)\right)^2\right] \\
&\leq\ n^{-a}\ \exp\left[\sum_{i=1}^{n}D\left((a/c)\log(n)(h/n)^{1/2}r_i(s)\right)^2\right] \\
&\leq\ n^{-a}(1+o(1)),
\end{aligned}$$

where in the last step (23) has been used, see also Condition A4. With the same arguments one gets a bound for $\Pr(\ (nh)^{1/2}[\widehat{\mathcal{H}}_{LC}\widetilde{d}^*_{LC}(s)] < -c)$. This gives that for $a, c > 0$

$$\Pr(\ (nh)^{1/2}|\widehat{\mathcal{H}}_{LC}\widetilde{d}^*_{LC}(s)| > c) \leq n^{-a}(2+o(1)).$$

This implies that for all constants $b > 0$ and finite sets $S_n \subset S$ with n^b elements we have that

$$\sup_{s\in S_n}|\widehat{\mathcal{H}}_{LC}\widetilde{d}^*_{LC}(s)| = o_P((nh)^{-1/2}). \tag{52}$$

Because ε_i has a bounded Laplace transform [see (27)] it holds that $\sup_{1\leq i\leq n}|\varepsilon_i| = O_P(\log(n))$. Because of assumption (28) this implies that for a constant C' large enough

$$\sup_{s\in S}|\frac{\partial}{\partial s}\widehat{\mathcal{H}}_{LC}\widetilde{d}^*_{LC}(s)| = o_P(n^{C'}).$$

Therefore for an appropriate choice of S_n we can achieve that

$$\sup_{s\in S}\inf_{t\in S_n}|\widehat{\mathcal{H}}_{LC}\widetilde{d}^*_{LC}(s) - \widehat{\mathcal{H}}_{LC}\widetilde{d}^*_{LC}(t)| = o_P((nh)^{-1/2}). \tag{53}$$

Claim (51) follows from (52) and (53). This shows (31). Note that the above argument can be extended to allow for only a finite number of moments on ε_i using a truncation argument, provided stronger assumptions on the operator $\widehat{\mathcal{H}}_{LC}$ are made. We come now to the proof of (30). Because of (46) we have to show

$$\sqrt{nh}\widetilde{d}_{LC}(s)s_n^{LC}(s)^{-1} \Longrightarrow N(0,1) \tag{54}$$

in distribution. We argue now that

$$\widetilde{d}_{LC}(s) - \widetilde{d}^*_{LC}(s) = o_P((nh)^{-1/2}). \tag{55}$$

This follows as in the proof of (48) where now $S = \{s\}$. Note that now, because S is finite, the proof of (51) follows immediately from (23) and the boundedness of σ_i^2, see (A1). Because of (55) and (26) For the proof of (54) it suffices to show

$$\sqrt{nh}\widetilde{d}^*_{LC}(s)s_n^{LC}(s)^{-1} \Longrightarrow N(0,1) \tag{56}$$

in distribution. Note that

$$\widetilde{d}^*_{LC}(s) = \sqrt{h/n}\sum_{i=1}^{n} R_i(s)\varepsilon_i.$$

Now, claim (56) follows with a standard version of the central limit theorem from (25) and the existence of a uniform bound for $E\varepsilon_i^{2+\delta}$, see (A1). This finishes the proof of Theorem 1. ■

Proof of Theorem 2. Claims (37) and (38) can be shown as in the proof of Theorem 1. Under the additional assumption of B3 one gets that $M(s)$ is asymptotically equivalent to a diagonal matrix with diagonal elements: $n^{-1}\sum_{i=1}^n\sum_{r=1}^{m_i} b_i(\tau_{ir})^2K_h(s-\tau_{ir})$ and $n^{-1}\sum_{i=1}^n\sum_{r=1}^{m_i} b_i(\tau_{ir})^2K_h(s-\tau_{ir})[(\tau_{ir}-s)/h]^2$. Using expansion (38) for $S=\{s\}$ one gets that $\widehat{d}_{LL}(s)$ has the same asymptotic variance as $\widehat{d}_{LC}(s)$. ■

Sketch Proof of (42). It suffices to show that

$$\frac{1}{n}\sum_{i=1}^{n}\{(X_i-\widehat{\overline{X}}_i)(X_i-\widehat{\overline{X}}_i)^T-(X_i-\overline{X}_i)(X_i-\overline{X}_i)^T\}\xrightarrow{p}0 \tag{57}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\overline{X}_i-\widehat{\overline{X}}_i)\varepsilon_i\xrightarrow{p}0 \tag{58}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-\widehat{\overline{X}}_i)\left[(\widehat{\overline{p}}_i-\overline{p}_i)-\alpha_0^T(\widehat{\overline{X}}_i-\overline{X}_i)\right]\xrightarrow{p}0, \tag{59}$$

since $\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_i)(X_i-\overline{X}_i)^T\}\xrightarrow{p}\Omega$ and $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-\overline{X}_i)\varepsilon_i\Rightarrow N(0,\Phi)$ by standard laws of large numbers and central limit theorems for independent random variables respectively. The proof of (57)-(59) is similar to Robinson (1988). Clearly, (57) follows directly from the uniform expansion for $\widehat{d}$. The main detail in the proof of (58) concerns the quadratic form

$$\sum_{i=1}^{n}\sum_{\ell=1}^{n}\overline{w}_{i\ell}\varepsilon_i\eta_\ell, \tag{60}$$

where $\eta_\ell=X_\ell-\overline{X}_\ell$ and

$$\overline{w}_{i\ell}=\sum_{j=1}^{m_i}b_i(\tau_{ij})\frac{\sum_{r=1}^{m_i}b_i(\tau_{\ell r})K_h(\tau_{ij}-\tau_{\ell r})}{\sum_{\ell=1}^{n}\sum_{r=1}^{m_i}b_\ell(\tau_{\ell r})^2K_h(\tau_{ij}-\tau_{\ell r})}.$$

The magnitude of (60) follows from standard kernel arguments. The bias terms in (59) are of order $\sqrt{n}h^4$. ■

REFERENCES

Amihud, Y. and Mendelsohn, H. (1991) Liquidity, Maturity, and the Yields on U.S. Treasury Securities, The Journal of Finance, 46, 1411-1425.

Anderson, N., Breedon, F., Deacon, M., Derry, A. and Murphy, G. (1996). Estimating and interpreting the yield curve. John Wiley Series in Financial Economics and Quantitative Analysis. Chichester.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and J. A. Wellner (1993). Efficient and adaptive estimation for semiparametric models. The John Hopkins University Press, Baltimore and London.

Campbell, J.Y., Lo, A.W. and MacKinlay, A.C. (1997). The econometrics of financial markets. Princeton University Press.

Campbell, J.Y., and Shiller, R.J. (1991). Yield spreads and interest rate movements: A birds eye view. Review of Economic Studies 58, 495-514.

Chambers, D. R., Carleton, W. T. and Waldman, D. W. (1984), A New Approach to Estimation of the Term Structure of Interest Rates, Journal of Financial and Quantitative Analysis, 19, 233-269.

Chesher, A. (1997). Diet Revealed: Semiparametric Estimation of Nutrient Intake-Age Relationships. Journal of the Royal Statistical Society, Series A. 160, 389-428.

Dahlquist, M., and Svensson, L.E.O. (1996). Estimating the term structure of interest rates for monetary policy analysis. Scandinavian Journal of Economics 98, 163-183.

Engle, R.F., Granger, C.W.J., Rice, J. and Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales, Journal of the American Statistical Association, 81, 310-320.

Engsted, T. and Tanggaard, C. (1995). The predictive power of yield spreads for future interest rates: Evidence from the Danish term structure. Scandinavian Journal of Economics 97, 145-159.

Eom, Y. H., Subrahmanyam, M., and Uno, J. (1998). Coupon effects and the pricing of Japanese government bonds: an empirical analysis. Journal of Fixed Income, September, 69-86

Estrella, A. and Mishkin, F. S. (1998). Predicting U.S. recessions: financial variables as leading indicators. Review of Economics and Statistics 80, 45-61.

Fan, J. (1992). Design-Adaptive Nonparametric Regression. Journal of the American Statistical Association, 87, 998-1004.

Fisher, M. E., Nychka, D. and Zervos, D. (1995). Fitting the term structure of interest rates with smoothing splines. Federal Reserve Bank Finance and Economics Discussion paper no 95-1.

Frankel, J. A. and Lown, C. S. (1994). An indicator of future inflation extracted from the steepness of the interest rate yield curve along its entire length. Quarterly Journal of Economics 109, 517-530.

Gozalo, P. L., and Linton, O. B. (1998). Local nonlinear least squares. Cowles Foundation Discussion Paper no 1075.

Härdle, W. (1990). Applied Nonparametric Regression. Cambridge University Press

Härdle, W. and Mammen, E. (1993). Comparing Nonparametric versus Parametric Regression Fits. Annals of Statistics, 21, 1926-1947.

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

Hausman, J. A., and Newey, W. K. (1995). Nonparametric estimation of exact consumers surplus and deadweight loss. Econometrica 63, 1445-1476.

Heath, D., R. Jarrow, and A. Morton (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. Econometrica 60, 77-105.

Horowitz, J.L. (1998). Semiparametric methods in econometrics. Springer-Verlag: Berlin.

Hull, J., and A. White (1990). Pricing interest rate derivaitve securities. Review of Financial Studies 3, 573-592.

Jordan, J. V. (1984). Tax effects in term structure estimation. Journal of Finance 39, 393-406.

Klein, R. W. and Spady, R. H. (1994). An Efficient Semiparametric Estimator for Binary Choice Models. Econometrica 61, 387-421.

Lighthill, M. J. (1958). Introduction to Fourier Analysis and Generalized Functions. Cambridge University Press, London.

Linton, O. B. (1998). Efficient estimation of generalized additive nonparametric regression models. To appear, maybe, in Econometric Theory.

Linton, O. B., Mammen, E. and Nielsen, J. P. (1997). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. Available at http:/www.econ.yale.edu/~linton.

McCullough, J. H. (1971). Measuring the term structure of interest rates. The Journal of Business, 44, 19-31.

McCullough, J. H. (1975). The Tax-Adjusted Yield Curve. The Journal of Finance, 30, 811-830.

Mammen, E. (1991). Nonparametric regression under qualitative smoothness assumptions. The Annals of Statistics 19, 741-759.

Matzkin, R.L. (1994). Restrictions of economic theory in nonparametric methods. in the Handbook of Econometrics volume 4, eds. R.F. Engle and D.F. McFadden. pp 2523-2558.

Nelson, C.R. and Siegel, A. F. (1987). Parsimonious Modelling of Yield Curves, Journal of Business, 60, 473-489.

Newey, W. K. and Powell, J. L. (1988). Instrumental variables estimation for nonparametric regression models. Manuscript.

Nielsen, J. P. and Linton, O. B. (1998). An optimization interpretation of integration and backfitting estimators for separable nonparametric models. Journal of the Royal Statistical Society, Series B, 60, 217-222.

Nychka, D., Wahba, G., Goldfarb, S. and Pugh, T. (1984). Cross-validated spline methods for the estimation of the three-dimensional tumor size distributions from observations on two-dimensional cross-sections. Journal of the American Statistical Association, 79, 832-846.

Opsomer, J. D. and D. Ruppert (1997). Fitting a bivariate additive model by local polynomial regression. Annals of Statistics 25, 186 - 211.

O'Sullivan, F. (1986). Ill posed inverse problems (with discussion). Statistical Science 4, 503-527.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992). Numerical Recipes in C, 2. ed., Cambrdige University Press

Riesz, F. and Sz.-Nagy, B. (1990). Functional Analysis. Dover, New York.

Robinson, P. M. (1988). Root-N-Consistent Semiparametric Regression. Econometrica, 56, 931-954.

Rust, J. (1997). Using randomization to break the curse of dimensionality. Econometrica 65, 487-516.

Schaefer, S. M. (1981). Measuring a tax-specific term structure of interest rates in the market for British government securities. The Economic Journal, 91, 415-438.

Shea, G. S. (1984). Pitfalls in Smoothing Interest Rate Term Structure Data: Equilibrium Models and Spline Approximations. Journal of Financial and Quantitative Analysis, 19, 253-269.

Svensson, L. E. O. (1994). Estimating and Interpreting Forward Interest Rates: Sweden 1992-4, CEPR Discussion Paper 1051.

Tanggaard, C. (1992). Kernel smoothing of discount functions. Aarhus School of Business Working paper no 92-8.

Tanggaard, C. (1997). Nonparametric smoothing of Yield Curves. Review of Quantitative Finance and Accounting 9, 251-267.

Thisted, R. A. (1988). Elements of Statistical Computing, Chapman and Hall Ltd, London.

Tsybakov, A. B. (1986). Robust reconstruction of functions by the local approximation method. Probability and Information Transmission, 22, 133–46.

Vasicek, O.A. and Fong, H.G. (1982). Term Structure Modelling using Exponential Splines. The Journal of Finance, 37, 339-348.

Wahba, G. (1990). Spline Models for Observational Data. CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia.

Weinstock, R. (1951). Calculus of variations. Dover. New York.

Figures

Figure 1. Estimated yields, November 1990 - October 1996.

Figure 2. Root mean squared error, November 1990 - October 1996.

Figure 3. Yield curve spread, November 1990 - October 1996.

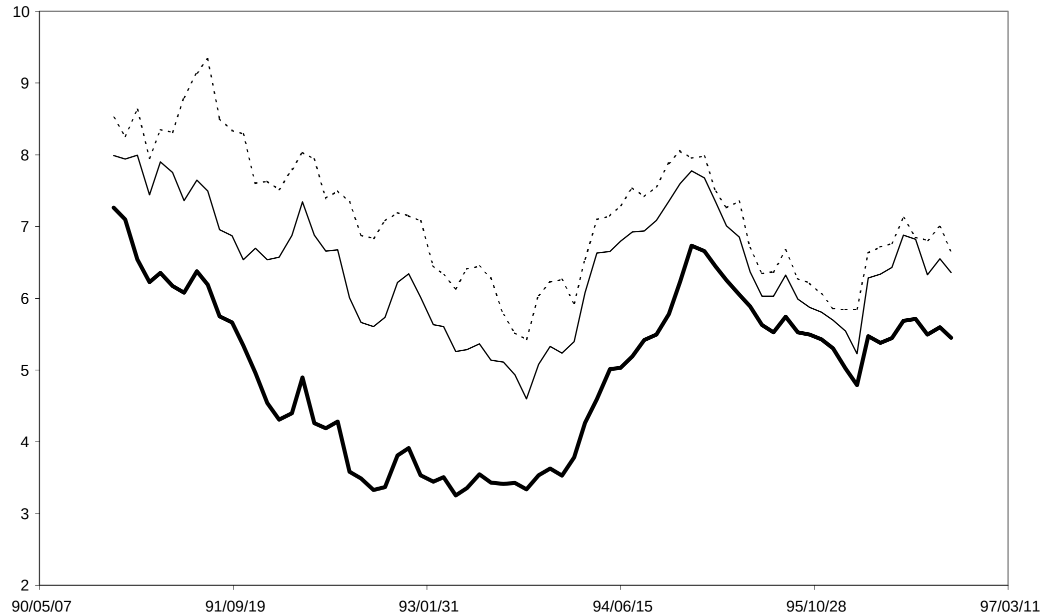
Figure 4. Yield curve with 95% confidence intervals, October 15, 1996.

Figure 5. Yield and forward curves, October 15, 1996.

Figure 6. Discount function with 95% confidence intervals, October 15, 1996.

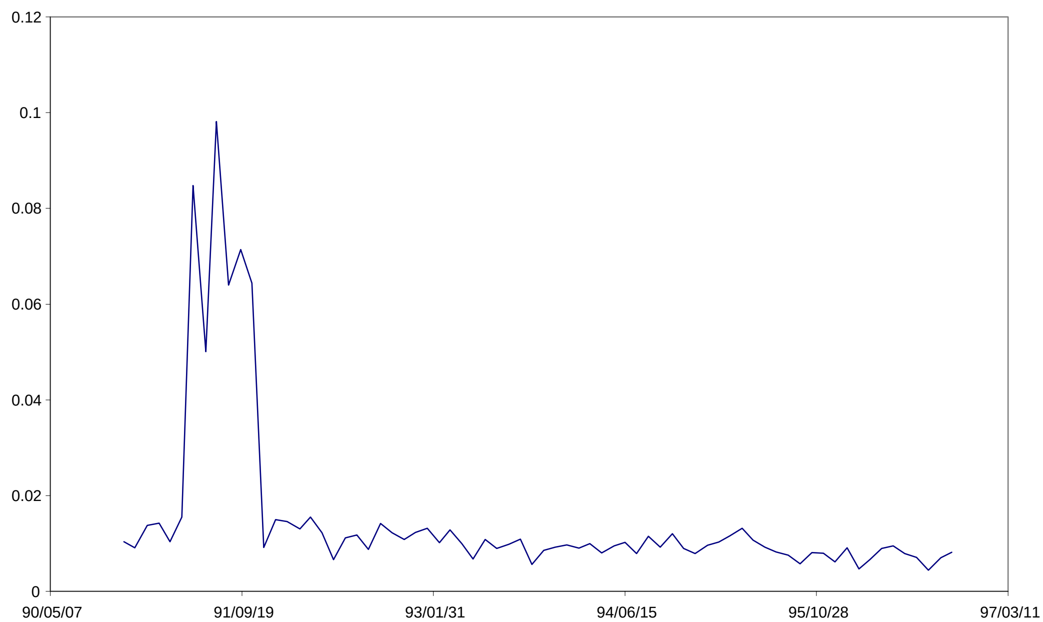
Figure 7. Discounted values against price, October 15, 1996.

Figure 8. Residuals against time to maturity, October 15, 1996.

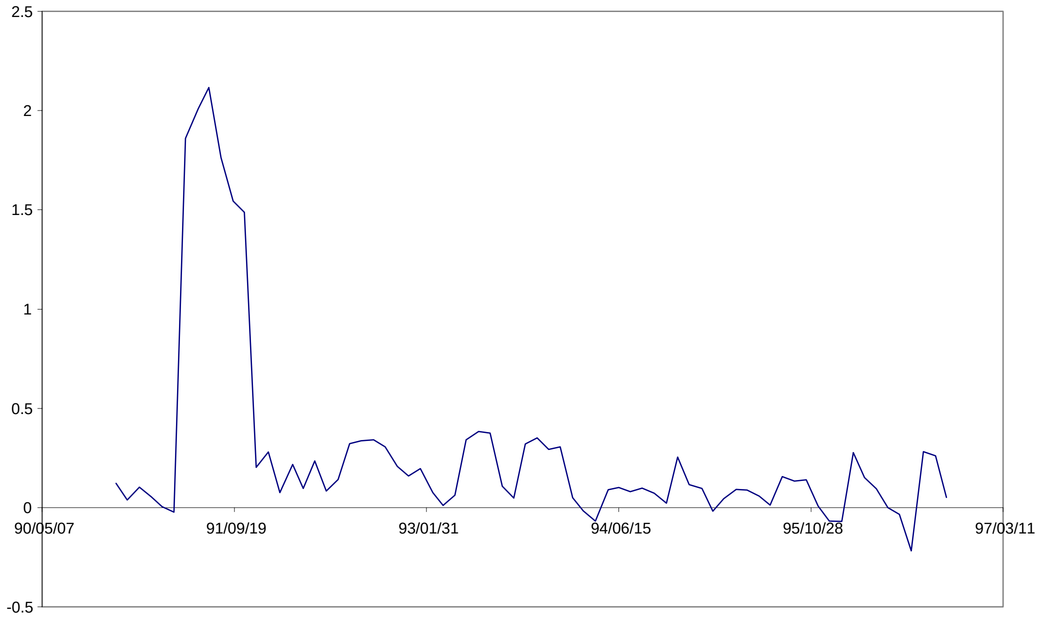


Estimated yields, November 1990 - October 1996.

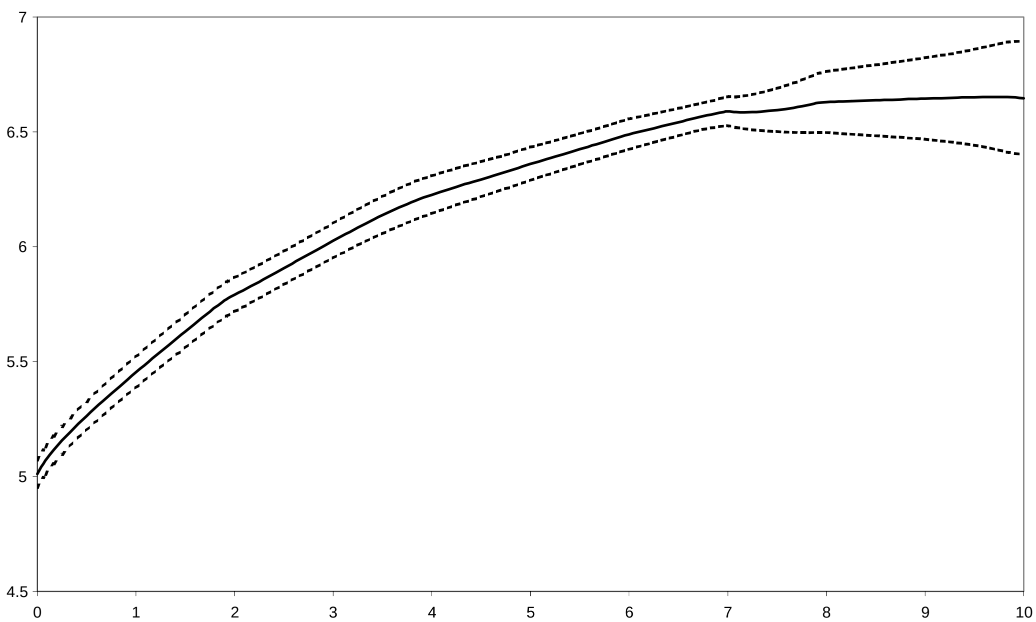
Note: The thick solid line is the 0 year yield, the thin solid line is the 5 year yield, while the dashed line is the 10 year yield.



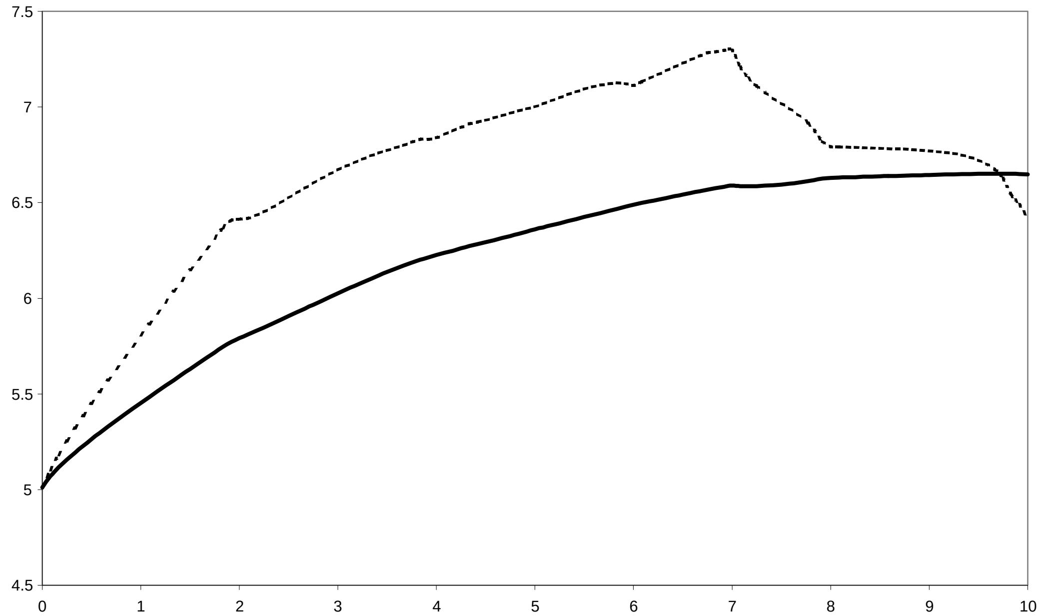
Root mean squared error, November 1990 - October 1996.



Yield curve spread, November 1990 - October 1996.
Note: The spread is between the 10 and the 7.5 year yield.

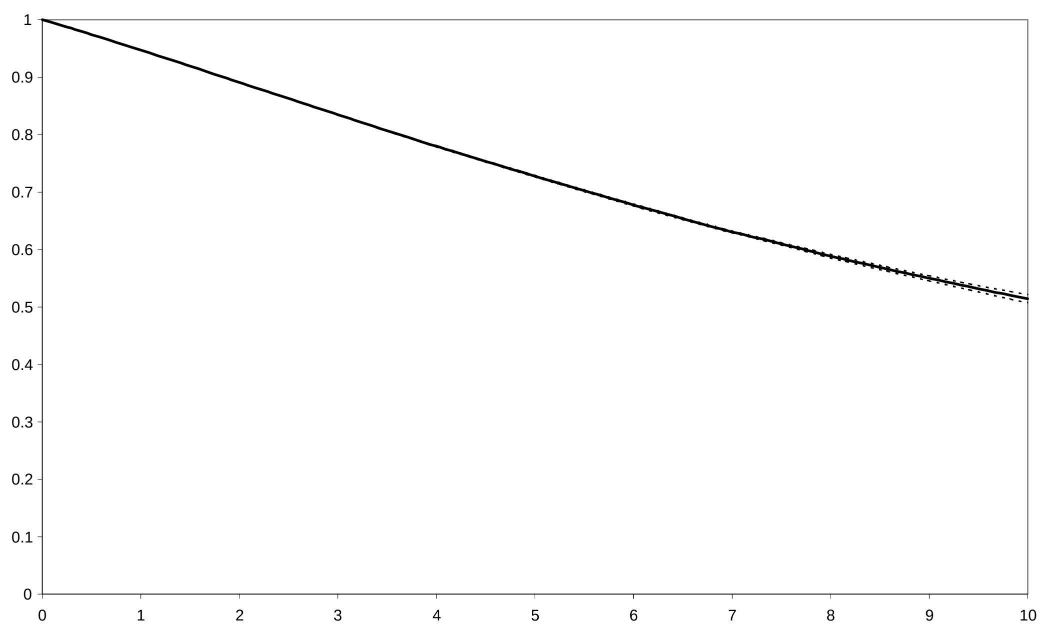


Yield curve with 95% confidence intervals, October 15, 1996.

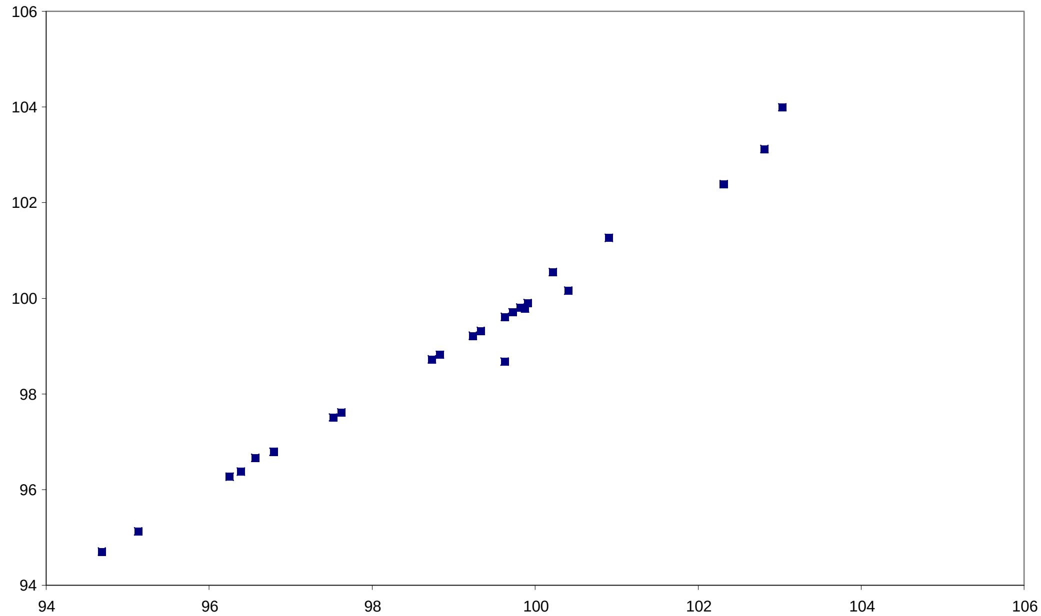


Yield and forward curves, October 15 1996.

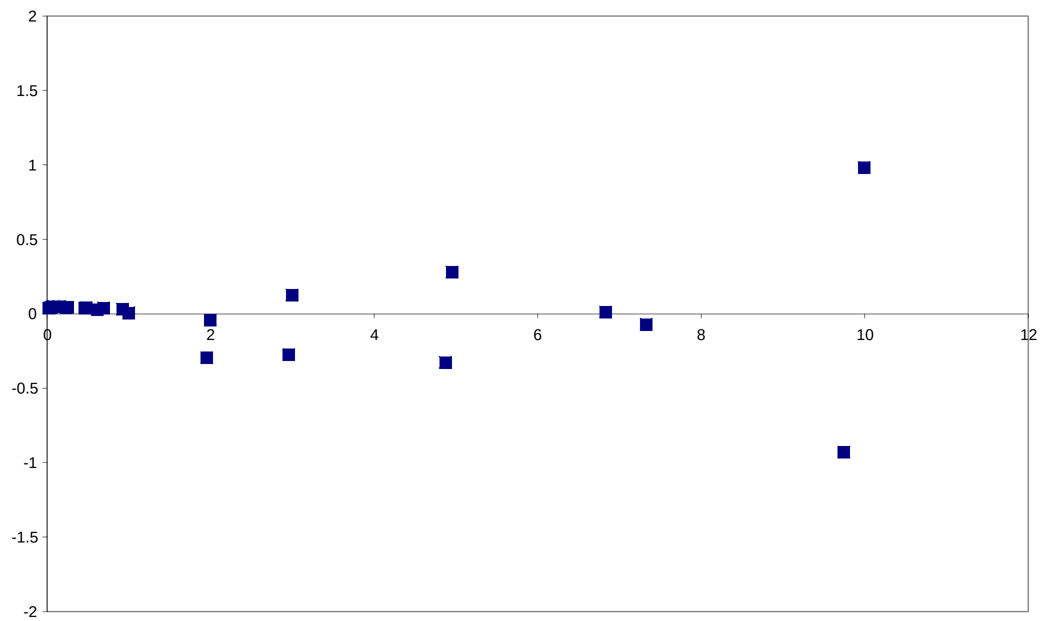
Note: The solid line is the yield curve. The dashed line is the forward curve.



Discount function with 95% confidence intervals, October 15, 1996.



Discounted values against price, October 15, 1996.



Residuals against time to maturity, October 15, 1996.