

中国利率期限结构的局部线性迭代模型研究*

文兴易 黎实

(西南财经大学统计学院, 成都, 611130)

摘要:利率期限结构问题是金融学的一个基本问题。用国外广为应用的样条估计方法估计国内利率期限结构时,会受到国内利率数据中异常点的影响,导致拟合扭曲。本文从局部线性回归非参数统计方法出发,提出适合我国利率市场特点的局部线性迭代模型。选取2010年6月17日上海证券交易所国债数据,使用局部线性迭代模型对我国利率期限结构进行了实证分析,并与样条方法相比较。结果表明局部线性迭代建模方法能够合理地估计我国的利率期限结构,有效地降低异常数据地影响,拟合效果比样条估计方法更准确,更好。从而为我国利率期限结构的研究提供了一种全新的方法理论。

关键字:局部线性迭代模型;非参数估计;利率期限结构

中图分类号:F833

文献表示码: A

The Term Structure of Interest Rates in China with Local linear Iteration Modeling

Abstract: The term structure of interest rate problem is one of the basic questions in finance. On the basis of Local Polynomial Theory, the article raises the iteration local linear model which is suitable for characteristics of local Chinese interest rates market. On the basis of the statistics which are chosen from Shanghai Stock Exchange Treasury data, the paper uses iteration local linear model to do empirical analysis of term structure of interest rate in China and makes a comparison with Spline Function Method. The results show that iteration local linear model method can estimate term structure of interest rate in China reasonably, and reduce the influences of abnormal data effectively. What's more, fitting effect is much better and more precise than the spline estimation method. Then, it provides a completely new method for estimating term structure of interest rate in our country.

Key Words: Local Polynomial, Local linear iteration, Spline Estimation, Term Structure of Interest Rate

*[投稿日期]: 2010-12-30

[基金项目]: 国家自然科学基金2010年度项目(71071130); 教育部人文社会科学重点基地2006年度重大项目(06JJD790026); 西南财经大学“211工程”三期统计学重点学科建设项目。

[作者简介]: 文兴易(1983-)男,西南财经大学统计学院数量经济博士研究生,电子邮箱:bonowen@foxmail.com。黎实(1955-)男,教授,经济学博士,西南财经大学统计学院。

引言

利率期限结构 (Term Structure of Interest Rates) 是指具有相同信用风险、不同期限的金融工具在某一时点上的收益率 (Yield) 与到期期限 (Maturity) 之间的关系及变化规律。当金融工具选用资金时, 利率的期限结构则专指某个时点上不同期限的即期利率与到期期限间的关系, 这种关系可图示为曲线描述, 以反映不同期限的资金供求关系, 揭示市场利率的总体水平和变化方向。因此, 利率期限结构不仅是市场重要的定价准则, 而且还是科学制定财政和货币政策, 完善国债发行和管理的重要依据。随着金融的不断创新, 利率期限结构在揭示通货膨胀和宏观经济未来走向等方面也具有明显的优势。因此研究利率期限结构对发展和完善我国资本市场具有重要的理论意义和现实意义。

20 世纪 70 年代以来, 国外学者对利率期限结构进行了大量的研究, 发展出一大批估计利率期限结构的模型。这些模型中, 主流模型可分为三类: 一是以 Cox、Ingersoll、Ross (CIR) (1985) 和无套利模型 (1992) 为代表的经济理论模型; 二是以 NS 和 NSS 模型为代表的参数模型, 三是多项式样条和 B 样条为代表的非参数模型等。近年来, 国内学者跟随上述三类模型方法对中国的利率期限结构进行实证分析和比较。研究表明, 大多数研究成果认为, 尚处于不断完善完善的国内利率市场约束限制了前两类模型 (经济理论模型和参数模型) 的应用。因此国内文献对利率期限结构的研究大都集中在套用国外样条估计的非参数模型上。

然而中国的国债市场不同于西方发达的债券市场, 样条估计的非参数模型在实际应用中也存在这样或那样的制约。就估计而言, 市场机制设计、税收、流动性溢价和虚假交易等诸多因素往往会引起某些债券的定价错误, 导致债券数据出现异常点, 这势必对样条函数的估计造成影响。再加之我国债券品种尚不够丰富, 可供观测的样本量受限, 债券期限分布不均匀, 并且大都为付息债券等因素, 将必然会加重异常点的影响效应, 造成拟合曲线更加的扭曲, 甚至会估计出负的利率水平。因此, 直接套用国外的样条模型的做法不一定是合适的。当然, 国内也有不少学者也注意到了在中国直接应用样条模型的局限, 或是利用剔除异常样本、或是稳健估计的样条、或是平滑样条等来减小异常点的影响, 试图提高样条模型在中国的应用性。但是, 笔者以为, 目前国内学者关于样条模型的研究工作, 大都是对原有模型进行的缝补工作, 并没有改变样条函数模型的本质的本质和统计特性, 因而多数研究工作的效果并不十分理想。

本文试图结合中国利率市场特点, 跳出和摒弃样条函数的束缚, 采用局部线性迭代模型非参数估计方法对利率期限上的每个点建模, 以期达到如下三个期望目标, 为中国利率期限结构的研究提供一种全新的方法和理论: 一是在估计方法上, 拟合精度比原有方法效果更好; 二是在统计性质上, 具有优良的统计性质, 弥补原有方法纯粹从数据上拟合缺乏统计性质的缺点; 三是在实际应用上, 有效地降低中国债券市场上异常点对曲线拟合的扭曲, 使得拟合结果更加符合中国债券市场的实际情况。

本文余下部分的结构安排如下: 第二部分回顾文献并阐述研究的内容和出发点; 第三部分解释研究设计, 包括基本模型设定、模型改进、样本选择和数据处理方法; 第四部分利用局部线性迭代估计模型和样条估计模型分别进行实证分析和比较, 第五部分给出结论和进一步研究方向。

1 文献综述

目前,国外对利率期限结构的研究方法主要分为两大类:经济理论方法和数量方法。经济理论方法以 Cox、Ingersoll、Ross(1985)提出的 CIR 均衡模型和 Vasicek(1977)、Ho(1986)和 Heath(1992)等提出的无套利模型为代表,通过一些经济上的假设,对利率短期的随机行为进行刻画和建模。此类模型以有效市场假说为基础,使得市场能够随时保持瞬时的无套利性,一旦这个前提和市场不符,应用便受限。事实上,国内债券市场似乎难以满足有效市场假说前提,目前关于这类模型在国内债券市场应用的研究鲜见。

数量方法是无论市场情况如何,都回归到利率期限结构的本质上,利用市场上可以观察到的债券价格数据来拟合利率期限结构。目前主要的估计方法有参数方法和非参数方法。参数模型估计利率期限结构,最初是由 Nelson 和 Siegel(1987)提出的(简称 NS 法)。NS 法主要采用指数衰减型函数对远期利率进行估计。其中的参数分别表示利率的长期水平、短期水平、收益率曲线的斜率以及驼峰形状。尽管 NS 法在形状相对简单的收益率曲线中能够得到较好的估计结果,但对于复杂的利率期限结构,NS 法最大的不足是拟合优度较差。由此, Svensson 扩展了 Nelson 和 Siegel 方法的方法(简称 NSS 法),增加了加强估计双驼峰状收益率曲线的能力,提高了估计的效果。NS 模型和 NSS 模型均采用了确定形式的拟合函数,参数估计较为简单,但这两个模型对收益率曲线长短期两端部分的数据变化十分敏感,这两个部分如果出现异常值都很有可能对整个利率期限结构的估计产生很大的扭曲,稳定性较差。因此,NS 模型和 NSS 模型在发展还不够完善的国内债券市场中的应用,也少有文献报告。

与参数估计方法相比,非参数方法的不同之处主要是对拟合函数不是采用确定的形式,而是在一定程度的有所放宽,拟合函数的具体形式主要采用有剥离息票法和样条法。

剥离息票法最初是由 Fama 和 Bliss(1987)提出的,该方法从付息债券中剥离出息票,并利用递归运算的方法估计零息票债券的到期收益率水平。但剥离息票法假定两个相邻期限之间的利率服从线性关系,并要求估计样本中各息票债券的付息日相同或相近,这些假设与现实并不太符合。同时,实证研究也表明利用剥离息票法拟合出的利率期限结构其短期部分精确度较高,而中长期部分的拟合优度则较差。

样条法是由 McCulloch(1971)提出的。该方法将利率期限结构表示成贴现函数的形式,贴现函数是到期时间的连续函数,通过样条函数估计贴现函数,从而达到拟合整个利率期限结构的目的。其实质上就是利用正交函数族对利率期限结构逼近和拟合。McCulloch(1975)认为三次样条函数估计效果要好于二次样条函数,从而成为标准的利率期限结构估计方法。但是由三次样条函数估计的远期利率不是很稳定,会出现较大的波动,此后很多学者在此基础上对样条函数进行改进,Vasicek 和 Fong(1982)提出利用指数样条函数拟合贴现函数。Fisher 和 Zeros(1982)在 McCulloch(1975)的三次回归样条法基础上提出了用三次平滑样条法来拟合远期瞬间利率曲线。Kteley(1991)使用 B 样条函数来逼近贴现函数,等等。

在国内,姚长辉和梁跃军(1998)、杨大凯和王欢(1999)、陈雯和陈浪南(2000)以及杨春鹏和曹兴华(2002)对我国的利率期限结构做了早期的研究,但这些研究的对象均是国债的到期收益率,由于国内债券市场上存在的国债大都为付息债券,到期收益率曲线并不是真正意义上的利率期限结构。对此,郑振龙和林海(2003)采用剥离息票法和样条估计方法,

利用上交所的附息国债价格来估计真正意义上的中国静态利率期限结构。随后,国内学者利用国外文献中各种样条函数的改进模型,结合中国国债数据对国内利率市场进行了大量的研究:王晓芳和刘凤根(2004)、周荣喜和车君(2010)利用三次多项式样条函数估计研究;何启志和何建敏(2008)选用多项式样条函数和指数样条函数估计;刘灿和易璐(2004)、仝晓燕和程希骏(2006)、朱淑英和刘伶(2010)、吴泽福和吴捷(2010)分别利用 B 样条函数拟合国债利率期限结构;胡海鹏和方兆本(2009)用平滑样条方法拟合;付曼丽和董荣杰(2002)分别利用多项式样条、B 样条函数、NS 模型、NSS 模型对中国利率市场做实证研究等。

样条估计中对异常点的处理方法方面,仝晓燕和程希骏(2006)、程希骏和萧楠(2008)剔除异常点后再用样条函数方法估计;胡海鹏和方兆本(2009)用改进的平滑样条方法进行估计;李熠熠和缪柏其(2010)在三次多项式样条函数的基础上用最小一乘准则估计;吴泽福和吴捷(2010)用 L1 样条与改进的 B 样条方法估计等。

笔者以为,目前国内关于弥补样条函数的思路 and 做法,至少会造成以下方面的不足:(1)剔除异常点的方法,在国内债券品种和数量较少的条件下,会导致样本数量进一步减少,流失样本信息量,使得估计偏差变大,不准确;(2)其余对异常点的处理方法,在本质上看,可认为只是在原有的样条估计方法基础上所进行的修补,这样一来,利用正交函数族拟合利率曲线的样条方法的本质,并没得以改变。那么,在这种本质下,异常点依然会对整段函数的估计产生不良的统计影响,甚至造成拟合曲线的扭曲;(3)样条函数拟合的思路,从统计的角度看,在某种意义上应可视为是一种纯粹从数据上拟合利率曲线的思路,而这种思路 and 做法似乎缺乏统计分析方法的支撑,尚不可进行诸如假设检验等在内的若干统计分析,这也势必会限制该方法的应用。

为此,笔者试图跳出样条函数拟合利率期限结构曲线思路的束缚,采用局部线性迭代模型的思路,对利率期限曲线上的每个点进行建模,进而达到克服样条函数的思路 and 做法上述三个方面不足的期望目标,以期弥补样条函数拟合缺乏统计性质的缺陷,形成具有优良的大样本统计性质、能够对各种假设做统计检验、应用面更为广泛的利率期限结构的局部线性迭代模型,为中国利率期限结构的研究提供一种全新的方法和理论。

2 研究设计

2.1 局部线性迭代模型

1. 基本模型

局部线性回归最初是由 Stone(1977,1980,1982)、Cleveland(1979)提出的,随后经 Fan(1992,1993a), Fan 和 Gijbels(1992)以及 Ruppert 和 Wand(1994)等的发展和完善,现今在金融、统计、医学、物理等多个领域有着广泛的应用。

一般的建模过程是对观测数据 $(X_i, Y_i), i = 1, 2, \dots, n$ 寻找被解释变量 Y 对解释变量 X 的函数关系:

$$Y_i = m(X_i) + u \quad (1)$$

其中 u 是白噪声过程, $m(x) = E(Y | X = x)$ 是需要估计和推断的回归函数,也是建模过程的关键。

局部线性回归与一般的参数模型不同，并没有事先假设回归函数 $m(x)$ 满足某种特定的参数形式，通过估计参数达到估计回归函数的目的。也不同于样条类非参数模型，把函数分为几段，用几个样条函数对各段数据分别拟合。数据的异常点会对某段函数的估计造成很大的影响。局部线性回归是从函数映射的定义出发，对定义域内的每一个点 x_0 ，利用观测数据，找到回归函数对应的映射 $\hat{m}(x_0)$ ，从而估计出整个回归函数 $\hat{m}(x)$ 。

对 $\hat{m}(x_0)$ 估计，局部线性回归首先将回归函数 $m(x)$ 在 x_0 的很小的邻域 h 内利用泰勒规则展开：

$$m(x) = m(x_0) + m'(x_0)(x - x_0) + \cdots + \frac{1}{p!} m^{(p)}(x_0)(x - x_0)^p \quad x \in x_0 \pm h \quad (2)$$

其中，邻域大小 h 称为窗宽， p 表示泰勒展式的阶，将 (2) 代入 (1) 后得：

$$Y_i = m(x_0) + m'(x_0)(X_i - x_0) + \cdots + \frac{1}{p!} m^{(p)}(x_0)(X_i - x_0)^p \quad X_i \in x_0 \pm h \quad (3)$$

令 $\beta_k = \frac{1}{k!} m^{(k)}(x_0)$, $k = 1, 2, \dots, p$ 则 (3) 可表示为：

$$Y_i = \beta_0 + \beta_1(X_i - x_0) + \cdots + \beta_p(X_i - x_0)^p \quad X_i \in x_0 \pm h \quad (4)$$

定义权函数 $K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right)$ ，其中函数 $K(x)$ 是核函数，然后利用加权最小二乘法的思想，由

$$\min_{\beta_0, \beta_1, \dots, \beta_p} \sum_{i=1}^n [Y_i - \sum_{k=0}^p \beta_k (X_i - x_0)^k]^2 K_h(X_i - x) \quad (5)$$

可以得到 β_k , $k = 1, 2, \dots, p$ 的估计，进而由等式(4)可得估计 $\hat{m}(x_0) = \beta_0$ ，再由 x_0 的任意性和函数映射的定义，估计出整个回归函数 $\hat{m}(x)$ 。

2. 多项式阶的选择

Fan 和 Gijbels(1996)认为等式(2)中泰勒展式阶数 p 的选取，如果太大，估计的偏差会减小，但是方差会变大；反之，如果太小，估计的偏差会变大，但是方差会变小。所以， p 的选取要视估计的目的而定。估计 $m(x)$ 的 ν 阶倒数 $m^{(\nu)}(x)$ 时，选择 $p = \nu + 1$ 就可以了。因此在后面估计 $m(x)$ 中，本文选取 $p = 1$ 。

3. 窗宽的选择

窗宽 h 的选择在局部线性迭代模型估计中同样起着相当重要的作用，如果窗宽选取太大，与 x_0 距离较远的观测点也参与了局部回归，这会造成局部线性回归的偏差会较大；如果窗宽选取得过小，与 x_0 距离较近的观测点没能参与局部回归，会导致局部线性回归的方差较大。因此寻求合适的窗宽也是局部线性迭代模型估计最重要的任务之一。在这里，我们采用交叉核实方法 (Cross-Validation) 对最优窗宽进行选择：对每一个观测点 $x = X_{i_0}$ ，首先在样本中剔除该观测点 (X_{i_0}, Y_{i_0}) ，然后将剩下的 $n - 1$ 个观测点在 $x = X_{i_0}$ 处进行局部线性迭代模型回归，得到 Y_{i_0} 的估计值 \hat{Y}_{i_0} ，最后比较交叉核实函数

$$CV(h) = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (6)$$

的大小，选择使得 $CV(h)$ 达到最小的窗宽 h ，并且这个窗宽而且是渐近最优的。

4.核函数的选择

对核函数的选择，要满足如下四个条件：①定义域为 $[-1,1]$ ；②非负性： $K(x) > 0$ ；③对称性： $\int_{-1}^1 xK(x)dx = 0$ ；④ $\int_{-1}^1 K(x)dx = 1$ 。并且离观测点 x_0 距离越近，所赋予的权数就越大，相反就比较小。Fan 和 Gijbels(1996)认为选取 Epanechnikov 核函数，即 $K(x) = 0.75(1-x^2)$ 时候估计的效果最好，本文后面的估计中也选取该核函数。

2.2 局部线性迭代建模估计中国利率期限结构

对付息国债数据 $(P_i, t_i), i = 1, 2, \dots, n$ ，利用贴现函数可以把价格表示成未来现金流的贴现：

$$P_i = (100 + C)\delta(t_i) + C\delta(t_i - 1) + C\delta(t_i - 2) + \dots + C\delta(t_i - [t_i]) \quad (7)$$

其中， $\delta(t)$ 是贴现函数， t_i 为付息债券的到期时间， C 为每期利息， P_i 为债券市场价格， $[t_i]$ 为小于 t_i 的最大整数。

需要指出的是，我们不能直接采用局部线性估计的思路估计式 (7)。这是因为，对函数关系式 $y = f(x)$ ，若已知被解释变量 y 和解释变量 x 的样本数据，对任意点 x 假设在其某个很小的局部邻域内函数满足线性关系 $a + bx$ （或二次、三次多项式关系），则可利用局部线性估计思路，通过 $\min_{a,b} \sum_{i=1}^n (Y_i - ax_i + b)^2 K_h(x)$ 估计出 a 和 b ，近似地估计出函数值 $\hat{f}(x)$ ，进而基于对任意点 x 都可以估计出函数的映射值的理论，也就估计出了整个函数 $f(x)$ 。

但式 (7) 却是类似于 $y = f(x) + f(x - c_1) + f(x - c_2) + \dots$ 类的函数关系式，其中 c_1, c_2, \dots 为常数。若继续采用局部多项式的估计方法，对任意点 x 而言，即使在 x 某个很小的局部邻域内函数满足线性关系 $a + bx$ 的假设成立，但在 $x - c_i$ ($i = 1, 2, 3, \dots$) 的附近是否满足上述假设则是不知道的，进而无法得到关于 $f(x - c_i)$ 的估计。若我们假设在 $x - c_i$ ($i = 1, 2, \dots$) 的某个局部邻域内函数也满足线性关系，但对于此假设，笔者以为又过于强烈，且在多次试算过程中发现其估计效果非常不好，并且估计方法也无法利用

$\min_{a,b} \sum_{i=1}^n (Y_i - ax_i + b)^2 K_h(x)$ 用核函数加权的最小二乘思路。就笔者目前所知而言，对于

$y = f(x) + f(x - c_1) + f(x - c_2) \cdots$ 型的函数关系，尚无直接采用局部线性回归思路进行估计的文献报告。为此，笔者应用剥离息票法中迭代的思想，利用迭代的局部线性迭代建模的方法，对上述模型进行了估计。估计的基本思路为，应用剥离息票法中迭代的思想，将债券到期时间分组为 0-1 年，1-2 年，2-3 年……等。考虑到对于 (7) 而言，需要估计的是定价方程 $P_i = (100 + C)\delta(t_i) + C\delta(t_i - 1) + C\delta(t_i - 2) + \cdots + C\delta(t_i - [t_i])$ 中的贴现函数 $\delta(t)$ 。对于到期时间为 0-1 年分组的债券，定价方程为 $P_i = (100 + C)\delta(t_i)$ ，也就是

$\frac{P_i}{(100 + C)} = \delta(t_i)$ ，显然可直接采用局部线性回归的思路估计出 $\delta(t)$ 在 0-1 的值；对到期

时间为 1-2 年的债券，定价方程为 $P_i = (100 + C)\delta(t_i) + C\delta(t_i - 1)$ ，由于 $t_i - 1$ 是在区间 [0,1]

范围内，用上一步的估计代替 $\delta(t - 1)$ ，后得到 $\frac{P_i - C\delta(t_i - 1)}{(100 + C)} = \delta(t_i)$ ，左边全是已知的，

又可以用局部线性回归估计 $\delta(t)$ 在 1-2 的值；以此类推则可以得到任意区间上的 $\delta(t)$ 。具体步骤如下：

首先将数据集 $M = \{(P_i, t_i)\}$ 中的元素按 t_i 从小到大排序，然后对 $0 < t_i < 1$ 的数据，即集合 $\{(P, t) | (P, t) \in M, 0 < t < 1\}$ 中的元素，由于满足关系

$$P_i = (100 + C)\delta(t_i) \quad (8)$$

可以解出 $\delta(t_i)$ 的值，得到数据对 $(\delta(t_i), t_i)$ ，再利用局部线性迭代模型建模方法，可以估计出贴现函数 $\delta(t), t \in [0, 1]$ ，记为 $\hat{\delta}_1(t)$ 。

对 $1 < t_i < 2$ 的数据，即集合 $\{(P, t) | (P, t) \in M, 1 < t < 2\}$ 中的元素，满足关系：

$$P_i = (100 + C)\delta(t_i) + C\delta(t_i - 1) \quad (9)$$

由于 $t_i - 1 \in [0, 1]$ ，所以可以用已知函数 $\hat{\delta}_1(t_i - 1)$ ，代替 $\delta(t_i - 1)$ 估计。从而等式(9)，可以表示成：

$$P_i = (100 + C)\delta(t_i) + C\hat{\delta}_1(t_i - 1) \quad (10)$$

在等式(9)中只有 $\delta(t_i)$ 是未知的，可以解出 $\delta(t_i)$ ，从而得到 $t_i \in [1, 2]$ 的数据对 $(\delta(t_i), t_i)$ ，再结合前面估计的 [0,1] 上的数据对，这样又可以利用局部线性迭代模型建模方法，估计出贴现函数 $\delta(t), t \in [0, 2]$ ，记为 $\hat{\delta}_2(t)$ 。

如此反复，对任意的到期时间 t ，利用迭代的局部线性迭代模型建模方法可以估计出贴现函数 $\hat{\delta}(t)$ 。然后利用

$$r(t) = -\frac{\ln \delta(t)}{t} \quad (11)$$

便可以估计出任何到期时间的利率 $r(t)$ ，即整个利率期限结构。

2.3 局部线性迭代建模的统计性质

对附息国债数据 (P_i, t_i) ，当 $0 < t_i < 1$ ， $i=1, 2, \dots, n$ 时，贴现函数满足关系式(8)，利用局部线性迭代模型回归方法可以得到贴现函数在区间 $[0, 1]$ 上的估计值 $\widehat{\delta}_1(t)$ 。由 Elias 和 Fan(1997)，当 $n \rightarrow \infty$ ， $h \rightarrow 0$ 且 $nh \rightarrow \infty$ 时， $\widehat{\delta}_1(t_i)$ 是服从渐近正态分布：

$$\sqrt{nh} \left\{ [\widehat{\delta}_1(t_i) - \delta_1(t_i)] - \frac{e_1 h^2 \delta_1''(t_i)}{2} S^{-1} \mu \right\} \xrightarrow{d} N(0, e_1 \sigma(t_i) S^{-1} \widetilde{S} S^{-1} e_1^T / f(t)) \quad (12)$$

其中 $\delta_1(t)$ 为贴现函数在区间 $[0, 1]$ 上的投影， h 为窗宽，向量 $e_1 = (1, 0, 0, \dots, 0)$ ，

$$\mu = (\mu_2, \mu_3)^T, S = \begin{pmatrix} \mu_0, \mu_1 \\ \mu_1, \mu_2 \end{pmatrix}, \mu_j = \int_{-\infty}^{+\infty} u^j K(u) du, \widetilde{S} = \begin{pmatrix} v_0, v_1 \\ v_1, v_2 \end{pmatrix}, v_j = \int_{-\infty}^{+\infty} u^j K^2(u) du, f(t)$$

为随机变量 t 的分布函数。

也就是

$$p \lim E \widehat{\delta}_1(t_i) - \delta_1(t_i) = p \lim \frac{e_1 h_1 \delta_1'(t_i)}{2} S^{-1} \mu = 0 \quad (13)$$

$$p \lim \text{Var} \widehat{\delta}_1(t_i) = p \lim \frac{1}{n_1 h_1} e_1 \sigma(t_i) S^{-1} \widetilde{S} S^{-1} e_1^T / f(t) = 0 \quad (14)$$

因此， $\widehat{\delta}_1(t)$ 是 $\delta_1(t)$ 的一致估计，并且满足渐近正态分布。

当 $1 < t_j < 2$ ， $j=1, 2, \dots, n$ 时，贴现函数满足关系式(9)。按照局部线性迭代建模的思想，用 $\widehat{\delta}_1(t_i)$ 代替 $\delta(t_i - 1)$ ，再利用局部线性迭代模型估计方法得到贴现函数 $\delta(t)$ 在区间 $[1, 2]$ 上函数值的估计 $\widehat{\delta}_2(t)$ ，笔者将在附录中证明，当 $n \rightarrow \infty$ ， $h \rightarrow 0$ 且 $nh \rightarrow \infty$ 时：

$$\sqrt{nh} \left\{ [\widehat{\delta}_2(t_i) - \delta_2(t_i)] - \frac{e_1 h^2 \delta_2''(t_i)}{2} S^{-1} \mu \right\} \xrightarrow{d} N(0, e_1 \sigma(t_i) S^{-1} \widetilde{S} S^{-1} e_1^T / f(t)) \quad (15)$$

由(15)可得 $\widehat{\delta}_2(t)$ 服从渐进正态分布，并且同理可得 $p \lim E \widehat{\delta}_2(t_i) - \delta_2(t_i) = 0$ ， $p \lim \text{Var} \widehat{\delta}_2(t_i) = 0$ ，因此 $\widehat{\delta}_2(t)$ 也是 $\delta_2(t)$ 的一致估计。

重复使用以上的方法可得，用局部线性迭代建模方法估计的任意区间上的体现函数都是原贴现函数的一致估计，并且满足渐近正态分布。

2.4 蒙特卡洛模拟

在这一节中，笔者将通过蒙特卡洛模拟的方法，讨论局部线性迭代模型建模方法与样条估计方法的拟合优度，并研究局部线性迭代的统计性质。

1. 拟合优度模拟

为了模拟的方便，我们假设所有的债券到期时间均小于两年，由等式(7)，我们可以得到如下的数据生成过程：

$$p_i = (100 + c)\delta(t_i) + c\delta(t_i) + \varepsilon_i \quad \text{当 } 1 < t_i \leq 2 \quad (16)$$

$$p_i = (100 + c)\delta(t_i) + \varepsilon_i \quad \text{当 } 0 < t_i \leq 1 \quad (17)$$

其中，假设所有债券按相同的固定利率付息，即 $c = 2.5$ ；随机扰动项服从标准正态分布，即 $\varepsilon_i \sim N(0,1)$ ；到期时间 t_i 为区间 $(0, 2]$ 上的均匀随机抽样； n 为样本个数。我们利用蒙特

卡洛方法，对贴现函数取 $\delta(x) = e^{-x}$ ， $\delta(x) = \frac{1}{x^2 + 2x + 1}$ ， $\delta(x) = x^2$ ， $\delta(x) = \cos x$ 以及

样本个数 $n = 50$ ， $n = 100$ ， $n = 200$ ， $n = 500$ 的情况，按照()和()式的数据生成过程，生成数据对 (p_i, t_i) 。然后分别利用局部线性迭代建模和样条估计方法估计出贴现函数 $\hat{\delta}_1(t_i)$ 和 $\hat{\delta}_2(t_i)$ 。(图 1——图 4)

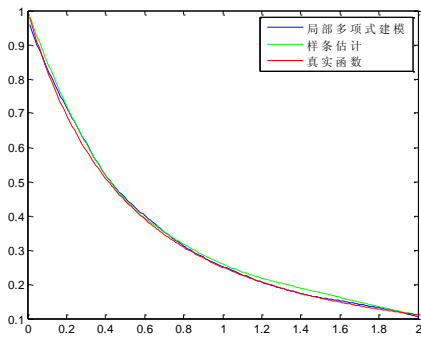


图 1 $\delta(x) = e^{-x}$

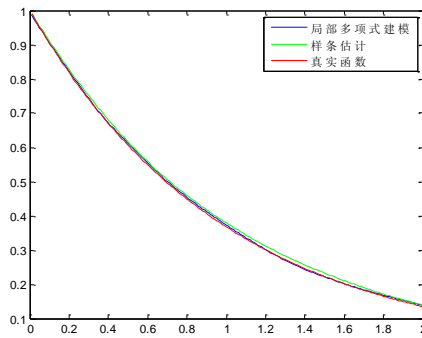


图 2 $\delta(x) = \frac{1}{x^2 + 2x + 1}$

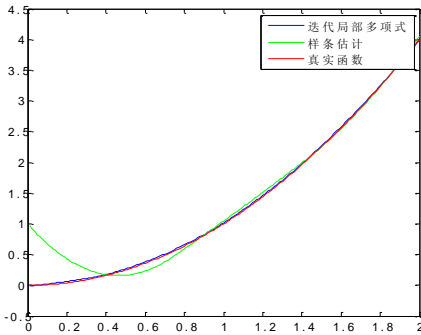


图 3 $\delta(x) = x^2$

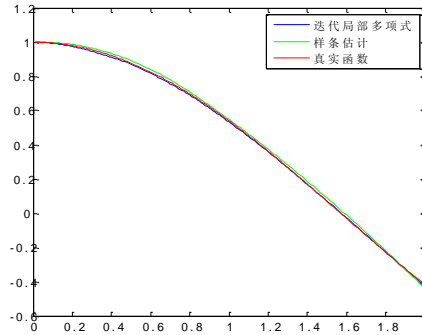


图 4 $\delta(x) = \cos x$

从图 1-图 4，可以看出，局部线性迭代建模和样条估计都可以拟合出不同函数曲线，但局部线性迭代模型估计和真实曲线更接近，拟合效果更好。

为了说明那种方法拟合的效果更好，我们让上面模拟步骤重复 $K = 100$ 次重复之后，通过对估计值和真实值的比较，计算平均绝对偏差 (MADE)

$$MADE_j = \frac{1}{K} \sum_{k=1}^K \frac{1}{n} \sum_{i=1}^n |\hat{\delta}_j(t_i) - \delta(t_i)| \quad \text{其中 } j = 1, 2; \quad (18)$$

来说明两种估计方法的拟合优度。(表 1)

表 1 两种估计方法拟合优度的比较

	$\delta(x) = e^{-x}$		$\delta(x) = \frac{1}{x}$		$\delta(x) = x^2$		$\delta(x) = \cos x$	
	$MADE_1$	$MADE_2$	$MADE_1$	$MADE_2$	$MADE_1$	$MADE_2$	$MADE_1$	$MADE_2$
$n = 50$	0.0037	0.0111	0.006	0.0128	0.0136	0.112	0.0044	0.0114
$n = 100$	0.0035	0.0113	0.0061	0.0125	0.0148	0.119	0.0044	0.0110
$n = 200$	0.0034	0.0112	0.0059	0.0122	0.0151	0.1219	0.0041	0.0112
$n = 500$	0.0033	0.0111	0.0061	0.0124	0.0153	0.1224	0.0042	0.0111

注： $MADE_1$:局部线性迭代建模拟合的平均绝对偏差； $MADE_2$:样条估计的平均绝对偏差。

从表 1 中可以看出，对不同的贴现函数，通过局部线性迭代建模方法计算的平均绝对偏差明显小于样条方法的平均绝对偏差。也就是说，局部线性迭代建模方法拟合的曲线与真实曲线更加接近，拟合效果更好。另外，从表 5 还可以看出，对不同的样本数量，局部线性迭代估计建模方法的拟合效果变化不大，因此该方法在国债品种较少，样本数量较小的中国利率市场上也有很大的实用性。

2. 统计性质模拟

我们将通过蒙特卡洛模拟的方法，说明局部线性迭代建模估计具有渐近正态的优良统计性质。不失一般性，选取点 $t = 1.5$ 。首先按照(16)和(17)式的数据生成过程，生成数据对 (p_i, t_i) ，然后利用局部线性迭代建模方法估计出贴现函数 $\hat{\delta}(1.5)$ 的值。重复 50000 次后，可以得到 50000 个 $\hat{\delta}(1.5)$ 的估计，通过统计估计值出现在不同的区间上的个数，模拟出 $\hat{\delta}(1.5)$ 分布的密度函数（图 5）。

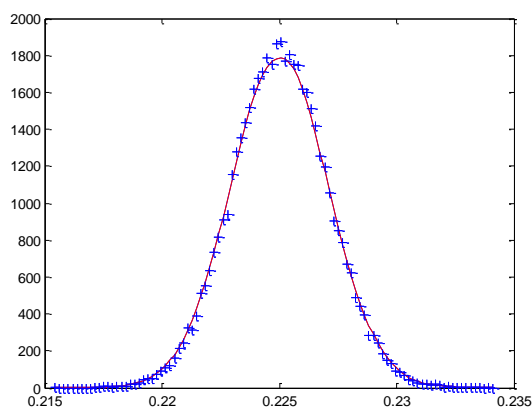


图 5 局部线性迭代的渐近正态分布

从图 5 中可以看出，当样本数趋于无穷大的时候，局部线性迭代建模估计将会趋于正态分布。也说明了局部线性迭代具有优良的统计性质，可以做假设检验，能够应用到更广泛的领域。

2.5 样本与数据

本文选取 2010 年 6 月 17 日上海证券交易所的 36 种付息国债交易价格数据（6 月 17 日共有 37 种国债挂牌交易，其中债券 010706 到期时间为 26 年和其他债券相差甚远，所含信

息比较少，对估计意义不大，故将其排除在外)。由于 2002 年 3 月 25 日以后，国债交易价格是以扣除累计利息的净价法表示，因此首先要加上这些债券的累计利息，得出真实的息票债券价格。2010 年 6 月 17 日上海证券交易所的国债交易如表 1 所示，数据来源于上海证券交易所网站。

表 2 中国上海证券交易所国债交易情况

2010-6-17

债券代码	年利率 (%)	剩余期限 (年)	全价 (元)	债券本金 (元)	付息方式
010307	2.66	0.18	102.57	100	1
010511	2.14	0.34	102.37	100	1
010311	3.5	0.42	103.02	100	1
010605	2.4	0.91	100.72	100	1
010404	4.89	0.94	103.5	100	1
010612	2.72	1.16	102.29	100	1
010407	4.71	1.19	107.62	100	1
010110	2.95	1.27	103.35	100	1
010618	2.48	1.36	104.8	100	1
010112	3.05	1.37	103.25	100	1
010410	4.86	1.44	107.23	100	1
010203	2.54	1.84	101.01	100	1
010705	3.18	1.85	100.49	100	1
010505	3.37	1.94	102.02	100	1
010509	2.83	2.19	103.18	100	1
010513	3.01	2.44	102.99	100	1
010601	2.51	2.7	101.36	100	1
010606	2.62	2.94	100.17	100	1
010613	2.89	3.21	105.3	100	1
010308	3.02	3.25	104.38	100	1
010620	2.91	3.45	101.62	100	1
010701	2.93	3.64	101.06	100	1
010707	3.74	3.94	110.26	100	1
010501	4.44	4.7	108.15	100	2
010603	2.8	5.78	98.82	100	2
010616	2.92	6.28	101.06	100	2
010703	3.4	6.77	100.82	100	1
010710	4.4	7.03	116.26	100	2
010213	2.6	7.27	97.64	100	2
010512	3.65	10.42	101	100	2
010107	4.26	11.13	109.07	100	2

010619	3.27	11.42	101.91	100	1
010303	3.4	12.84	98.43	100	2
010504	4.11	14.92	105.21	100	2
010609	3.7	16.04	101.77	100	2
010713	4.52	17.18	101.89	100	2

注：付息方式 1 表示每年付息一次；2 表示每年付息两次

3 实证分析

笔者选取 2010 年 6 月 17 日上海证券交易所的国债交易价格数据，分别采用局部线性迭代回归估计方法和现在国内广为流行的样条估计方法对我国的利率期限结构进行静态的估计分析。并用交叉核实检验对两种方法进行比较。

3.1 基于局部线性迭代估计和样条估计的结果

分别采用局部线性迭代回归估计方法和样条估计方法，利用表 2 中的数据对中国市场中的利率期限结构进行估计。其中在局部线性迭代回归时，选用 Epanechnikov 核函数，利用交叉核实法选择窗宽，利用一阶多项式拟合贴现值和利率曲线。利用样条方法估计时，选择的节点数 $k = 4$ 。拟合的贴现值、利率水平以及利率曲线见表 3、表 4 和图 6：

表 3 局部线性迭代回归估计贴现值和利率水平

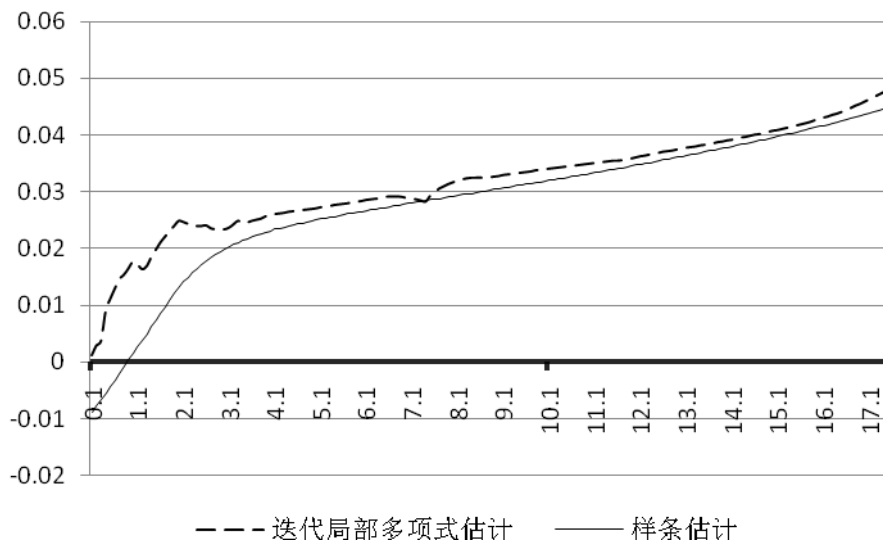
期限	贴现值	利率水平	期限	贴现值	利率水平	期限	贴现值	利率水平
0.18	0.9991	0.0049	1.85	0.9435	0.0314	5.78	0.8309	0.0321
0.34	1.0023	-0.0066	1.94	0.9549	0.0238	6.28	0.8338	0.0289
0.42	0.9954	0.0111	2.19	0.9489	0.0239	6.77	0.7909	0.0346
0.91	0.9836	0.0182	2.44	0.9423	0.0244	7.03	0.8588	0.0217
0.94	0.9867	0.0142	2.7	0.941	0.0225	7.27	0.7984	0.031
1.16	0.9693	0.0268	2.94	0.9267	0.0259	10.42	0.6757	0.0376
1.19	0.9828	0.0146	3.21	0.9412	0.0189	11.13	0.6701	0.036
1.27	0.9753	0.0197	3.25	0.9274	0.0232	11.42	0.6856	0.033
1.36	0.9985	0.0011	3.45	0.9051	0.0289	12.84	0.6176	0.0375
1.37	0.9724	0.0204	3.64	0.8994	0.0291	14.92	0.561	0.0387
1.44	0.9765	0.0165	3.94	0.9593	0.0105	16.04	0.5413	0.0383
1.84	0.9606	0.0218	4.7	0.8715	0.0293	17.18	0.4179	0.0508

表 4 样条方法估计贴现值和利率水平

期限	贴现值	利率水平	期限	贴现值	利率水平	期限	贴现值	利率水平
0.18	1.0004	-0.0023	1.85	0.9796	0.0111	5.78	0.8566	0.0268
0.34	1.0004	-0.001	1.94	0.9773	0.0118	6.28	0.8388	0.028

0.42	1.0002	-0.0004	2.19	0.9707	0.0136	6.77	0.8216	0.029
0.91	0.9968	0.0035	2.44	0.964	0.015	7.03	0.8125	0.0295
0.94	0.9965	0.0037	2.7	0.9568	0.0164	7.27	0.8042	0.03
1.16	0.9936	0.0055	2.94	0.9499	0.0175	10.42	0.6994	0.0343
1.19	0.9932	0.0058	3.21	0.942	0.0186	11.13	0.6768	0.0351
1.27	0.9919	0.0064	3.25	0.9408	0.0188	11.42	0.6677	0.0354
1.36	0.9904	0.0071	3.45	0.9348	0.0195	12.84	0.6242	0.0367
1.37	0.9902	0.0072	3.64	0.929	0.0202	14.92	0.5634	0.0385
1.44	0.9889	0.0078	3.94	0.9196	0.0213	16.04	0.5321	0.0393
1.84	0.9799	0.011	4.7	0.8947	0.0237	17.18	0.5013	0.0402

图 6 两种方法拟合中国利率期限结构



由图 6，我们可以看出，两种方法估计出的利率曲线形状大致相同，都是一条略微向上倾斜的曲线，随着到期时间的增加，利率水平也不断增大，呈长期利率高于短期利率的正向势头。符合流动性溢酬的假设。两种方法估计的短期利率变化较大，而中长期相对比较平缓，这也和短期债券流动性强，交易方便，久期较短风险小等实际特点相符合。因此，局部线性迭代建模方法和样条估计方法一样都能都对国内利率期限结构进行比较合理的估计。

另外，分别从两种估计方法上来看。样条估计方法估计的短期利率，特别是在期限很小的时候为负数，与实际不相符合。这正是因为样条估计方法是对每段区间用一个确定的样条函数拟合，而国内债券市场数据存在的异常值会对该段期限结构的估计造成很大的影响。选用局部线性迭代建模方法，是对利率期限上的每个点建模，从而可以有效的降低异常值对拟合曲线的影响。估计出来的利率水平全部为正，更加符合实际。所以局部线性迭代建模方法更适合我国利率市场的特点，对中国市场的利率期限结构估计效果更好。

3.2 两种估计方法比较

本文利用交叉核实检验法，对两种方法的拟合效果进行比较：首先对每一对关于到期时

间 T_{i_0} 和债券价格 P_{i_0} 的观测数据 (T_{i_0}, P_{i_0}) ，在样本中剔除该观测点；然后利用剩下的 $n-1$ 个样本数据，分别采用局部线性迭代回归和样条估计方法估计出整个贴现函数 $\hat{\delta}_{-i_0}(T)$ ；再由方程(7)估计出到期期限为 T_{i_0} 的债券价格 \hat{P}_{i_0} （见表 5）；最后比较估计价格和真实价格的差异（见图 7），计算交叉核实函数：

$$CV = \frac{1}{n} \sum_{i=1}^n (P_i - \hat{P}_i)^2 \quad (19)$$

通过比较 CV 值来比较两种方法的拟合效果。 CV 值越小，表明估计的债券价格和真实价格之间的平均差异越小，估计越准确，模型拟合效果越好。

表 5 两种方法估计的真实价格

期限	价格 1	价格 2	真实价格	期限	价格 1	价格 2	真实价格
0.18	102.7925	102.6291	102.57	3.21	104.549	103.1184	105.3
0.34	102.3203	101.7342	102.37	3.25	105.0491	103.816	104.38
0.42	103.6827	103.1149	103.02	3.45	104.2038	103.7542	101.62
0.91	102.4732	100.9091	100.72	3.64	103.7248	102.634	101.06
0.94	104.8991	103.0914	103.5	3.94	105.0191	103.7768	110.26
1.16	105.086	103.6661	102.29	4.7	110.4314	109.2346	108.15
1.19	108.7	107.3431	107.62	5.78	101.9457	100.888	98.82
1.27	105.2753	103.8369	103.35	6.28	102.2102	100.9294	101.06
1.36	103.9519	102.1141	104.8	6.77	105.1948	104.5161	100.82
1.37	105.2972	103.6253	103.25	7.03	111.9971	109.8484	116.26
1.44	108.6036	106.8021	107.23	7.27	99.4688	99.9916	97.64
1.84	102.9955	100.6111	101.01	10.42	105.2126	103.2552	101
1.85	104.3342	102.2784	100.49	11.13	111.1617	109.3184	109.07
1.94	104.3245	101.8638	102.02	11.42	100.9014	99.1992	101.91
2.19	105.1223	103.0533	103.18	12.84	99.8361	98.0741	98.43
2.44	104.853	103.1339	102.99	14.92	104.3265	102.3823	105.21
2.7	102.5339	100.9493	101.36	16.04	97.331	95.2635	101.77
2.94	102.1752	101.1446	100.17	17.18	111.4625	110.9831	101.89

注：价格 1 是用样条估计方法估计的债券价格，价格 2 使用迭代局部多形式建模方法估计的债券价格

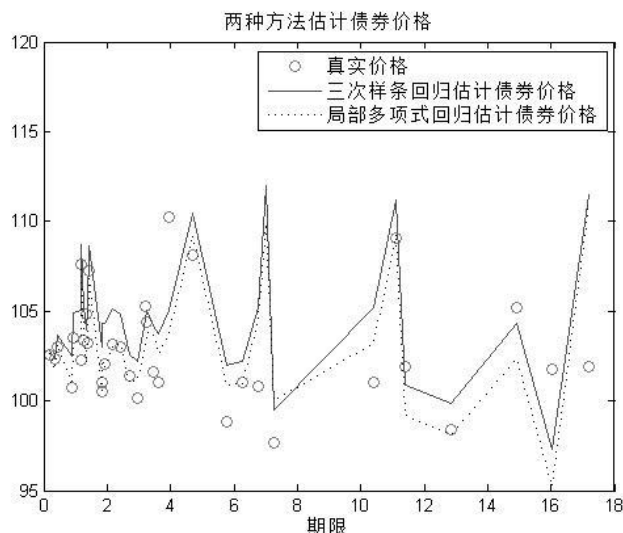


图 7 两种方法估计债券价格比较

表 6 两种方法交差核实检验结果

	局部线性迭代回归	样条估计方法
CV 值	7.79	8.28

由图 7，我们可以看出，采用两种方法估计的债券价格，局部线性迭代回归估计的债券价格比样条估计方法与真实价格更加接近，拟合效果更好。并且由表 5 交差核实检验的结果，也可以看出用局部线性迭代回归方法计算出的交叉核实函数为 7.79，用样条估计方法得到的交叉核实函数为 8.28。局部线性迭代回归的核实函数要小于样条估计方法，因此局部线性迭代估计方法的拟合效果比样条估计更优。

4 结 论

本文根据国内债券市场上债券品种不多，可供观测的样本量有限，期限分布不均匀，大都为付息债券，且定价偏差较大，样本异常点较多等特点，跳出对国外发达国家估计方法的沿用，在局部线性回归非参数统计方法的基础上，提出了适合我国债券市场的局部线性迭代建模方法。选取 2010 年 6 月 17 日上海证券交易所国债数据，用局部线性迭代建模实证分析了我国利率期限结构，并用交差核实检验与样条方法相比较。得出结论主要有：（1）局部线性迭代建模和现有样条估计两种方法估计的利率曲线形状大致相同，均为略向上倾斜的曲线，和实际相吻合。表明局部线性迭代建模方法是合理的。（2）国内利率数据的异常点会对样条估计方法拟合结果造成扭曲，短期利率出现负的利率水平，与实际不相符合。而局部线性迭代建模可以有效的降低异常点的影响，更适合国内利率市场。（3）利用交差核实检验说明局部线性迭代模型拟合比样条估计更加准确，效果更好。（4）从统计性质上来说，样条估计只是纯粹数量上的拟合，缺乏统计性质，而局部线性迭代模型具有渐进正态的优良统计性质，方法更加可靠，应用更加广泛。

总之，局部线性迭代建模方法能够更加准确、合理的估计出国内的利率期限结构，有效的降低国债数据中的异常点的影响，且具有优良的统计性质，弥补了原有方法纯粹从数量上拟合的不足，更加可靠，更加适应中国债券市场的特点。

参考文献

- [1] Bliss R R, 1997, Testing term structure estimation methods [J], *Advances in Futures and Options Research*, 9(1):197-231.
- [2] Elias M, Fan J, 1997, Local Polynomial Estimation of Regression Functions for Mixing Processes [J], *Scandinavian Journal of Statistics*, 24(2):165-179.
- [3] Fama E F, Bliss R R, 1987, The information in long-maturity forward rates [J], *American Economic Review*, 77(04):680-692.
- [4] Fan J, Gijbels I, 1996, *Local Polynomial Modeling and Its Application*[M], Chapman&Hall
- [5] Fisher M, Nychka D, Zervos D, 1995, Fitting the Term Structure of Interest Rates with Smoothing Splines [R], Washington:Division of Research and Statistics, Federal Reserve Board, 1-15.
- [6] McCulloch J H, 1971, Measuring the term structure of interest rates [J], *Journal of Business*, 44(1):19-31.
- [7] McCulloch J H, 1975, The tax-adjusted yield curve [J], *Journal of Finance*, 30(3):811-830.
- [8] Nelson C R, Siegel A F, 1987, Parsimonious modeling of yield curves [J], *Journal of Business*, 60(4):473-489.
- [9] Shea G S, 1984, Pitfalls in smoothing interest rate term structure data: Equilibrium models and spline approximations [J], *Journal of Finance and Quantitative Analysis*, 19(3):253-269.
- [10] Svensson L, 1995, Estimating forward interest rates with the extended Nelson & Siegel method [J], *Sveriges Riksbank Quarterly Review*, 3(1):13-26.
- [11] Vasicek O A, Fong H G, 1982, Term structure modeling using exponential splines [J], *Journal of Finance*, 37(2):339-356.
- [12] Waggoner D, 1997, *Spline Methods for Extracting Interest Rate Curves from Coupon Bond Prices* [R], Georgia: Federal Reserve Bank of Atlanta, 1-10.
- [13] 陈雯, 陈浪南, 2000, 国债利率期限结构:建模和实证[J], *世界经济*, 23(8):24-28.
- [14] 程希骏, 萧楠, 2008, 基于稳健估计的样条函数法对国债利率期限结构的拟合[J], *中国科学技术大学学报*, 36(12):1275-1280
- [15] 傅曼丽, 董荣杰, 屠梅曾, 2005, 国债利率期限结构模型的实证比较, *系统工程*, 23(8):56-61.
- [16] 何启志, 何建敏, 陈珊珊, 2008, 利率期限结构指数样条模型实证研究[J], *管理科学*, 21(1): 100-107.
- [17] 胡海鹏, 方兆本, 2009, 中国利率期限结构平滑样条拟合改进研究[J], *管理科学*, 12(1): 101-111.
- [18] 李熠熠, 潘婉彬, 缪柏其, 2010, 基于最小一乘准则的三次样条对利率期限结构的拟合[J], *数理统计与管理*, 29(1):170-174
- [19] 李竹渝, 鲁万波, 龚金国, 2007, *经济、金融计量学中的非参数估计技术*[M], 科学出版社.
- [20] 刘灿, 易璐, 2004, 深沪两市国债收益率期限结构的实证研究-B 样条函数法[J], *证券市场导报*, 14(2): 36—141.
- [21] 鲁万波, 李竹渝, 2006, 波动性的非参数局部线性迭代模型估计[j], *四川大学学报*, 43(2): 243-248
- [22] 全晓燕, 程希骏, 2007, 基于 B 样条对国债利率期限结构的实证研究[J], *系统工程*, 25(3): 17-22.
- [23] 王晓芳, 刘凤根, 韩龙, 2005, 基于三次样条函数的中国国债利率期限结构曲线构造[J], 23(6):: 85-89.
- [24] 吴泽福, 吴捷, 2010, 基于 L1 样条与 B 样条改进的利率期限结构估计优化[J], *华侨大学学报 (哲学社会科学版)*, 2: 49-57.

- [25] 杨春鹏, 曹兴华, 2002, 我国国债收益率曲线的构造与实证研究[J], 投资与证券, 12(10):80-82.
- [26] 杨大楷, 王欢, 1999, 关于我国国债收益率曲线的再研究[J], 扬州大学税务学院学报, 4(3):50-54.
- [27] 姚长辉, 梁跃军, 1998, 我国国债收益率曲线的实证研究[J], 金融研究, 19(08):12-18.
- [28] 郑振龙, 林海, 2003, 中国市场利率期限结构的静态估计[J], 武汉金融, 4(3):33-36.
- [29] 周荣喜, 车君, 2010, 基于利率期限结构模型的净现值公式[J], 北京化工大学学报(自然科学版), 37(1):130-134.
- [30] 朱淑英, 刘伶, 2010, 中国国债的利率期限结构的实证研究: B-样条法[J], 法制与经济, 3:62-67.

附 录

局部线性迭代估计渐进正态性((15)式)的证明:

对附息国债数据 $(P_i, t_i), i=1, 2, \dots, n$, 当 $1 < t_j < 2$, $j=1, 2, \dots, n$ 时, 我们首先令:

$$\hat{y}_i = \frac{P_i - C\hat{\delta}_1(t_i - 1)}{100 + C}, y_i = \frac{P_i - C\delta(t_i - 1)}{100 + C}, Y = (y_1, y_2, \dots, y_n)^T, \hat{Y} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)^T$$

$$\beta = (\beta_0, \beta_1)^T, \beta_0 = \delta_2(t), \beta_1 = \delta_2'(t), X_2 = \begin{pmatrix} 1 & t_1 - T_0 \\ 1 & t_2 - T_0 \\ \vdots & \vdots \\ 1 & t_n - T_0 \end{pmatrix}, W_2 = \text{diag}\{K_h(t_i - T_0)\},$$

$$H = \text{diag}\{1, h\}, s_{n,j} = \frac{1}{n} \sum_{i=1}^n \left(\frac{t_i - t}{h}\right)^j K_h(t_i - t), t_{n,j}^* = \frac{1}{n} \sum_{i=1}^n \left(\frac{t_i - t}{h}\right)^j K_h(t_i - t)(Y_i - \delta_2(t_i)),$$

$$S_n = \begin{pmatrix} s_{n,0}, s_{n,1} \\ s_{n,1}, s_{n,2} \end{pmatrix}, t_n^* = \begin{pmatrix} t_{n,0}^* \\ t_{n,1}^* \end{pmatrix}$$

由 Fan(1996)可得:

$$\begin{aligned} \hat{\beta} &= (X_2^T W_2 X_2)^{-1} X_2^T W \hat{Y} \\ &= (X_2^T W_2 X_2)^{-1} X_2^T W (\hat{Y} - Y + Y) \\ &= (X_2^T W_2 X_2)^{-1} X_2^T W Y + (X_2^T W_2 X_2)^{-1} X_2^T W (\hat{Y} - Y) \end{aligned} \quad (20)$$

将 $\delta_2(t_i)$ 在点 t 按照泰勒规则展开后得:

$$\begin{aligned} \delta &= (\delta_2(t_1), \delta_2(t_2), \dots, \delta_2(t_n))^T \\ &= X_2 \beta + \frac{\delta_2''(t)}{2!} ((t_1 - t)^2, \dots, (t_n - t)^2)^T + o_p(h^2) \end{aligned} \quad (21)$$

于是:

$$\begin{aligned} (X_2^T W_2 X_2)^{-1} X_2^T W Y &= (X_2^T W_2 X_2)^{-1} X_2^T W (Y - \delta + \delta) \\ &= (X_2^T W_2 X_2)^{-1} X_2^T W (Y - \delta) + (X_2^T W_2 X_2)^{-1} X_2^T W \delta \end{aligned} \quad (22)$$

由 Ruppert 和 Wand(1994)可知 $X_2^T W_2 X_2 = nHS_n H$, 将(18)式代入到(19)前半部分中, 通过矩阵的乘法, 有:

$$\begin{aligned} (X_2^T W_2 X_2)^{-1} X_2^T W(Y - \delta) &= n^{-1} H^{-1} S_n^{-1} H^{-1} n H t_n^* \\ &= H^{-1} S_n^{-1} t_n^* \end{aligned} \quad (23)$$

另外将(18)式代入到(19)后半部分中, 通过矩阵的乘法可得:

$$\begin{aligned} &(X_2^T W_2 X_2)^{-1} X_2^T W \delta \\ &= (X_2^T W_2 X_2)^{-1} X_2^T W \left\{ X_2 \beta + \frac{\delta_2''(t)}{3!} ((t_1 - t)^2, \dots, (t_n - t)^2)^T + o_p(h^2) \right\} \\ &= \beta + \frac{\delta_2''(t)}{3!} (X_2^T W_2 X_2)^{-1} X_2^T W ((t_1 - t)^2, \dots, (t_n - t)^2)^T + o_p(h^2) \\ &= \beta + \frac{\delta_2''(t)}{3!} n^{-1} H^{-1} S_n^{-1} H^{-1} n H \mu + o_p(h^2) \\ &= \beta + \frac{\delta_2''(t)}{3!} H^{-1} S_n^{-1} \mu + o_p(h^2) \end{aligned} \quad (24)$$

因此, 综合(17)-(21)式可得:

$$H(\hat{\beta} - \beta) - \frac{\delta_2''(t)}{2!} S_n^{-1} \mu = S_n^{-1} t_n^* + H(X_2^T W_2 X_2)^{-1} X_2^T W(\hat{Y} - Y) \quad (25)$$

下面我们将会说明, 当 $n \rightarrow \infty$, $h \rightarrow 0$ 且 $nh \rightarrow \infty$ 时(23)式成立。

$$p \lim \sqrt{nh} H(X_2^T W_2 X_2)^{-1} X_2^T W(\hat{Y} - Y) = 0 \quad (26)$$

事实上, 由于 $\hat{Y} - Y = \frac{C}{100+C} (\delta_1(t_1 - 1) - \hat{\delta}_1(t_1 - 1), \dots, \delta_1(t_n - 1) - \hat{\delta}_1(t_n - 1))^T$, 所以

$$\begin{aligned} &\sqrt{nh} H(X_2^T W_2 X_2)^{-1} X_2^T W(\hat{Y} - Y) \\ &= \frac{C n^{-1}}{100+C} S_n^{-1} H^{-1} X_2^T W_2 \sqrt{nh} (\delta_1(t_1 - 1) - \hat{\delta}_1(t_1 - 1), \dots, \delta_1(t_n - 1) - \hat{\delta}_1(t_n - 1))^T \end{aligned} \quad (27)$$

由 Elias 和 Fan(1997), 可知

$$\sqrt{nh} [\hat{\delta}_1(t_i) - \delta_1(t_i)] = S_n^{-1} t_n^* + \frac{e_1 h^2 \delta_1''(t_i)}{2} S^{-1} \mu \quad (28)$$

将(25)代入到(24), 并通过矩阵运算可得:

$$\sqrt{nh} H(X_2^T W_2 X_2)^{-1} X_2^T W(\hat{Y} - Y) = -\frac{C}{100+C} S_n^{-1} (V_1 + V_2) \quad (29)$$

其中 V_1 和 V_2 均为 2×1 维的列向量, 第 $i=1, 2$ 个元素为:

$$V_{1i} = e_1 S_n^{-1} t_n^* \frac{1}{n} \sum_{k=1}^n \left(\frac{t_k - t}{h} \right)^i K_h(t_k - t) = e_1 S_n^{-1} t_{n,i}^*$$

$$V_{2i} = h^2 \frac{1}{n} \sum_{k=1}^n \left(\frac{t_k - t}{h} \right)^i K_h(t_k - t) e_1 \frac{\delta_1''(t_k)}{2!} S^{-1} \mu = h^2 \frac{e_1 S^{-1} \mu}{2!} \frac{1}{n} \sum_{k=1}^n \left(\frac{t_k - t}{h} \right)^i K_h(t_k - t) \delta_1''(t_k)$$

由 Elias 和 Fan(1997)中定理 3 可得, 当 $n \rightarrow \infty$, $h \rightarrow 0$ 且 $nh \rightarrow \infty$ 时有:

$$\sqrt{nh} S_n^{-1} t_n^* \xrightarrow{d} N(0, \sigma^2(t) S^{-1} \tilde{S} S^{-1} / f(t)) \quad (30)$$

$$p \lim S_n = f(t) S \quad (31)$$

也就是说 $p\lim S_n^{-1}t_n^* = 0$, S_n 为 $O_p(1)$ 。再加上 $s_{n,i}$ 和 $\frac{e_1 S^{-1} \mu}{2!} \frac{1}{n} \sum_{k=1}^n \left(\frac{t_k - t}{h}\right)^i K_h(t_k - t) \delta_1''(t_k)$

为 $O_p(1)$, 所以 $p\lim V_{1i} = 0$ 且 $p\lim V_{2i} = 0$ 。因此由(26)式可得(23)式成立。

那么, 由(22), (23)和(27)式, 当 $n \rightarrow \infty$, $h \rightarrow 0$ 且 $nh \rightarrow \infty$ 时, 可得

$$\sqrt{nh} \left(H(\hat{\beta} - \beta) - \frac{\delta_2''(t)}{2!} S_n^{-1} \mu \right) = \sqrt{nh} S_n^{-1} t_n^* \xrightarrow{d} N(0, \sigma^2(t) S^{-1} \tilde{S} S^{-1} / f(t)) \quad (32)$$

(QED)