YIELD CURVE ESTIMATION BY KERNEL SMOOTHING METHODS

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

April 3, 2000

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. We investigate the finite sample performance of our method, in comparison with other well-established methods, in a small simulation experiment.

Some key words: Coupon bonds; Kernel Estimation; Hilbert Space; Nonparametric regression; Term structure estimation; Yield curve; Zero coupon.
Journal of Economic literature classification: C14, G12

[01]LSE and Yale University. Address for correspondence: Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: lintono@lse.ac.uk. Web page: http://econ.lse.ac.uk/staff/olinton/. Research was supported by the National Science Foundation, the Danish Social Science Research Council, and the North Atlantic Treaty Organization. Thanks go to Frank Diebold and two referees, to Olaf Bunke, Greg Connor, Stewart Hodges, Andrew Jeffrey, Thong Nguyen, and Peter Phillips for helpful comments.

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 the 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, 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).[1]

[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]Codan, 60 Gammel Kongevej, DK-1790 Copenhagen V, Denmark. Email: npj@codan.dk. Research is supported by Bergiafonden and the Danish Social Science Research Council.

[4] Department of Finance, The Aarhus School of Business, 4 Fuglesangs All, DK-8210 Aarhus V, Denmark. E-mail: cat@hha.dk. Research is supported by the Danish Social Science Research Council.

[1]Some recent studies support this, for example Amihud and Mendelson (1991). They found that close to maturity

The statistical problem we address is to estimate the function $d(\cdot)$ from a sample of coupon paying bonds. 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.

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 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. 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. We propose a new class of methods for estimating $d(\cdot)$ based 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 first interpret the function d as the solution of some population mean squared error criterion. We then smooth the sample version of this to obtain an empirical criterion function,

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.

which is regular enough to provide consistent estimates. 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.[2] We also give a 'backfitting' interpretation to our procedure, as in Opsomer and Ruppert (1997) and Mammen, Linton and Nielsen (1999). We establish the convergence of our iterative scheme and establish the asymptotic properties of the estimator. We obtain a representation of our estimator as an infinite series with declining coefficients, which thereby provides 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 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. With

[2]See Rust (1997) for discussion of some alternative methods.

minor modifications we can provide new estimators in all these situations and find their asymptotic distribution.

In section 2 we discuss [for reasons of completeness] smoothing of pure discount bonds. 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, i.e., the object of interest is the discount function, while subsection 3.2 gives the local linear extension. In subsection 3.3 we describe the local constant exponential version and its one-step approximation. We present the asymptotic properties of our methods in section 4. In section 5 we provide a small simulation study that compares our method with some alternatives. Proofs are given in the appendix. In the sequel all integrals are Lebesgue integrals, and run from $-\infty$ to $+\infty$ unless otherwise stated.

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 are an error 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 any 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

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

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 $\widehat{d}(\tau)$ minimize $Q_n(\theta)$ with respect to θ. 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; we shall also suppose throughout that $K \geq 0$. 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) when ε_i are independent.

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 is necessary because not all time points of interest have some payment coming due. 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.[3] 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 a random sequence satisfying $E[\varepsilon_i] = 0$, $i = 1, \ldots, n$. To our knowledge, all previous treatments of this problem have assumed that ε_i are independent random variables, possibly heterogenous in distribution. We shall allow for correlation across i [i.e., across bonds]; this is because it is expected that there is a common source for some of the pricing errors which make the error terms. However, we will need to assume that the correlation is weak enough for a central limit theorem to be applied.

[3] 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.

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.[4] 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 any minimizer [with respect to $\theta(\cdot)$] of the limiting criterion function

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

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

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

[4]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.

and let $\widehat{\theta}(\cdot)$ minimize this criterion with respect to functions $\theta(\cdot) \in \Theta$, where Θ is the class of all functions for which the resulting integral is well defined.[5] 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.[6] In our case, however, it is essential to start with the global criterion function.

We define the following localized sample criterion function

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

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

$$\sum_{i=1}^{n} \{p_i - \sum_{j=1}^{m_i} b_i(\tau_{ij}) \theta(s_{ij})\}^2 \prod_{l=1}^{m_i} K_h(s_{il} - \tau_{il}) \tag{8}$$

is preferentially weighting bonds whose maturities $\tau_{i1}, \ldots, \tau_{im_i}$ are close to the sequence $\{s_{il}\}$. Although the local criterion (8) does impose the requirement that $\theta(\cdot)$ is constant in a small neighbourhood 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

[5] For positive K, the integral is well-defined (possibly equal to infinity) for all measurable functions $\theta(\cdot)$.

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

single point. The averaging in (7) imposes the necessary structure.[7] We define our estimator $\widehat{d}_{LC}(\cdot)$ as any minimizer of $Q_n(\theta)$ with respect to $\theta(\cdot) \in \Theta$. 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

$$\begin{aligned} 0 &= \sum_{i=1}^{n}\sum_{r=1}^{m_i}\int\left\{p_i - \sum_{j=1}^{m_i}\mathbf{1}[j\neq r]b_i(\tau_{ij})\widehat{d}_{LC}(s_{ij}) \; - b_i(\tau_{ir})\widehat{d}_{LC}(s)\right\} \\ &\quad b_i(\tau_{ir})\prod_{j=1}^{m_i}\mathbf{1}[j\neq r]\left\{K_h(s_{ij}-\tau_{ij})\; ds_{ij}\right\}K_h(s-\tau_{ir}) \\ &= \sum_{i=1}^{n}\sum_{r=1}^{m_i}p_i b_i(\tau_{ir})K_h(s-\tau_{ir}) - \sum_{i=1}^{n}\sum_{r=1}^{m_i}\sum_{\substack{j=1\\ j\neq r}}^{m_i} b_i(\tau_{ij})b_i(\tau_{ir})K_h(s-\tau_{ir})\int \widehat{d}_{LC}(t)K_h(t-\tau_{ij})\; dt \\ &\quad -\sum_{i=1}^{n}\sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(s-\tau_{ir})\widehat{d}_{LC}(s), \end{aligned} \tag{9}$$

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

$$\widehat{d}_{LC}(s) = \overline{d}(s) + \int \widehat{H}_{LC}(s,t)\widehat{d}_{LC}(t)dt, \tag{10}$$

where

$$\overline{d}(s) = \frac{\sum_{i=1}^{n}\sum_{r=1}^{m_i} p_i b_i(\tau_{ir})K_h(s-\tau_{ir})}{\sum_{i=1}^{n}\sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(s-\tau_{ir})}, \tag{11}$$

$$\widehat{H}_{LC}(s,t) = -\frac{\sum_{i=1}^{n}\sum_{r=1}^{m_i}\sum_{j\neq r}^{m_i} b_i(\tau_{ir})b_i(\tau_{ij})K_h(s-\tau_{ir})K_h(t-\tau_{ij})}{\sum_{i=1}^{n}\sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(s-\tau_{ir})}, \tag{12}$$

The expression (10) is a linear integral equation involving the intercept $\overline{d}$ and the linear operator $\widehat{H}_{LC}$, both of which quantities are fairly simple functions of the data [ratios of sample averages of kernel weighted data]. This structure is important in deriving the asymptotic properties of $\widehat{d}_{LC}$.

[7] Note also that by a change of variable

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

which is asymptotically equivalent to (5).

Relation (10) suggests the following iteration for the calculation of $\widehat{d}_{LC}$:

$$\widehat{d}_{LC}^{[a+1]}(s) = \overline{d}(s) + \int \widehat{H}_{LC}(s,t)\widehat{d}_{LC}^{[a]}(t)dt, \quad a = 0, 1, \ldots, \tag{13}$$

where the starting value $\widehat{d}_{LC}^{[0]}$ is equal to $\overline{d}$, say. The integration in (13) can be computed numerically. We discuss this further in the application section below, see also Mammen, Linton, and Nielsen (1999) 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

$$\sum_{i=1}^{n}\sum_{r=1}^{m_i}\{p_i - b_i(\tau_{ir})\theta\}^2 K_h(s - \tau_{ir}). \tag{14}$$

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 (14) 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 (14).

Our iterative method (13) is called successive approximation. For a detailed discussion we refer the reader to Kantorovich and Akilov (1964) and Luenberger (1969). 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 $e_k = \widehat{\mathcal{H}}_{LC}e_{k-1}/\|\widehat{\mathcal{H}}_{LC}e_{k-1}\|_2$, and the initial condition e_0 is chosen appropriately. Compare with the calculation of the absolutely maximum eigenvalue of a finite order matrix (in this case, e_0 should not be orthogonal to the eigenvector corresponding to the maximal eigenvalue). 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

$$\|g\|^2 = \sum_{i=1}^{n} \int \left\{ \sum_{j=1}^{m_i} b_i(\tau_{ij}) g_i(s_{ij}) \right\}^2 \prod_{j=1}^{m_i} \{K_h(s_{ij} - \tau_{ij})\ ds_{ij}\}\,. \tag{15}$$

If one puts $\eta_i(\cdot)$ as equal to $p_i[\sum_{j=1}^{m_i} b_i(\tau_{ij})]^{-1}$ [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 $\|\widehat{\mathcal{H}}_{LC}\| < 1$.[8]

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 $p_i \mapsto p_i / f(\tau_{im_i})$ and $b_i(\tau_{ij}) \mapsto b_i(\tau_{ij})/f(\tau_{im_i})$. Therefore, estimation proceeds as above with the transformed data.

[8] 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 (15). This means that

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

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.

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

$$\sum_{i=1}^{n} \int \left\{ p_i - \sum_{j=1}^{m_i} b_i(\tau_{ij})[d(s_{ij}) + (\tau_{ij} - s_{ij})d_1(s_{ij})] \right\}^2 \prod_{j=1}^{m_i} \{K_h(s_{ij} - \tau_{ij})\, ds_{ij}\} \tag{16}$$

with respect to (d, d_1). The local linear smooth is now defined as $\widehat{d}_{LL}$. Again, the minimizer of (16) can be easily calculated by an iterative algorithm. To solve the minimization in (16) we proceed as in the last section and put $d = \widehat{d}_{LL} + \epsilon\delta_s$ and $d_1 = \widehat{d}_{LL,1} + \epsilon\delta_s$ and differentiate with respect to ϵ and set $\epsilon = 0$. We end up with the following system of equations:

$$\begin{pmatrix} \widehat{d}_{LL}(s) \\ h\widehat{d}_{LL,1}(s) \end{pmatrix} = \begin{pmatrix} \overline{d}(s) \\ \overline{d}_1(s) \end{pmatrix} + \int \widehat{H}_{LL}(s,t) \begin{pmatrix} \widehat{d}_{LL}(t) \\ h\widehat{d}_{LL,1}(t) \end{pmatrix} dt, \tag{17}$$

where

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

and

$$M(s) = \frac{1}{n}\sum_{i=1}^{n}\sum_{r=1}^{m_i} \begin{pmatrix} b_i(\tau_{ir})^2 K_h(s-\tau_{ir}) & b_i(\tau_{ir})^2 \frac{\tau_{ir}-s}{h} K_h(s-\tau_{ir}) \\ b_i(\tau_{ir})^2 \frac{\tau_{ir}-s}{h} K_h(s-\tau_{ir}) & b_i(\tau_{ir})^2 \frac{(\tau_{ir}-s)^2}{h^2} K_h(s-\tau_{ir}) \end{pmatrix} \tag{18}$$

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

$$\begin{pmatrix} \widehat{d}_{LL}^{[a+1]}(s) \\ h\widehat{d}_{LL,1}^{[a+1]}(s) \end{pmatrix} = \begin{pmatrix} \overline{d}(s) \\ \overline{d}_1(s) \end{pmatrix} + \int \widehat{H}_{LL}(s,t) \begin{pmatrix} \widehat{d}_{LL}^{[a]}(t) \\ h\widehat{d}_{LL,1}^{[a]}(t) \end{pmatrix} dt. \tag{19}$$

As starting value $(\widehat{d}_{LL}^{[0]}, h\widehat{d}_{LL,1}^{[0]})$ one can choose $(\overline{d}, \overline{d}_1)$ for example. In this case the operator $\widehat{H}_{LL}$ is a 2 by 2 matrix, but otherwise is very similar to $\widehat{H}_{LC}$.

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 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 $d(0) = 1$. 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

$$\sum_{i=1}^{n} \int \left\{ p_i - \sum_{j=1}^{m_i} b_i(\tau_{ij}) \exp\{-\tau_{ij} y(s_{ij})\} \right\}^2 \prod_{j=1}^{m_i} \{K_h(s_{ij} - \tau_{ij})\ ds_{ij}\}, \tag{20}$$

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 $\epsilon = 0$) a single first order condition that defines the estimator. Unfortunately, the resulting first order condition is a nonlinear integral equation. Although there are a number of solution methods applicable here, our knowledge about the convergence of such methods is incomplete.

We suggest the following approach based on the local nonlinear least squares criterion function in θ

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

where

$$p_{ik} = p_i - \sum_{\substack{j=1 \\ j \neq k}}^{m_i} b_i(\tau_{ij}) \exp\{-\tau_{ij} y(\tau_{ij})\}.$$

Now let $\widehat{y}_{LCE}^{[a]}$ be given and define $\widehat{y}_{LCE}^{[a+1]}(t)$ as the minimizer of (21) with p_{ik} replaced by

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

Taking some initial value $\widehat{y}_{LCE}^{[0]}$, we proceed $a = 1, 2, \ldots$. The nonlinear optimization problem (21) can itself be approximately solved using Newton's method or Fisher scoring. We expect that reasonable implementations of this method will converge, but have no proof of this.

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 $\widehat{y}_{LCE}^{[1]}(t)$, which use $\widehat{y}_j(t)$ as initial condition and terminate at $a = 1$. In this case, of course, there is no issue about the convergence of the algorithm. Furthermore, the statistical properties of this estimator can be derived, as in Linton (1997,1998). Finite step methods have been studied in a general context by Robinson (1988), but the basic idea goes back to Fisher (1925).

A local linear exponential version of this algorithm can be defined similarly, and the same comments regarding asymptotic properties and one-step estimators apply.[9]

[9]For the local linear exponential, we minimize

3.4 Estimating Related Quantities

3.4.1 Estimating The Yield and Forward curve from the Discount Function

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

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

3.4.2 Estimating the discount function and forward curve from the Yield 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, $\widehat{y}'_{LCE}(t) = \int K_h'(t-s)\widehat{y}_{LCE}(s)ds$.

$$\sum_{i=1}^{n}\sum_{k=1}^{m_i}\left\{p_{ik}^{[a]} - b_i(\tau_{ik})\exp(-\tau_{ik}[\theta_0 + (\tau_{ik}-s)\theta_1])\right\}^2 K_h(s-\tau_{ik}),$$

where

$$p_{ik}^{[a]} = p_i - \sum_{\substack{j=1\\ j\neq k}}^{m_i} b_i(\tau_{ij})\int \exp\{-\tau_{ij}[\widehat{y}_{0,LLE}^{[a]}(t) + \widehat{y}_{1,LLE}^{[a]}(t)(\tau_{ij}-t)]\}K_h(t-\tau_{ij})dt.$$

Further details are available from the authors upon request.

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. As we show, the asymptotic variances of these procedures described above are identical [when estimating the same quantity], although the bias is certainly influenced by the chosen functional form.[10]

4.1 Local Constant Smoothing

Our asymptotics will be stated for the case of nonrandom τ_{ij} and m_i. This includes some random designs as special cases, when one works in the conditional distribution. We also give an example of a random design that satisfies our conditions. We concentrate on the case of a nonrandom design because our model (4) arises in other applications that would require another stochastic model for the design. Specifically, the influential paper of Engle et al. (1986)) falls in our framework. Note also, that our framework permits variable bandwidth methods with minimal changes in the assumptions. Therefore, we prefer to state the theorem under conditions that may work also in these applications. Our conditions are 'high level' − they are defined mostly in terms of the operator $\widehat{\mathcal{H}}_{LC}$ rather than the primitives of the data.

4.1.1 Assumptions for Theorem 1

(A1) Model (4) holds. Furthermore, we suppose that the error variables $\varepsilon_1, \ldots, \varepsilon_n$ are of the form $\varepsilon_i = \sum_{j=1}^{n} Q_{ij} u_j$, where $u_1, \ldots, u_n$ are independent mean zero variables with $E|u_j|^{2+\delta} \leq C^{2+\delta}$ for n large enough with constants $C, \delta > 0$. The values τ_{ij}, Q_{ij} and m_i are deterministic and they, the distributions of u_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 fulfils $n^{-1+\gamma_1} \leq h \leq n^{-\gamma_2}$ 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 $p = 2$.

[10] For comparison, Gozalo and Linton (1998) showed 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.

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

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

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

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

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

where

$$R_j(t) = \sum_{i=1}^{n} Q_{ij} \frac{\sum_{l=1}^{m_i} b_i(\tau_{il}) K_h(t-\tau_{il})}{\frac{1}{n}\sum_{i=1}^{n}\sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(t-\tau_{ir})}.$$

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 variables u_i have a finite Laplace transform

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

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 $s \in S$

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

where

$$R(t) = \frac{1}{n}\sum_{i=1}^{n} R_i(t).$$

4.1.2 Discussion

Assumptions (23)-(26) implicitly impose some restrictions on the Q_{ij}, i.e., the cross-sectional correlation, however, our assumptions are rather general and do not impose any specific structure. We now discuss a particular model, due to Conley (1999), which we think might be of interest here. He supposes that there is a variable z that drives the correlation between the $\varepsilon's$. Let $D_n = \{D_n(i,j)\}$ be the $n(n-1)/2$ vector of interpoint distances, where $D_n(i,j) = \|z_i - z_j\|$, and let $\Sigma(D_n)$ be an $n \times n$ covariance matrix such that $E\varepsilon\varepsilon' = \Sigma(D_n)$, where $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$. The variables z_i are observable and ancillary. The example $z_i = z(b_i, \tau_i)$ for some function $z(\cdot)$ and vectors $\tau_i = (\tau_{i1}, \ldots, \tau_{im_i})'$ and $b_i = (b_i(\tau_{i1}), \ldots, b_i(\tau_{im_i}))'$ is quite convenient; according to this, two bonds are similar if their nominal cash flows are similar. Suppose now that the matrix $Q_n = (Q_{ij})_{i,j=1}^n = Q(D_n)$ satisfies

$$\Sigma(D_n) = Q(D_n)Q(D_n)' = \begin{bmatrix} c_1^2 + C(0) & C(D_n(1,2)) & \cdots & C(D_n(1,n)) \\ & c_2^2 + C(0) & & C(D_n(2,n)) \\ & & \ddots & \vdots \\ & & & c_n^2 + C(0) \end{bmatrix}.$$

Provided the function C is continuous at zero and decreases to zero rapidly as the distance increases, we can expect the central limit theorem to operate for standardized averages of $\varepsilon_i's$.

Our results can be used to treat random τ_{ij}, Q_{ij} and m_i. In this case, one has to assume that our conditions hold conditionally given τ_{ij}, Q_{ij} and m_i with probability tending to one. We discuss now the special case that $Q_{ii} = 1$ and $Q_{ij} = 0$ for $i \neq j$ (in which case, $\varepsilon_1, \ldots, \varepsilon_n$ is an independent sample). In particular, if one assumes that the variables m_i are i.i.d. from some distribution with bounded integer 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.

Consider the following special case:

1. m_i is i.i.d. from some discrete distribution with frequency function f_m and support $\{1, \ldots, m_0\}$ for some fixed integer m_0.

2. The random variable τ_{i1} is continuously distributed on the interval $[0, 0.5]$ with density f_τ, which is strictly positive on its support. We then define

$$\tau_{ij} = \tau_{i1} + \frac{j-1}{2}, \ j = 1, \ldots, m_i.$$

3. The coupon rate c_i is randomly distributed on the interval $[\epsilon_1, \epsilon_2]$, where $0 < \epsilon_1 < \epsilon_2 < 1$, with density f_c; it is independent of m_i and τ_{i1}. The payment function satisfies

$$b_i(\tau_{ij}) = \begin{cases} c_i & \text{if } j < m_i \\ \\ 1 + c_i & \text{if } j = m_i. \end{cases}$$

Then, condition (A2) holds under a further condition on f_m, f_τ, and f_c.[11] However, this condition is rather complicated to state; instead we shall consider a special case, which gives better intuition. Suppose further that $m_i = m$ for all i and that $b_i(\tau_{ij}) = \beta_j$ for all i, j. Then, 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 be met in practice because the redemption value is usually considerably larger than the coupons. For an analytical example, suppose that $\beta_j = c < 1$ for $j = 1, \ldots, m-1$ and that $\beta_m = 1 + c$, then $\|\mathcal{H}_{LC}^\kappa\|_2 < 1$ for κ large enough, provided $c(m-1)/(1+c) < (\sqrt{5}-1)/2$. 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.[12]

We give now a rough discussion why the other conditions in A3 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

$$\int \widehat{H}_{LC}(s,u) R_i(u) \, du \approx \int H_{LC}(s,u) R_i(u) \, du \approx H_{LC}(s, \tau_{i1}) / f_\tau(\tau_{i1}).$$

[11] It is also easy to show that a sufficient condition for the function d in (5) to be unique on a given interval is that there is a positive fraction of bonds that mature in each half-yearly sub-interval.

[12] In cases where this condition might be violated we can always ensure compliance by weighting the criterion function to ensure that the 'average bond' satisfies this condition.

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.

4.1.3 Main Result

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

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

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

$$\sup_{s\in S}\left|\widehat{d}_{LC}(s) - d(s) - \beta_n^{LC}(s) - \sum_{i=1}^n \delta_{ni}(s)\varepsilon_i\right| = o_P((nh)^{-1/2}). \tag{31}$$

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 $p = \infty$.

For the special case of an independent sample $\varepsilon_1, \ldots, \varepsilon_n$ (i.e., $Q_{ii} = 1$ and $Q_{ij} = 0$ for $i \neq j$), we get the following formula for $s_n^{LC}(s)^2$

$$s_n^{LC}(s)^2 = \frac{h\frac{1}{n}\sum_{i=1}^n \left[\sum_{r=1}^{m_i} b_i(\tau_{ir})K_h(s-\tau_{ir})\right]^2 \ \mathrm{var}[\varepsilon_i]}{\left[\frac{1}{n}\sum_{i=1}^n \sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(s-\tau_{ir})\right]^2}.$$

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).[13] Under additional regularity conditions [like twice continuous differentiability of d] 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

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

[13] For example, in the independent error case, let

$$\widehat{v}_n(s)^2 = \sum_{i=1}^n \delta_{ni}^2(s)\widehat{\varepsilon}_i^2,$$

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).

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

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

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

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

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

where now

$$R_j(t) = \sum_{i=1}^{n} Q_{ij} M(s)^{-1} \sum_{l=1}^{m_i} b_i(\tau_{il}) K_h(t - \tau_{il}) \begin{pmatrix} 1 \\ (\tau_{il} - t)/h \end{pmatrix}.$$

Furthermore, it holds that

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

The matrix $S_n^{LL}(s)$ is defined in the statement of Theorem 2.

(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 variables have a finite Laplace transform

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

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 $s \in S$

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

where

$$R(t) = \frac{1}{n}\sum_{j=1}^{n} R_j(t).$$

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

$$
\begin{aligned}
S_n^{LL}(s) &= nh \text{ var} \left[M(s)^{-1} \frac{1}{n} \sum_{i=1}^{n} \sum_{r=1}^{m_i} b_i(\tau_{ir}) K_h(s - \tau_{ir}) \varepsilon_i \begin{pmatrix} 1 \\ (\tau_{ir} - s)/h \end{pmatrix} \right], \\
\beta_n^{LL}(s) &= \sum_{k=0}^{\infty} \widehat{\mathcal{H}}_{LL}^k \beta_n^{*,LL}(s), \\
\beta_n^{*,LL}(s) &= M(s)^{-1} \sum_{i=1}^{n} \sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(s - \tau_{ir}) [d(\tau_{ir}) - d(s) - (\tau_{ir} - s) d'(s)] \begin{pmatrix} 1 \\ (\tau_{ir} - s)/h \end{pmatrix} \\
&\quad + M(s)^{-1} \sum_{i=1}^{n} \sum_{r=1}^{m_i} \sum_{j=1, r \neq j}^{m_i} b_i(\tau_{ir}) b_i(\tau_{ij}) K_h(s - \tau_{ir}) \int K_h(t - \tau_{ij}) [d(\tau_{ij}) - d(t) \\
&\quad - (\tau_{ij} - t) d'(t)] \, dt \begin{pmatrix} 1 \\ (\tau_{ir} - s)/h \end{pmatrix}.
\end{aligned}
$$

For the special case of an independent sample $\varepsilon_1, \ldots, \varepsilon_n$ (i.e., $Q_{ii} = 1$ and $Q_{ij} = 0$ for $i \neq j$) we get under the additional assumption of B3 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

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

For the special case of an independent sample $\varepsilon_1, \ldots, \varepsilon_n$ (i.e., $Q_{ii} = 1$ and $Q_{ij} = 0$ for $i \neq j$) we get the following formula for $S_n^{LL}(s)$

$$
\begin{aligned}
S_n^{LL}(s) &= hM(s)^{-1} \frac{1}{n} \sum_{i=1}^{n} \text{ var} [\varepsilon_i] \begin{pmatrix} v_{i,0}^2(s) & v_{i,0}(s) v_{i,1}(s) \\ v_{i,0}(s) v_{i,1}(s) & v_{i,1}^2(s) \end{pmatrix} M(s)^{-1}, \\
v_{i,j}(s) &= \sum_{r=1}^{m_i} b_i(\tau_{ir}) K_h(s - \tau_{ir}) \left[\frac{\tau_{ir} - s}{h} \right]^j \qquad \text{for } i = 1, \ldots, n \text{ and } j = 0, 1.
\end{aligned}
$$

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

4.3 Local Exponential Smoothing

It is not possible at this stage to give a treatment of the convergence of the algorithm or the asymptotic properties of the local exponential methods. Our previous results made heavy use of the Hilbert space structure, which is not immediately available here. Instead, we shall examine the statistical properties of the one-step estimators, which are themselves linear approximations to the nonlinear method. The asymptotics of the one-step methods follows from the arguments given in Linton (1998).

We just outline the result for the one-step local constant exponential method $\widehat{y}^{[1]}_{LCE}(s)$. We first of all define an infeasible estimator whose properties will determine those of $\widehat{y}^{[1]}_{LCE}(s)$. Let $\widehat{y}^{*}_{LCE}(s)$ be the minimizer of the criterion (21) with respect to θ. 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

$$\widehat{y}^{*}_{LCE}(s) = \{y(s) + \beta^{*}_{n}(s) + \alpha^{*}_{n}(s)\}\{1 + o_p(1)\}, \tag{39}$$

where

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

Here, $\alpha^{*}_{n}(s)$ is a sample average of independent [or more generally, weakly dependent] random variables and will satisfy a central limit theorem, while $\beta^{*}_{n}(s)$ is a bias term which can be expected to be of order h^2 under the usual conditions.[14]

In a previous sub-section we established the uniform asymptotic expansion

$$\widehat{y}_{LC}(s) = \{y(s) + \beta^{LC}_{n}(s) + \alpha^{LC}_{n}(s)\}\{1 + o_p(1)\}, \tag{40}$$

where $\alpha^{LC}_{n}(s)$ is a sum of mean zero independent [or more generally weakly dependent] random variables, and hence satisfies a central limit theorem, while $\beta^{LC}_{n}(s)$ is a deterministic bias function. Combining the expansions for $\widehat{y}^{*}_{LCE}(s)$ and $\widehat{y}_{LC}(s)$, we obtain after some fairly standard manipulations that

$$\widehat{y}^{[1]}_{LCE}(s) = \{\widehat{y}^{*}_{LCE}(s) + \beta^{**}_{n}(s)\}\{1 + o_p(1)\} \tag{41}$$

[14] In fact,

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

uniformly in s, where $\beta_n^{**}(s)$ is a deterministic bias term.[15] Note that the stochastic terms from $\widehat{y}_{LC}(s)$ do not appear in the leading terms of $\widehat{y}_{LCE}^{[1]}(s)$, i.e., the asymptotic variance of $\widehat{y}_{LCE}^{[1]}(s)$ is the same as that of $\widehat{y}_{LCE}^{*}(s)$; this follows by the same arguments used in Linton (1998). The bias of $\widehat{y}_{LCE}^{[1]}(s)$ is different from that of $\widehat{y}_{LCE}^{*}(s)$ and includes additional terms. In conclusion, $\widehat{y}_{LCE}^{[1]}(s)$ is asymptotically normal with variance $\mathrm{var}(\alpha_n^*(s))$.[16]

Clearly, the biases of our procedures will in general be complicated functions of the data distribution, and very hard to estimate. We can say that the local constant estimator is approximately unbiased when d is a constant function, while the local linear estimator is approximately unbiased when d is a linear function. Similarly, the local constant exponential will be approximately unbiased when y is a constant function, while the local linear exponential will be approximately unbiased when y is a linear function. This is perhaps one reason why the exponential versions seem to do better - the yield curve is much closer to a constant or linear function than is the discount function.

5 Simulation study

In this section we investigate the properties of the kernel yield curve estimator in a numerical example. We study the performance of the local constant exponential (LCE) and local linear exponential (LLE) estimators, and compare them with several well-known competitors.

We generated the data using the Nelson-Siegel yield curve model [Nelson and Siegel (1987)]. This model has been in several empirical studies of term structure estimation methods [Bliss (1996) and Dahlquist and Svensson (1996)]. We chose the extended version

$$y(t) = \beta_0 + \beta_1 \left[\frac{1 - \exp(-t/\tau_1)}{t/\tau_1} \right] + \beta_2 \left[\frac{1 - \exp(-t/\tau_2)}{t/\tau_2} - \exp(-t/\tau_2) \right].$$

[15] The bias term is

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

[16] It is expected that further iterations have the same asymptotic variance, but that the bias formula involves additional terms.

Table 1 summarizes the three special cases we use here. In the first version, the curve is concave and monotonically increasing, whereas the last two cases have a stationary point.

We measure the performance of estimators by the root integrated mean squared error

$$IRMSE = \int_a^b \sqrt{\mathrm{E}\left[\widehat{\xi}(t) - \xi(t)\right]^2} dt, \tag{42}$$

where ξ is either the yield curve, the forward curve, or the discount function, while $\widehat{\xi}$ is an estimator. Here, a, b are constants. We also report results using the integrated mean absolute error

$$IMAE = \int_a^b \mathrm{E}\left|\widehat{\xi}(t) - \xi(t)\right| dt. \tag{43}$$

These expected values are found by running a large number of Monte Carlo simulations of the true yield curve and then averaging across the integrated squared errors [absolute errors] from the corresponding estimator.

We use a design that resembles closely the structure of the US Treasury market. US Treasury notes are issued with original maturities of 2,3,5, and 10 years. We exclude the long term bonds with original time to maturity of 30 years. We [rather arbitrarily] select the coupon so as to make the note trade at par or close to. T-Bills [zero coupons] are issued with original time to maturity 1 to 12 months. We generate prices for 12 Bills with time to maturity less than 1 year and 24 issues with time to maturity larger than 1 year.

Prices are generated by discounting the future cashflow and adding a random error, ε_i, generated as $\varepsilon_i \sim \mathrm{N}[0, \sigma_i^2]$, where the noise term is uncorrelated between bonds and the variance may vary from bond to bond.

We consider two cases. The first case is when the variance is constant across maturities. In the second case we follow the convention of letting the standard deviation be proportional with the bonds duration, i.e., $\sigma_i^2 = \sigma \times duration_i$ [see Bliss (1996) for example].[17]

In general, a constant bandwidth does not perform well, because the payment time distribution is far from uniform. Accordingly, we use a non-constant bandwidth. For each s define $s_i = |s - m_i|, .i =$

[17]The duration is defined as a weighted average of the future payment dates. The formula is $duration_i = \sum_{i=1}^{m_i} \tau_{ij}\varpi_{ij}$, where $\varpi_{ij} = (pv_i/pv)$ is the relative contribution of the payment at time time j to the total discounted value of the bond. For simulation purposes we used the discounted values under the true yield curve. In practical applications one can use the Macauley duration where discounting is done using the redemption yield.

$1, \ldots, n$. Then we choose as the bandwidth a quantile of the distribution of s.[18] For maturity intervals such as the interval $[0, 1]$,where relatively more bonds are redeemed, the bandwidth will become smaller than in intervals with fewer bonds such as in the long end of the yield curve.[19] For the example of this paper we use the 33.33% quantile. The bandwidths determined by this method range from 0.16 years for short maturities to 6.25 years for the longest maturities. This way of determining bandwidth could be supplemented by an empirical measure of goodness of fit such as generalized cross-validation.

Besides the kernel smoothers we computed three other estimators. The first one is the extended Nelson Siegel procedure. Second, we estimate a cubic regression spline for the yield curve with knots $1/12, 1.5, 3, 5$, and 10 years. Finally, we estimate the cubic smoothing spline of Tanggaard (1997).[20] The first alternative serves as a kind of benchmark. Depending on the heteroscedasticity structure one can not hope to get a better fit than this, when in fact, the true curve is the ENS, and it is interesting to see how close a nonparametric estimator can come to this.

Tables 3 and 4 clearly demonstrates that fitting the Nelson-Siegel model, when the true curve is the NS, is better than anything else. This is the case for all error measures.[21] In the more realistic cases when we do not assume a known functional form, we see that the local linear exponential performs better than the local constant exponential version. That is true irrespective of the performance measure. Furthermore the LLE model perform better than any of the other curve estimators, the only exception being the regression spline which performs better for the first version of the NS-curve. However, the regression spline performs remarkedly worse for versions 2 and 3 of the NS-curve. This is consequence of one of the major problems with regression splines (splines with fixed knots): the

[18]This is one of several examples of bandwidth formulas we tried. We also tried $h(s) = \beta_0 \mathrm{f}(s)^{-\beta_1}$, where f is an estimate of the density of the payment times, while the parameters β_0, β_1 are chosen to minimize the cross-validation criterions. We did some experiments but for several reasons, but found that it can not be used as a general purpose bandwidth estimator. A major reason for this is the computational problems.

[19]Graphs of the estimated curves, clearly indicated that this method produced too high bandwidths for maturities < 1 year. Therefore, we arbitrarily set the bandwidth to 1/6 year for these short maturities. This is explained by the fact that bonds in this interval are zero coupon bonds, which should be smoothed less than the longer maturity bonds. This changes only affects the estimates in the short end and had no substantial effect on the conclusions of this study.

[20]The spline smoother of Tanggaard is essentially the same model as the one suggested by Fisher at al. (1995). It depends on a smoothing constant which is determed empirically by a generalised cross-validation type of smoother.

[21]The errors in estimation of the discount are so small that it is difficult to assess the difference between the estimators.

optimal placement of knots depends heavily on the curvature properties of the curve one is interested in approximating. There are to our knowledge no good ways of estimating the knot placement in this kind of model.

Given the superior curvature properties of cubic splines it is a bit surprising how relatively poor the spline smoother (SS) performs. In a study - not reported here - of the performance under best possible circumstances, i.e., when the degree of smoothing is best possible, we found that the spline smoother performed better than any of the other curves. One explanation of these results is therefore that this is related to the lack of a good bandwidth estimator.

Finally, the results also demonstrate that the error in estimating the forward curve is an order of magnitude larger than for the yield curve. The errors in estimating the discount curve are measured on another scale and can not be compared with the yield curve errors. However, panel III of tables 3 and 4 show that the discount curve alone can hardly be used to discriminate among competing estimators.

Overall this example shows that kernel methods is a powerful method of estimating zero-coupon term structures. In most examples the estimation error is smaller than for competing methods. The problems with this and other nonparametric methods lies in good methods for empirical estimation of the degree of smoothing.

6 Concluding Comment

One nice feature of our projection approach to estimation is that has a sensible intepretation under misspecification; we are finding the function d that is closest to the price vector in the chosen metric.

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

$$\sup_{s\in S} |\int \widehat{H}_{LC}(s,t)g(t)\ dt| \leq \sup_{s\in S} |\int \widehat{H}_{LC}(s,t)^2\ dt|^{1/2}\|g\|_2 \leq C\|g\|_2. \tag{44}$$

Note now that

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

and

$$\widehat{d}_{LC}(s) - E\widehat{d}_{LC}(s) = \arg\min_{v} \sum_{i=1}^{n} \int \left\{ \varepsilon_i - \sum_{j=1}^{m_i} b_i(\tau_{ij}) v(s_{ij}) \right\}^2 \prod_{j=1}^{m_i} \{K_h(s_{ij} - \tau_{ij})\ ds_{ij}\}. \tag{46}$$

For the proof of (31) we show first that

$$E\widehat{d}_{LC}(s) - d(s) = \beta_n^{LC}(s). \tag{47}$$

For the proof of (47) note first that equation (45) implies that

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

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 (47) 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 (46) 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

$$\sup_{s\in S} |\sum_{k=1}^{\infty} \widehat{\mathcal{H}}^k_{LC} \widetilde{d}^*_{LC}(s)| = o_P((nh)^{-1/2}) \tag{48}$$

holds. This shows

$$\sup_{s\in S} |\widetilde{d}_{LC}(s) - \widetilde{d}^*_{LC}(s)| = o_P((nh)^{-1/2}). \tag{49}$$

Therefore, (47) and (48) imply (31).

For the proof of claim (48) we will show

$$\sup_{s\in S} |\widehat{\mathcal{H}}^k_{LC} \widetilde{d}^*_{LC}(s)| \le \rho^k R_n, \tag{50}$$

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 (48) because of

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

For the proof of (50) we will show that

$$\|\widehat{\mathcal{H}}_{LC} \widetilde{d}^*_{LC}\|_2 = o_P((nh)^{-1/2}), \tag{51}$$

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

We argue now that (51) and (52) imply (50). 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 (44) 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 $k \ge 2$ with $R_n = \|\widehat{\mathcal{H}}_{LC} \widetilde{d}^*_{LC}\|_2/\rho^2$. With (51) and (52) we get (50). So for the proof of (31) it remains to show (51) and (52). For the proof of (51) we get using (A1) and (23)

$$E\|\widehat{\mathcal{H}}_{LC} \widetilde{d}^*_{LC}\|_2^2 = E\, \frac{1}{n^2} \sum_{i,j=1}^n \int \widehat{H}_{LC}(s,t) \widehat{H}_{LC}(s,v) R_i(t) R_j(v)\, dv\, dt\, ds\, u_i u_j$$

$$
\begin{aligned}
&= \frac{1}{n^2}\sum_{i=1}^{n}\int \widehat{H}_{LC}(s,t)\widehat{H}_{LC}(s,v)R_i(t)R_i(v)\ dv\ dt\ ds\ \ \mathrm{var}[u_i] \\
&\leq \frac{1}{n^2}\sum_{i=1}^{n}\int \left|\int \widehat{H}_{LC}(t,v)R_i(v)\ dv\right|^2\ dt\ C \\
&= o((nh)^{-1}).
\end{aligned}
$$

This shows (51). For the proof of (52) we write

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

where

$$
r_i(s) = \int \widehat{H}_{LC}(s,v)R_i(v)\ dv.
$$

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

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

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

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

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

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

This follows from the uniform bound for the Laplace transform of u_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

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

Because u_i has a bounded Laplace transform [see (27)] it holds that $\sup_{1\leq i\leq n}|u_i| = O_P(\log(n))$. 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

$$\sup_{s\in S}\inf_{t\in S_n} |\widehat{\mathcal{H}}_{LC}\widetilde{d}^*_{LC}(s) - \widehat{\mathcal{H}}_{LC}\widetilde{d}^*_{LC}(t)| = o_P((nh)^{-1/2}). \tag{54}$$

Claim (52) follows from (53) and (54). This shows (31). Note that the above argument can be extended to allow for only a finite number of moments on u_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 (47) we have to show

$$\sqrt{nh}\widetilde{d}_{LC}(s)s_n^{LC}(s)^{-1} \Longrightarrow N(0,1) \tag{55}$$

in distribution. We argue now that

$$\widetilde{d}_{LC}(s) - \widetilde{d}^*_{LC}(s) = o_P((nh)^{-1/2}). \tag{56}$$

This follows as in the proof of (49) where now $S = \{s\}$. Note that now, because S is finite, the proof of (52) follows immediately from (23) and the boundedness of σ_i^2, see (A1). Because of (56) and (26) for the proof of (55) it suffices to show

$$\sqrt{nh}\widetilde{d}^*_{LC}(s)s_n^{LC}(s)^{-1} \Longrightarrow N(0,1) \tag{57}$$

in distribution. Note that

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

Now, claim (57) follows with a standard version of the central limit theorem from (25) and the existence of a uniform bound for $Eu_i^{2+\delta}$, 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: $n^{-1}\sum_{i=1}^{n}\sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(s-\tau_{ir})$ and $n^{-1}\sum_{i=1}^{n}\sum_{r=1}^{m_i} b_i(\tau_{ir})^2 K_h(s-\tau_{ir})[(\tau_{ir}-s)/h]^2$. Using expansion (38) for $S = \{s\}$ one gets that $\widehat{d}_{LL}(s)$ has the same asymptotic variance as $\widehat{d}_{LC}(s)$. ■

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.

Conley, T.G. (1999). GMM estimation with cross-sectional dependence. Journal of Econometrics 92, 1-46.

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.

Fisher, R.A. (1925). Theory of statistical estimation. Proceedings of the Cambridge Philosophical Society 22, 700-725.

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 estimation: Using parametric information nonparametrically. Cowles Foundation Discussion Paper no 1075. Forthcoming in the Journal of Econometrics.

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.

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

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

Kantorovich, L.V. and G.P. Akilov (1964). Functional analysis in normed spaces. Pergamom press, Oxford.

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. (1997). Efficient estimation of additive nonparametric regression models. Biometrika 84, 469-474.

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

Luenberger, D.G. (1969). Optimisation by vector space methods. Wiley, New York.

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

McCulloch, 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.

Mammen, E., Linton, O. B., and Nielsen, J. P. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. The Annals of Statistics.

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., Cambridge University Press

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

Robinson, P.M. (1988). The stochastic difference between econometrics statistics. Econometrica 56, 531-548.

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.

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.

Tables

	Special cases of the Nelson-Siegel model		
	y_1	y_2	y_3
β_1	0.065	0.04	0.0
β_2	-0.015	0.01	0.05
β_3	0.05	0.1	2
τ_1	0.5	6	0.75
τ_2	6	0.5	125

Table 1: Parameters of the extended Nelson-Siegel model.

Estimator	Description
NS	Extended Nelson-Siegel
LCE	Local constant exponential with variable bandwidth
LLE	Local linear exponential with variable bandwidth
RS	Regression spline with 5 fixed knots
SS	Smoothing spline. Smoothing constant determined by GCV

Notes: The knots for the regression spline are placed at maturities 1/12,1.5,3,5,10 years.

Table 2: Summary of estimators

Error measure	MSE			MAE		
I: Yield curves	$y_{NS,1}$	$y_{NS,2}$	$y_{NS,3}$	$y_{NS,1}$	$y_{NS,2}$	$y_{NS,3}$
$\widehat{y}_{NS}$	0.01	0.00	0.00	0.00	0.00	0.00
$\widehat{y}_{LCE}$	0.07	0.14	0.34	0.05	0.10	0.25
$\widehat{y}_{LLE}$	0.05	0.07	0.05	0.04	0.06	0.04
$\widehat{y}_{RS}$	0.04	0.31	0.07	0.01	0.09	0.02
$\widehat{y}_{SS}$	0.32	0.35	0.64	0.19	0.24	0.34
II: Forward curves	$f_{NS,1}$	$f_{NS,2}$	$f_{NS,3}$	$f_{NS,1}$	$f_{NS,2}$	$f_{NS,3}$
$\widehat{f}_{NS}$	0.01	0.01	0.01	0.00	0.01	0.00
$\widehat{f}_{LCE}$	0.46	0.95	3.62	0.37	0.79	2.60
$\widehat{f}_{LLE}$	0.32	0.67	0.47	0.20	0.34	0.21
$\widehat{f}_{RS}$	0.05	0.29	0.09	0.04	0.17	0.06
$\widehat{f}_{SS}$	0.45	0.67	0.72	0.37	0.57	0.59
III: Discount curves	$d_{NS,1}$	$d_{NS,2}$	$d_{NS,3}$	$d_{NS,1}$	$d_{NS,2}$	$d_{NS,3}$
$\widehat{d}_{NS}$	0.00	0.00	0.00	0.01	0.01	0.01
$\widehat{d}_{LCE}$	0.00	0.00	0.02	0.15	0.32	1.05
$\widehat{d}_{LLE}$	0.00	0.00	0.00	0.11	0.20	0.11
$\widehat{d}_{RS}$	0.00	0.00	0.00	0.02	0.07	0.03
$\widehat{d}_{SS}$	0.00	0.01	0.04	0.33	0.54	0.59

Notes: The table reports the estimation errors when estimating a yield curve based on simulated data from one of 3 versions of the extended Nelson-Siegel model. Estimation errors are summarised by the root mean squared error (square root of MSE), as well as the mean absolute error. For the yield curves and forward curves the scale is percentage points. The noise term is homoscedastic with standard deviation equal to 0.03. The number of bonds is $n = 36$. The numerical precision is better than 10^{-2}.

Table 3: Estimation errors with homoscedastic noise term.

Error measure	MSE			MAE		
I: Yield curves	$y_{NS,1}$	$y_{NS,2}$	$y_{NS,3}$	$y_{NS,1}$	$y_{NS,2}$	$y_{NS,3}$
$\widehat{y}_{NS}$	0.00	0.00	0.00	0.00	0.00	0.00
$\widehat{y}_{LCE}$	0.10	0.19	0.33	0.08	0.15	0.23
$\widehat{y}_{LLE}$	0.04	0.06	0.05	0.03	0.05	0.04
$\widehat{y}_{RS}$	0.04	0.29	0.07	0.03	0.25	0.06
$\widehat{y}_{SS}$	0.35	0.40	0.59	0.30	0.32	0.52
II: Forward curves	$f_{NS,1}$	$f_{NS,2}$	$f_{NS,3}$	$f_{NS,1}$	$f_{NS,2}$	$f_{NS,3}$
$\widehat{f}_{NS}$	0.00	0.00	0.01	0.00	0.00	0.01
$\widehat{f}_{LCE}$	0.48	0.90	3.35	0.37	0.74	2.34
$\widehat{f}_{LLE}$	0.29	0.61	0.72	0.20	0.38	0.35
$\widehat{f}_{RS}$	0.19	1.42	0.37	0.16	1.20	0.31
$\widehat{f}_{SS}$	1.36	1.08	1.94	1.04	0.92	1.58
III: Discount curves	$d_{NS,1}$	$d_{NS,2}$	$d_{NS,3}$	$d_{NS,1}$	$d_{NS,2}$	$d_{NS,3}$
$\widehat{d}_{NS}$	0.00	0.00	0.00	0.01	0.01	0.01
$\widehat{d}_{LCE}$	0.00	0.01	0.02	0.25	0.52	0.99
$\widehat{d}_{LLE}$	0.00	0.00	0.00	0.11	0.20	0.14
$\widehat{d}_{RS}$	0.00	0.01	0.00	0.10	0.84	0.20
$\widehat{d}_{SS}$	0.01	0.01	0.04	0.90	0.89	1.62

Notes: The table reports the estimation errors when estimating a yield curve based on simulated data from one of 3 versions of the extended Nelson-Siegel model. Estimation errors are summarised by the root mean squared error (square root of MSE), as well as the mean absolute error. For the yield curves and forward curves the scale is percentage points. The noise term is heteroscedastic with standard deviation equal to $0.01 \times duration$. The number of bonds is $n = 36$. The numerical precision is better than 10^{-2}.

Table 4: Estimation errors with heteroscedastic noise term.