

基于 Q-SCAD 的样条函数法拟联合国债利率期限结构

郭键鸿,程希骏,刘峰

(中国科学技术大学管理学院统计与金融系,安徽合肥 230026)

摘要:将惩罚分位数回归 SCAD 方法引入样条函数并以此来构建国债利率期限结构模型.该方法可以实现自动选取最优分位数,并同步完成模型中的节点选择和参数估计.样本外预测结果显示,与传统的方法相比,新方法可以有效地选择合适的模型,增加参数估计的稳健性,提高预测的精度,增强利率期限结构定价的准确度.

关键词:期限结构;样条函数;节点选择;Q-SCAD

中图分类号:F830.9 **文献标识码:**A **doi:**10.3969/j.issn.0253-2778.2015.03.012

引用格式: Guo Jianhong, Cheng Xijun, Liu Feng. Building the term structure of interest rates with splines based on Q-SCAD[J]. Journal of University of Science and Technology of China, 2015, 45(3): 254-258.

郭键鸿,程希骏,刘峰. 基于 Q-SCAD 的样条函数法拟联合国债利率期限结构[J]. 中国科学技术大学学报, 2015, 45(3): 254-258.

Building the term structure of interest rates with splines based on Q-SCAD

GUO Jianhong, CHENG Xijun, LIU Feng

(Dept. of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei 230026, China)

Abstract: Penalized quantile regression with the SCAD was introduced into the spline function to build the term structure of interest rates of treasury bills. This method can automatically select the optimal quantile and complete the model's knot selection and parameter estimation at the same time. The out-of-sample forecasting results show that, compared with the traditional methods, the new method can select the appropriate model effectively, increase the robustness of parameter estimation, improve forecast accuracy and enhance the pricing precision of the term structure of interest rates.

Key words: term structure; spline function; knot selection; Q-SCAD

0 引言

利率期限结构是金融市场上最基本也是最重要的经济变量之一,用样条函数来拟合债券的利率期限结构具有灵活性好、简单易行的特点,一直受到中

外研究者的广泛关注^[1-5]. 而对于样条函数的研究,一个核心问题就是节点的选择,包括节点的个数和节点的位置. McCulloch^[1]指出样条函数的节点个数近似为被估计样本个数的平方根,同时任意两个相邻节点间的债券个数要相近或相等.但是这种过

收稿日期:2013-10-10;修回日期:2014-01-25

基金项目:中国科学院知识创新工程重要方向项目(KJCX3-SYW-S02)资助.

作者简介:郭键鸿,男,1988年生,硕士.研究方向:金融工程. E-mail: gjh688@qq.com

通讯作者:程希骏,博士/副教授. E-mail: xjc@ustc.edu.cn

于主观的方法可能存在过度拟合的弊端. 孙增献等^[3]用分位数回归的样条函数法拟合利率期限结构,发现分位数为 0.5 时分位数回归比最小二乘回归得到的利率期限结构更稳定. 但是该文中样条函数的节点选择是人工确定的,分位数的选取还是显得主观. 李熠熠等^[4]用逐点删除方法选择节点,先均匀地取较多的点,然后再逐渐删去每个节点,根据 AIC 准则得到节点,之后再次进行参数估计. 该方法很大程度上提高了准确度,但是计算量比较大. 李熠熠等^[5]也曾用 LAD-LASSO 的方法来构建利率期限结构,虽然该方法可以同时选择节点和参数估计并且具有稳健性,但是构建的罚函数容易产生不连续的解,导致模型不稳定^[6],并且 LAD 仅仅是分位数回归的一种特殊情况.

为了克服这些不足,本文在吸收 Fan^[7]提出的 SCAD(smoothing clipped absolute deviation)理论的基础上,将惩罚分位数回归的 SCAD 方法(penalized quantile regression with the SCAD, Q-SCAD)^[8]引入样条函数中,通过最优分位数的自动选取,达到节点选择和参数估计的同步实现,并以此来构建利率期限结构. 与传统方法相比,首先,该方法避免了人为选择分位数和节点,将主观因素的影响降到最低,同时使得计算过程简洁和自动化;其次,在该模型中引入分位数回归的思想,则扩展了 LAD 回归方法;最后,与 LASSO 惩罚项相比,SCAD 惩罚项具有无偏性,且可以克服系统的误差. 结果表明,该模型具有很强的稳健性,同时还降低了异常值对参数估计的影响,避免了过度拟合,因而使得构建的利率曲线具有更好的平滑性.

1 模型的构建

设贴现因子 $B(t)$ 可表示为一个三次样条函数且满足

$$\{B_i(t) = d_i + c_i t + b_i t^2 + a_i t^3 \mid t \in [t_i, t_{i+1}], \\ i = 0, 1, \dots, L-1\} \quad (1)$$

式中, t 为到期时间, t_1, t_2, \dots, t_L 为样条节点, a_i, b_i, c_i, d_i 为待估参数. 为了保证分段函数的光滑性及在分段点处的连续性,样条函数还需满足如下的约束条件:

$$\left. \begin{aligned} B_i(t_i) &= B_{i+1}(t_i); \\ B'_i(t_i) &= B'_{i+1}(t_i), i = 0, 1, \dots, L-1; \\ B''_i(t_i) &= B''_{i+1}(t_i) \end{aligned} \right\} \quad (2)$$

式中, $B'_i(t_i), B''_i(t_i)$ 分别表示 $B(t)$ 的一阶导数和二

阶导数. 结合式(1)和(2)可将样条函数改写为如下的形式:

$$B(t) = \sum_{l=1}^{L+4} \beta_l P_l(t) \quad (3)$$

式中, β_l 为待估参数, $(\beta_1, \dots, \beta_L)$ 为节点参数,它的大小表示了此节点在样条函数中作用的大小, $(\beta_{L+1}, \dots, \beta_{L+4})$ 为 t^0, t^1, t^2 和 t^3 的参数; $\{P_l(t)\}$ 为 3 阶样条函数在 t 时刻组成的 $L+4$ 维的线性空间的基底,称之为功效基底(power basis),具有如下形式:

$$\left. \begin{aligned} P_l(t) &= (t - t_l)_+^3, l = 1, \dots, L; \\ P_{L+l}(t) &= t^{l-1}, l = 1, 2, 3, 4 \end{aligned} \right\} \quad (4)$$

我们考虑第 i 只付息债券的当前价格 Y_i , 其贴现定价的定价公式为

$$Y_i = \sum_{m=1}^{N_i} C_i(t_m) B(t_m) \quad (5)$$

式中, N_i 表示国债的付息次数, L 为预选的节点数目; $C_i(t_m)$ 表示国债在 t_m 时刻的付息额,末次支付即为本息和. 将式(3)代入式(5),可以得到

$$Y_i = \sum_{l=1}^{L+4} [\beta_l \sum_{m=1}^{N_i} C_i(t_m) P_l(t_m)] \quad (6)$$

对于式(6),一般情形下使用最小二乘法(LS)来进行参数估计,但是最小二乘法不是稳健的估计方法. 由于金融数据中异常值是比较常见的,并且它的误差分布与正态分布相比往往具有厚尾性,如果运用最小二乘方法来估计参数,会导致模型的预测效果不佳. 在最小二乘法的基础上加入惩罚项,最小化下式可以解决变量的选择问题:

$$\sum_{i=1}^n (Y_i - \sum_{l=1}^{L+4} [\beta_l \sum_{m=1}^{N_i} C_i(t_m) P_l(t_m)])^2 + n \sum_{l=1}^{L+4} p_\lambda(|\beta_l|) \quad (7)$$

式中, $\lambda > 0$ 为调整系数, n 为总只数.

当惩罚函数满足 $p_\lambda(|\beta|) = \lambda |\beta|$, 即为 L_1 惩罚函数时,式(7)为由 L_1 惩罚函数驱动的程序 Lasso^[6],但是该模型对于较大的系数 β 会造成不必要的偏倚,容易导致模型的偏差过大;当惩罚函数满足

$$p'_\lambda(\beta) = \lambda \{ I(\beta \leq -\lambda) - [(a\lambda - \beta)_+ / (a-1)\lambda] I(\beta > \lambda) \}$$

时,式(7)被称为 SCAD,在贝叶斯的观点下惩罚函数中的 a 取值为 3.7^[7],此时该模型具有稀疏性、连续性和无偏性的特点.

由于 SCAD 是在最小二乘法的基础上建立起

来的,不具有稳健性,Wu 等^[8]在此基础上提出 Q-SCAD 方法:

$$Q_n(\beta; \lambda, \tau) = \sum_{i=1}^n \rho(Y_i - \sum_{l=1}^{L+4} [\hat{\beta}_l \sum_{m=1}^{N_i} C_i(t_m) P_l(t_m)]) + n \sum_{l=1}^{L+4} p_{\lambda_n}(|\hat{\beta}_l|) \quad (8)$$

式中, ρ 是一个凸函数,当分位数 $\tau \in (0, 1)$ 且记 $x^+ = \max(x, 0)$ 时,

$$\rho_{\tau}(x) = \tau x^+ + (1 - \tau)(-x)^+$$

为 τ 的分位数回归. 对于一个给定的调整系数 λ_n , 用分位数回归方法估计的系数值为

$$\hat{\beta} = \hat{\beta}(\lambda_n) = \underset{\tau \in (0, 1), \beta}{\operatorname{argmin}} Q_n(\beta; \lambda_n) \quad (9)$$

式中, $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_{L+4})$.

这样,如果选择使得预测残差的绝对值总和最小的分位数为最优分位数,则可同时确定最终的节点

$$\tau_{\text{best}} = \underset{\tau \in (0, 1)}{\operatorname{argmin}} \sum_{i=1}^n (Y_i - \sum_{l=1}^{L+4} [\hat{\beta}_l \sum_{m=1}^{N_i} C_i(t_m) P_l(t_m)]) \quad (10)$$

对于调整系数 λ_n 的选择,详见文献[9]. 通过对最优分位数的自动选取,克服了人工主观确定分位数和节点的不足,使得节点选择和参数估计更加接近实际,同时使得计算过程更加简洁和智能化.

在模型求解方面,Fan^[7]给出用局部二次逼近方法,但该算法存在一个很大的缺点:在采用逐步后退的方法来进行变量选择时,一旦在中间迭代的过程中一个变量被删除,则在最终的模型中该变量将不会出现. 为了克服这一缺陷,本文使用文献[10]提出的 DCA 算法来进行计算.

2 实证分析

本文选取上海证券交易所 2014-08-01~2014-08-29 交易的固定利率付息国债的收盘价做实证研究(共 150 只,21 组,限于篇幅不予一一列出),数据

来源为上海证券交易所公布的行情数据. 首先,将每组国债按到期日由小到大分为短期、中期和长期 3 类,在每一类中随机地选取 1/6 的国债,合并起来共 25 只作为检验样本,其余的 125 只(5/6)作为训练样本,用于参数估计. 以数据观测日(如 2014-08-01)作为 0 时刻,其后每隔 1 或 2 a 的时刻都选作初始节点. 由于中短期国债样本较多,因此在较为靠近 0 时刻的时段里我们选择节点间隔为 1 a,远离 0 时刻的时段,节点间隔设置为 2 a,这样能够保证每 2 个节点之间到期的国债个数比较平均.

接下来对数据进行实证,不妨采用预测残差的绝对值总和作为衡量标准,结果如图 1 所示.

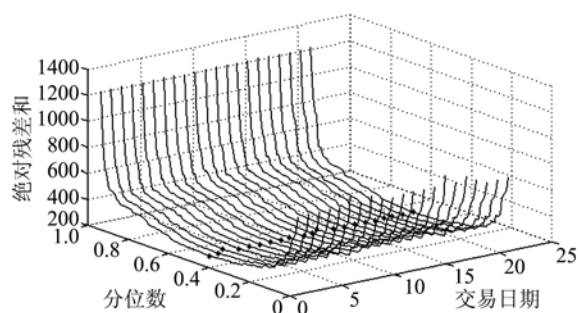


图 1 不同分位数下的绝对残差和

Fig. 1 The residuals' absolute sum with different quantile

我们选择使得预测残差的绝对值总和最小的分位数为最优分位数,由图 1 可知,21 组国债数据实证时的最优分位数平均值约为 0.46. 于是,可知在最优分位数取 0.46 时,模型的节点选择为(1, 5, 10, 14 a),见表 1(限于篇幅,仅列示以 2014-08-01 为观测日的第一组实证结果,以下同). 运用 Q-SCAD 方法获得的参数估计值见表 2.

为了比较模型的优劣,我们依据文献[7]依次选取 5 和 8 a 作为节点,分别采用最小一乘法 and 最小二乘法来回归,结果见表 3;同时,又采用 LAD-LASSO 方法进行节点选择和参数估计,结果见表 4.

表 1 Q-SCAD 节点选择 ($\tau = 0.46$)

Tab. 1 Q-SCAD variable selection ($\tau = 0.46$)

节点	2015-08-01	2016-08-01	2017-08-01	2018-08-01	2019-08-01	2020-08-01
对应参数估计	0.000 292 8	#	#	#	-0.000 367 7	#
节点	2022-08-01	2024-08-01	2026-08-16	2028-08-16	2030-08-16	
对应参数估计	#	0.000 144 5	#	-0.000 422 9	#	

【注】 # 表示值小于 1×10^{-6}

表 2 Q-SCAD 方法参数估计值
Tab. 2 Q-SCAD parameter estimation

参数	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
$\tau=0.46$	0.999 863 7	-0.028 815 1	-0.021 183 27	#	0.000 292 8	-0.000 367 7	0.000 144 5	-0.000 422 9

【注】 β_0, \dots, β_3 为 t^0, t^1, t^2 和 t^3 的参数, $\beta_4, \beta_5, \beta_6, \beta_7$ 为节点参数

表 3 LAD, LS 方法参数估计值
Tab. 3 LAD, LS parameter estimation

参数	β_0	β_1	β_2	β_3	β_4	β_5
LAD	0.994 601 9	-0.017 519 5	-0.006 003 0	0.000 562 5	-0.000 905 4	0.000 448 2
LS	0.999 535 4	-0.025 284 6	-0.004 130 1	0.000 411 3	-0.000 604 6	0.000 204 0

【注】 β_0, \dots, β_3 为 t^0, t^1, t^2 和 t^3 的参数, β_4, β_5 为节点参数

表 4 LAD-LASSO 方法参数估计值
Tab. 4 LAD-LASSO parameter estimation

参数	β_0	β_1	β_2	β_3	β_4	β_5	β_6
LAD-LASSO	0.998 615 0	-0.025 756 8	-0.001 882 5	-0.000 326 3	#	0.001 588 1	#

参数	β_7	β_8	β_9	β_{10}	β_{11}	β_{12}	β_{13}
LAD-LASSO	-0.003 170 2	0.004 546 9	-0.003 912 7	0.000 112 6	0.004 875 7	-0.005 862 0	0.001 363 2

【注】 β_0, \dots, β_3 为 t^0, t^1, t^2 和 t^3 的参数, $\beta_4, \beta_5, \dots, \beta_{12}, \beta_{13}$ 为节点参数

将用不同方法得到的系数 $\hat{\beta}$ 分别代入式(6), 可以估计出相应方法下的剩下 25 只国债的当天价格, 对其与真实价格相减后的残差取绝对值并求和, 可以得到不同方法下的残差的绝对值总和, 如图 2 所示. 我们还可以计算出即期利率和瞬时远期利率, 它们期限结构曲线如图 3 所示.

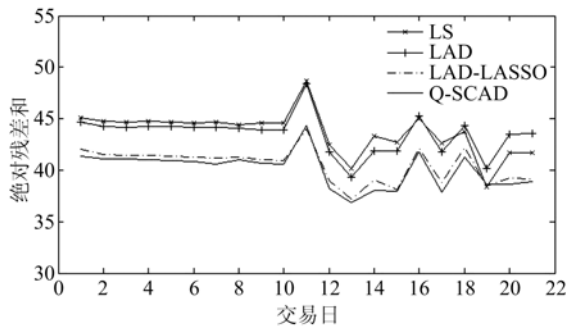


图 2 预测结果
Fig. 2 Forecast results

从图 2 可以看出, 以残差的绝对值总和为标准时, 在总体上看 Q-SCAD($\tau=0.46$)方法得到的国债值好于其他方法; 固定节点为 5 和 8 a 时, 比较 LAD 方法与 LS 方法预测的结果, 我们可以看出在异常点存在的情况下, 最小一乘法要优于最小二乘法; Q-SCAD($\tau=0.46$)方法比 LAD-LASSO 方法得到的结果要好, 说明惩罚分位数回归方法能更准确

地选择节点, 从而使得模型的预测效果更好.

从图 3(a), (d) 可以看出, 在期限较短(8 a 以内)时, Q-SCAD($\tau=0.46$)方法和 LAD-LASSO 方法拟合的利率期限结构都较好, 利率与期限成正向关系, 且远期利率与即期利率曲线走势平稳而光滑; 但随着期限的增加(8 a 以后), LAD-LASSO 方法的远期利率波动性陡然增大. 比较图 3(a)和图 3(b), (c), (d), 发现采用 LAD, LS 和 LAD-LASSO 方法拟合的利率期限结构波动性都较大, 曲线不光滑, 拟合效果不佳, 而用 Q-SCAD($\tau=0.46$)方法构建利率期限结构更符合实际情况, 期限结构的波动性比较小, 曲线形状比较光滑, 拟合效果较好, 更为合理.

3 结论

通过实证发现, 采用 Q-SCAD 方法构建样条函数来拟合我国国债的利率期限结构, 具有如下优点: 首先, 该方法可以自动选择最优分位数和同时实现参数估计与节点选择, 避免了人为事先设定的影响; 其次, 该模型具有一定的稳健性, 降低了异常值对参数估计的影响, 避免了过度拟合, 同时使利率曲线具有优良的平滑性; 最后, 在中短期内, Q-SCAD 方法构建的远期利率期限结构曲线形状与即期利率期限结构曲线形状的时间变化趋势几乎是一样的, 并且

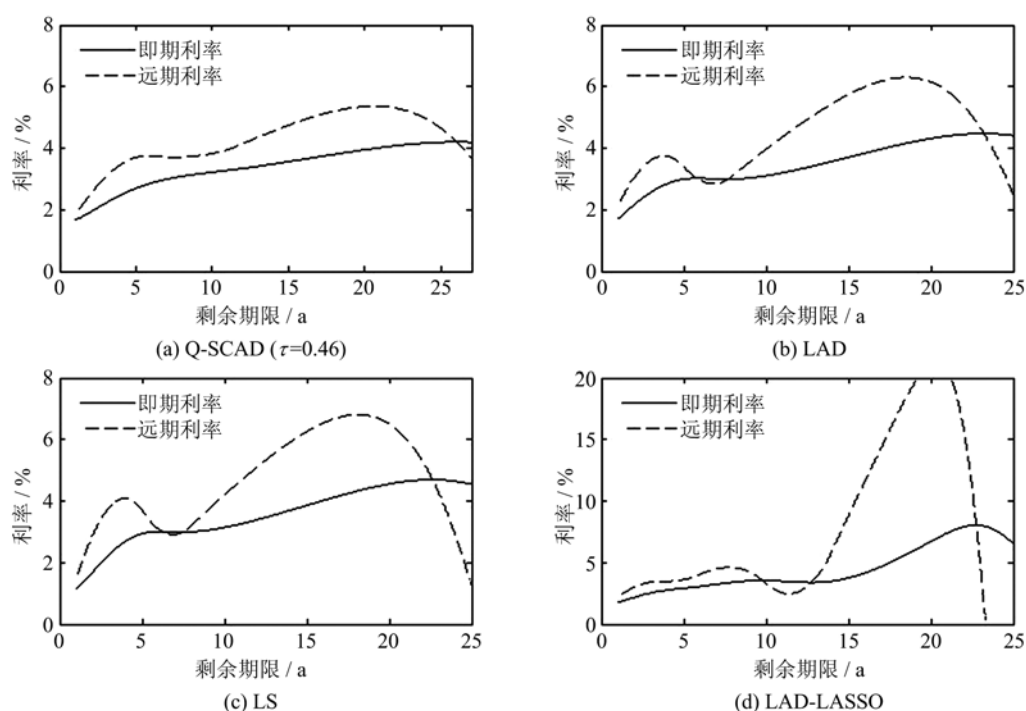


图 3 各种方法的即期利率与远期利率

Fig. 3 Spot rate and forward rate of different methods

比较光滑,更为贴近实际.随着长期国债和超长期国债发行数量的逐渐增多,对此类国债的利率期限结构的更加深入的研究是一个重要的课题,这也将是我们下一步的工作.

参考文献(References)

- [1] McCulloch J H. Measuring the term structure of interest rates[J]. The Journal of Business, 1971, 44(1): 19-31.
- [2] Xiao Nan, Cheng Xijun. Fitting term structure of interest rate of domestic bonds market by robust spline[J]. Systems Engineering, 2007, 25(6): 35-40.
萧楠,程希骏.抗差样条模型对我国国债利率期限结构的建模与实证[J].系统工程,2007,25(6):35-40.
- [3] Sun Zengxian, Cheng Xijun, Ma Lijun, et al. Fitting term structure of interest rate with splines based on quantile regression[J]. Systems Engineering, 2008, 26(11): 6-10.
孙增献,程希骏,马利军,等.基于分位数回归的样条函数法拟合国债利率期限结构[J].系统工程,2008,26(11): 6-10.
- [4] Li Yiyi, Pan Wanbin, Miao Baiqi. Knot selection of estimating the term structure with cubic spline function[J]. Systems Engineering Theory & Practice, 2009, 29(4): 28-33.
李熠熠,潘婉彬,缪柏其.基于三次样条的利率期限结构估计中的节点选择[J].系统工程理论与实践,2009,29(4): 28-33.
- [5] Li Yiyi, Pan Wanbin, Miao Baiqi. Knot selection of estimating the term structure with cubic spline function based on LAD-Lasso criterion [J]. Journal of University of Science and Technology of China, 2010, 40(6): 551-556.
李熠熠,潘婉彬,缪柏其.基于LAD-Lasso方法的利率期限结构拟合中的节点选择[J].中国科学技术大学学报,2010,40(6):551-556.
- [6] Fan Jianqing, Yao Qiwei. Nonlinear Time Series: Nonparametric and Parametric Methods [M]. New York: Springer, 2003.
- [7] Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties [J]. Journal of the American Statistical Association, 2001, 96(456): 1 348-1 360.
- [8] Wu Y, Liu Y. Variable selection in quantile regression [J]. Statistica Sinica, 2009, 19(2): 801-817.
- [9] Wang H, Li R, Tsai C L. Tuning parameter selectors for the smoothly clipped absolute deviation method[J]. Biometrika, 2007, 94(3): 553-568.
- [10] Le Thi Hoai A, Tao P D. Solving a class of linearly constrained indefinite quadratic problems by DC algorithms[J]. Journal of Global Optimization, 1997, 11(3): 253-285.