

西安交通大学

毕业设计（论文）

题 目 多折点回归模型的参数估计

数学与统计 学院 统计学 专业 统计 81 班

学生姓名 于越

学 号 2183210516

指导教师 付利亚

设计所在单位 西安交通大学

2022 年 6 月

西安交通大学

系(专业) 0621 统计学

系(专业) 主任 惠永昌

批准日期 2021 年 11 月 05 日

毕业设计(论文)任务书

数学与统计学院 学院 0621 统计学 专业 统计 81 班 学生 于越

毕业设计(论文)课题 多折点回归模型的参数估计研究

毕业设计(论文)工作自 2021 年 9 月 11 日起至 2022 年 6 月 22 日止

毕业设计(论文)进行地点: 校内

课题的背景、意义及培养目标

在实际数据建模过程中,因变量和协变量之间的关系在不同时间可能并不相同,因此产生了折点回归模型。对于折点中的参数和折点数进行估计在实际问题中有重要的应用价值。通过毕业设计,掌握折点回归模型中参数估计的方法,能够利用该模型对实际数据进行分析;熟悉科技论文的撰写要求;提高数据分析能力和编程能力,增强统计素养。

设计(论文)的原始数据与资料

见主要参考文献的第一篇文献。

课题的主要任务

1、阅读回归分析、折点模型的相关英文文献。

2、对文献中的方法进行复现。

3、通过数值模拟对不同的方法进行比较。

4、翻译一篇英文文献。

5、按照毕设要求撰写毕业论文。

课题的基本要求(工程设计类题应有技术经济分析要求)

无

完成任务后提交的书面材料要求(图纸规格、数量，论文字数，外文翻译字数等)

1、翻译一篇毕业论文的参考文献或与毕业论文密切相关的资料。译文字数 2000~3000 字，要求译文与原文内容相符。

2、根据毕业设计的撰写要求，完成一篇不少于 15000 字的毕业论文。

主要参考文献

1、Zhong.W., Wan C., Zhang W. (2021), Estimation and Inference for Multi-Kink Quantile Regression, Journal of Business Economic Statistics, DOI:10.1080/07350015. 2021.1901720.

2、Hansen B.E. (2017), Regression Kink With an Unknown Threshold, Journal of Business Economic Statistics, 35, 228 - 240.

3、Li C., Wei Y., Chappell R. and He X. (2011), Bent Line Quantile Regression With Application to an Allometric Study of Land Mammals Speed and Mass, Biometrics, 67, 242 - 249.

指导教师： 付利亚

接受设计(论文)任务日期： 2021 年 11 月 05 日

（注：由指导教师填写）

学生签名： _____

西 安 交 通 大 学

毕业设计(论文)考核评议书

数学与统计学院 学院 0621 统计学 专业 统计 81 班

指导教师对学生于越所完成的课题为多折点回归模型的参数估计研究的毕业设计(论文)进行的情况,完成的质量及评分的意见:该生从选题初期到最后的论文撰写过程中,每周都和导师主动沟通交流,遇到问题及时解决。该生学习态度端正,敏而好学,其毕设研究了单折点模型和多折点模型的参数和折点数估计及其检验,并根据模型和目标函数特征,提出几种改进方法。进一步地,为了评估改进方法和原有方法的性能,该生进行了大量的数值模拟试验,从同方差和异方差,正态分布和重尾分布,以及不同的折点数和不同的分位数进行比较,根据数值结果,得到有价值的建议和结论。该论文结构完整,内容充实,书写规范,是一篇优秀的毕设论文。

指导教师建议成绩: A+

指导教师: 付利亚

2022 年 06 月 08 日

毕业设计(论文)评审意见书

评审意见:该论文中系统性地对三个重要的折点回归模型进行了参数估计、折点效应检验、算法以及相关渐近性质的了解,在此基础上,针对模型本身折点位置处不光滑与分位数回归损失函数并非处处可导的特性,给出了对于多折点分位数回归模型的算法改进。论文研究的是当前领域的前沿问题,作者对相关问题进行了全面的了解并进行了改进。论文撰写逻辑严谨,表述清晰,展现了良好的思维能力与表述能力,符合优秀论文标准。

评阅教师建议成绩: A+

评阅人: 程晓青 职称: 副教授

2022 年 06 月 13 日

毕业设计(论文)答辩结果

数学与统计学院 学院

0621 统计学 专业

毕业设计(论文)答辩组对学生于越所完成的课题为多折点回归模型的参数估计研究的毕业设计(论文)经过答辩,其意见为该生论文研究了多折点分位数回归模型的参数估计和折点数估计,并分别对回归方程和目标函数提出改进方法。进一步地,通过大量的数值模拟对提出的五种改进方法进行比较评估。反映出该生较好地掌握了多折点分位数回归的基本思想和相关统计方法。论文工作量饱满,结构清晰合理,写作规范,图标清楚,达到本科毕业论文的要求。在答辩过程中,该生表述清楚,回答问题思路清晰,经答辩组成员无记名投票,一致通过于越的本科毕业论文答辩。

并确定成绩为 A+

毕业设计(论文)答辩组负责人

惠永昌

答辩组成员

付利亚

康尧

孙凯

曹相湧

2022 年 06 月 16 日

摘 要

变点问题在当今社会的各个领域都有所涉及。折点回归模型是研究变点问题的重要模型，可以有效刻画数据集变量之间的连续多段的线性变化。本文依据该模型的发展历程，系统性地对该模型在不断演变过程中的三个重要模型：单折点均值回归模型、单折点分位数回归模型与多折点分位数回归模型，进行了参数估计、折点效应检验、算法以及相关渐近性质的介绍。与此同时，主要针对模型本身折点位置处不光滑与分位数回归损失函数并非处处可导的特性，结合前人研究的方法，给出了对于多折点分位数回归模型的一些算法改进。这些改进算法主要应用梯度下降算法的框架，得到模型参数的有效估计。大量的数值模拟结果表明，本文中介绍的多折点分位数回归模型算法以及其相关的改进算法可以很好地完成模型参数估计的任务。具体的试验表现为在同方差、异方差两种情形以及随机误差项服从不同分布的情况下，对于不同的分位数水平，这些算法都可以很好地控制参数估计的偏以及均方误差。最后，本文还将上述方法用到实际数据分析，论证了这些算法的实际应用意义，同时说明这些算法对于不同的数据集各有优劣，因而在实际问题中需要依据具体问题选择合适的算法进行数据分析。

关 键 词：折点回归模型；分位数回归；参数估计；光滑化

ABSTRACT

The change-point problems have involved in all varieties of fields in contemporary society. Kink regression model is important for study of change-point problems, which can effectively depict continuous and piecewise linear variations between variables of dataset. According to the development history of this kind of model, this paper systematically introduces several aspects of three important evolving models: Single-kink mean regression model, Bent-line quantile regression model and Multi-kink quantile regression model, including parameter estimation, tests of kink effect, algorithms and theories of asymptotic property. Meanwhile, mainly aiming to characters of the model such as roughness on the location of kink and nondifferentiability of quantile regression loss function, combining with some previous research methods, this paper gives some algorithm improvements on multi-kink quantile regression model. These algorithm improvements mainly implement the framework of iterative gradient descent algorithm, which can achieve effective estimations of model parameters. A large number of numerical simulations show that the multi-kink quantile regression model and our proposed algorithm improvements can commendably fulfil the task of estimation of parameters. The specific experimental performance is that for both homoscedasticity case and heteroscedasticity case, for the conditions that random error terms subject to different distributions, for different quantile levels, these algorithms can control bias and MSE of parameters estimation in a good condition. Finally, this paper also applies the above methods to the actual data analysis, demonstrating the practical application significance of these algorithms. In the meanwhile, we explain that these algorithms have their own disadvantages and defects for different datasets, hence it's vital to select appropriate algorithms to analyze data based on the concrete issues.

Key Words: Kink Regression Model; Quantile Regression; Parameter Estimation; Smoothing

目 录

1	绪论.....	1
1.1	研究背景.....	1
1.2	前人工作与研究现状.....	2
1.3	本文文章结构.....	4
2	单折点回归模型.....	6
2.1	单折点均值回归模型.....	6
2.1.1	模型形式与假设.....	6
2.1.2	损失函数与估计方法.....	6
2.1.3	折点效应存在性检验.....	7
2.1.4	模型统计推断.....	9
2.2	单折点分位数回归模型.....	12
2.2.1	模型形式与假设.....	13
2.2.2	损失函数与估计方法.....	13
2.2.3	折点效应存在性检验.....	14
2.2.4	模型统计推断.....	15
3	多折点分位数回归模型.....	17
3.1	模型形式与假设.....	17
3.2	参数估计方法.....	18
3.2.1	损失函数.....	18
3.2.2	折点位置估计.....	18
3.2.3	折点数估计.....	20
3.2.4	方差估计的极限分布.....	22
3.3	模型折点效应检验.....	24
3.4	折点位置置信区间构造.....	26
4	算法改进.....	29
4.1	算法改进 1: 线性部分核光滑化.....	29
4.1.1	改进原理.....	29
4.1.2	算法流程.....	32
4.2	算法改进 2: Bent-Cable 平滑化方法.....	32
4.2.1	改进原理.....	32
4.2.2	算法流程.....	35

4.3 算法改进 3: 分位数回归损失函数的替换.....	35
4.3.1 改进原理.....	35
4.3.2 算法流程.....	37
4.4 算法改进 4: 基于损失函数替换的 Boosting 方法.....	37
4.4.1 改进原理.....	37
4.4.2 算法流程.....	38
4.5 算法改进 5: 诱导光滑化方法.....	39
4.5.1 改进原理.....	39
4.5.2 算法流程.....	41
5 数值模拟试验.....	42
5.1 单折点情况下的数值模拟.....	42
5.1.1 试验设置.....	42
5.1.2 数值模拟结果分析.....	43
5.2 多折点情况下的数值模拟.....	46
5.2.1 试验设置.....	46
5.2.2 数值模拟结果分析.....	47
6 实际数据分析.....	70
6.1 单折点回归模型.....	70
6.1.1 数据简介.....	70
6.1.2 结果展示与分析.....	70
6.2 多折点分位数回归模型.....	73
6.2.1 数据简介.....	73
6.2.2 结果展示与分析.....	73
7 结论与展望.....	79
致 谢.....	80
参考文献.....	81
附录 A 正文未展示的模拟试验结果展示.....	84
附录 B 外文文献原文.....	90
附录 C 外文文献译文.....	103

1 绪论

1.1 研究背景

线性回归模型作为统计学中的一大基础模型，已经被大众所熟知。该模型思想简单，容易实现，对于小数据量、简单的关系可以做出有效的描述；且该模型作为许多非线性模型的基础，十分容易理解，且具有很好的可解释性，有利于数据分析。但是在实际生活中，很多事物之间的关系并非是一成不变的。很多时候会观测到这样的现象：响应变量与解释变量在某个临界位置前后呈现出了截然不同的回归关系。这个临界位置被称为“变点”。

变点这一概念最早由 Page^[1] 在 1954 年发表于统计学期刊 *Biometrika* 上的第一篇关于变点分析的文章中提出，用于研究产品质量检测问题中的连续抽样检验，通过变点位置模拟产品检测的质量警戒线，从而剔除产品中的次品。在此之后，变点问题在统计学领域受到越来越多学者的关注，目前已经演变成统计学中的一个重要研究方向。

根据模型回归函数在变点处是否连续，变点模型被划分为折线模型和结构变点模型。在本文中主要研究的是折线模型。在折线模型中，变点位置又常常被称为“折点”。在折线模型中，解释变量与响应变量之间关系的变化主要体现在折点位置前后的线性模型的斜率上。以 Garland^[2] 文中对数尺度下成年哺乳动物的最大奔跑速度与体重之间的关系为例，从图1-1中可以很明显看出，对于质量不是特别大的动物而言，动物的最大奔跑速度会随着动物质量的增加而增加；但是当质量增加到一定程度时，动物的最大奔跑速度反而随着动物质量的增加而减小。这意味着在对数尺度下，动物的最大奔跑速度和质量之间可能存在线性关系，但是二者之间的回归关系在“折点”位置两侧存在着不同的斜率。事实上这与我们的生物学直觉是一致的。例如猎豹的质量要大于野兔，猎豹的最大奔跑速度也是大于野兔的；但是当动物质量超过一个临界时，比如大象的质量远大于猎豹，但是大象的最大奔跑速度显然不及猎豹。因此如何证实该数据集确实存在“折点”效应，如何估计出折点的位置以及模型的其他参数，这些都是值得考虑的问题。

与此同时，在实际问题中还可能碰到响应变量与解释变量之间的关系经历了不止一次变化的情形。例如，1880 年到 2020 年的年度全球气温极值变化数据集，如图1-2所示。该数据集描述了 1880 年到 2020 年 141 年间每年的气温极值随时间推移的变化。该数据集来自于美国国家航空航天管理局的网站 (www.giss.nasa.gov)。从图中可以明显看出，尽管这 141 年间的全球气温总体呈上升趋势，但是这一趋势并非一成不变。在约 1945 年之前，气温的异常值上升呈现一个较为缓慢的趋势，而在 1945 年到 1970 年之间，气温的异常值在一个较为稳定的数值附近波动。而在 1970 年后，这一数值又开始随时间推移呈上升趋势，并且此时的上升幅度远大于 1945 年之前。这一趋势最早由 Hansen 与 Lebedeff^[3] 发现，并且引发了后人诸多的研究。显然，在这一数据集当中，折

点的数量不止一个。变量之间的关系发生了不止一次变化。因此我们渴望引入一种可以表征多个折点的模型来刻画这种数据关系的变化，并估计出多个折点的位置以及模型的其他系数，从而挖掘出更多的数据信息。

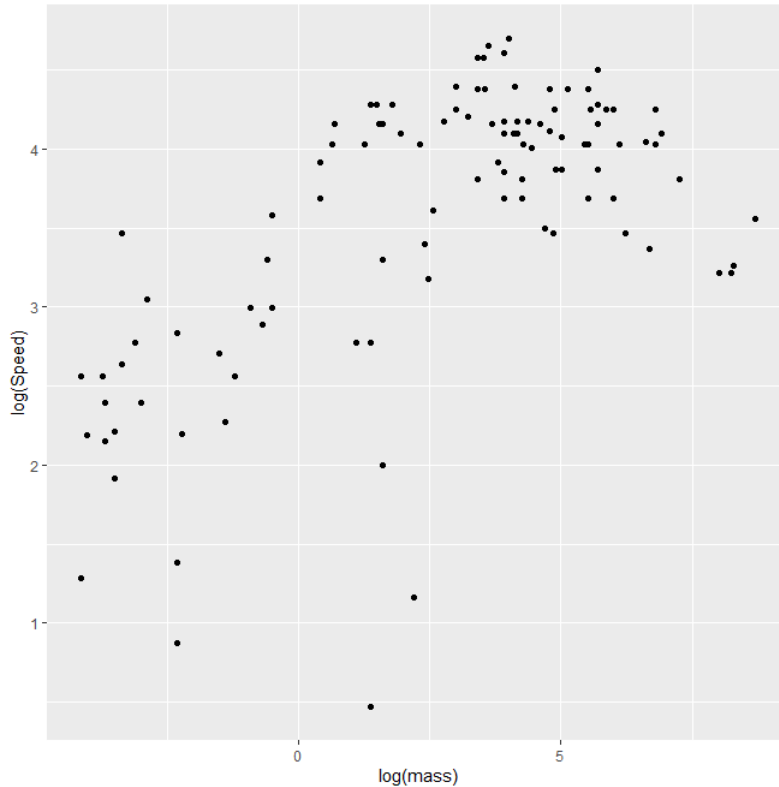


图 1-1 动物质量与最大奔跑速度的对数形式散点图

对于响应变量与解释变量之间经历了不止一次变化的数据集，也有前人采用平滑样条的方法，构建非参数回归模型进行分析，例如 Koenker 与 Schorfheide^[4] 对于全球温度极值的分析。但是，非参数回归模型存在两个问题：(a) 难以刻画出具体变量之间关系变化的位置，发掘出具体变化的相关信息，这给相关的决策留下了一些难题；(b) 非参数回归模型过于灵活，以至于模型的可解释性较差。由此作为一般线性回归模型与非参数回归模型的折中，折点式的回归模型应运而生，目前正处于不断发展的状态。

1.2 前人工作与研究现状

本节中主要介绍与折点回归模型相关的研究历史与现状。折点回归模型的历史最早可以追溯到 Quandt^[5, 6] 提出的连续分段线性模型与 Sprent^[7] 提出的两阶段线性回归模型。该模型在研究响应变量和解释变量之间存在的连续折线关系时被提出。顾名思义，这种折线关系在折点位置前后都是线性的，具有完全不同的斜率，且两段线性模型在折点位置处相交。在此模型基础上，Hinkley^[8, 9] 与 Feder^[10] 以及 Chappell^[11] 研究了这种分段式折线模型的统计推断问题。而他们的研究方法大都是基于最小二乘方法，旨在使用传统方法解决这一新兴模型的问题。后来，基于这种模型，Tong^[12, 13] 引入了

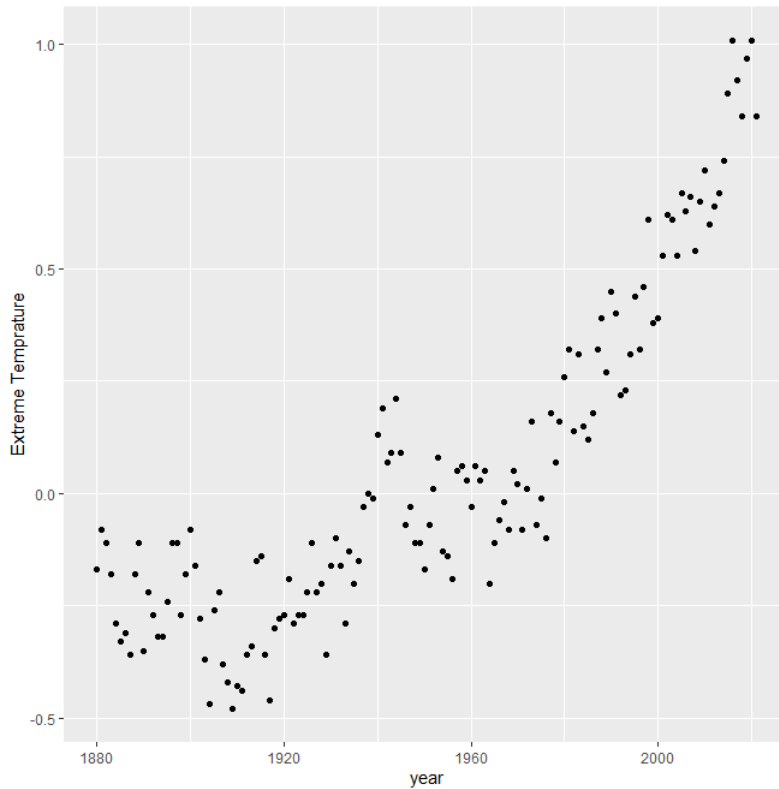


图 1-2 全球气温极值变化数据散点图

门限效应模型。这种模型可以被用于刻画非线性关系。Chan 与 Tsay^[14]在此基础上提出了连续门限模型。这一模型和分段式的线性回归模型有着相同的形式。

但是，运用这一模型的前提是数据集中存在着明显的回归关系的急剧突变。但是在实际操作过程中，人们难以判断数据集中是否存在着这一明显的突变。因此，研究的重心被渐渐地转移到了这一问题上来。这一问题被总结为“检验门限效应是否存在的问题”。关于门限效应检验的问题，开始于 Chan 与 Tong^[13]对于门限自回归效应的检验；接着 Hansen^[15]与 Lee 等人^[16]也各自对这一问题做出了自己的贡献。值得一提的是，Chiu 等人^[17,18]从缓和模型中存在的急剧突变的角度入手，受启发于 Tishler 与 Zang^[19]提出的“直线——二次曲线——直线”结构，创造性地提出了 BentCable 均值回归模型。他们利用这一结构，缓和模型的折点位置处的急剧突变，由此较为灵活地刻画了折点附近的模型变化。后续 Das 等人^[20]对于这一方法进行了相关的改进。该方法在本文中被应用到算法改进部分。

在 Chan 与 Tsay^[14]的研究基础上，Hansen^[21]对相关研究理论进行了拓展。Hansen^[21]系统性地给出了单折点均值回归模型的参数估计、方差估计、折点效应检验以及统计推断的方法，其中还包含了参数置信区间构造甚至是整个单折点均值回归模型的置信区间构造方法，是折点均值回归这个方向上目前的集大成者。但是，均值回归模型不能动态刻画响应变量和解释变量之间的关系。于是，Li 等人^[22]基于 Koenker 与 Bassett^[23]提出的分位数回归模型，以 1.1 节提到的动物质量与最大奔跑速度的数据集作为实证分析，创造性地提出了单折点分位数回归模型。这一模型可以对具有折点效应的数据集

的处于较大及较小分位数位置的数据进行系统的研究，很好地克服了单折点均值回归模型存在的缺陷。

但是正如之前所提到的，在实际数据中变量之间急剧变化的关系很有可能不止一处。此时，单折点分位数回归模型就不适用。于是，Zhong 等人^[24]在 Li 等人^[22]的研究基础上系统性地提出了多折点分位数回归模型。Zhong 等人^[24]的研究很好地利用了 Muggeo^[25]的折点位置估计的线性化技巧，做到了折点位置的准确估计；与此同时，他们对于折点数估计、参数估计、折点效应的检验以及置信区间的构造等问题也给出了自己的创新性方法。他们结合了 Bradley^[26]的自助法思想，给出了用于该问题参数估计的自助式迭代分段分位数回归算法（Bootstrap Restarting Iterative Segmented Quantile, BRISQ 算法）。该模型很好地做到了线性分位数回归模型与非参数分位数回归模型的折中，是值得推荐的回归分析方法，这在后文中会重点介绍。

除此之外，对于折点回归模型中折点位置处不光滑不可导的性质，Zhou 与 Zhang^[27]对于单折点回归模型给出了相关的光滑化方法。这一方法也用到了本文中对于多折点分位数回归方法的算法改进当中。对于这一问题，此前提到的 BentCable 回归模型也是一个良好的解决方法。而 Zhang 和 Li^[28]提出的折线期望分位数模型可以解决这一问题，该模型做到了损失函数处处可导，从而使参数估计更为简洁。

综上所述，统计学中变点问题这一领域的发展正变得愈发成熟，同时也值得进行更深入的探索。

1.3 本文文章结构

本文主要对折点回归模型进行系统研究。本文大致结构如下所示：

第二章，主要介绍折点回归模型中单折点回归模型的情况。首先从单折点均值回归模型出发，先大致介绍 Hansen^[21]的单折点均值回归模型的算法与基本原理。这一部分主要涵盖了单折点均值回归模型的参数估计、方差估计、折点效应检验、参数置信区间构造的原理与算法。接着从单折点均值回归模型过渡到单折点分位数回归模型，介绍了分位数回归情形下的原理与相关算法。

第三章，将单折点回归模型发展到多折点回归模型，对 Zhong 等人^[24]研究的多折点分位数回归模型算法与理论进行基本介绍。这一部分主要涉及参数估计、统计推断、折点效应检验这三个方面，重点在于折点位置估计中运用到的线性化技巧，BRISQ 算法以及在确定了数据集的折点效应存在的情况下折点数的估计。

第四章，本文运用前人提出的一些方法，对 Zhong 等人^[24]的多折点分位数回归模型的参数估计算法进行了一系列的改进。这一部分主要介绍这些改进的原理及算法。改进的思想主要集中于“对示性函数的光滑化”方面。

第五章，本文对第 2-4 章的算法进行了大量的模拟数值试验，尤其是对于第三章和第四章中涉及的多折点分位数回归模型参数估计算法及其改进算法，用于评估这些算法的性能。在这一部分，本文主要设置了同方差与异方差两种情形下的数据模型，基

于随机误差项的不同分布（标准正态分布、t 分布、标准 Laplace 分布）和不同的分位数水平进行试验。大量的数值模拟结果表明，所有的算法都可以胜任参数估计的任务，且不同的算法表现各有优劣。

第六章，本文运用两个实际数据集——我国不同城市的人均 GDP 与第二产业比重数据集和冈比亚女性年龄与肱三头肌皮褶厚度关系的数据集，分别针对单折点回归模型与多折点分位数回归模型进行了实证分析，进一步说明这些模型的实际应用意义。对于多折点分位数回归模型的实证分析，本人将此前 Zhong 等人^[24]的估计算法与改进算法都进行了应用，并且以此比较了对于这一数据集而言的算法优劣性。

本文的创新之处主要体现在第四章与第五章，即对于 Zhong 等人^[24]的多折点分位数回归模型的参数估计算法的多种改进算法的介绍与模拟试验。最后，本人对所有的方法进行了总结，并给出了这一课题的未来展望。

2 单折点回归模型

本章首先介绍单折点回归模型。本文主要从模型形式、模型假设、损失函数、估计方法等方面入手，对单折点均值回归模型^[21]与单折点分位数回归模型^[22]进行介绍。

2.1 单折点均值回归模型

2.1.1 模型形式与假设

假定样本量为 n ，响应变量为 y_i ，解释变量为 x_i 与 \mathbf{z}_i ，随机误差项为 e_i 。单折点均值回归模型具有以下形式：

$$y_i = \beta_1(x_i - \delta)_- + \beta_2(x_i - \delta)_+ + \beta_3^T \mathbf{z}_i + e_i, \quad i = 1, 2, \dots, n, \quad (2-1)$$

式中 δ 为折点所在位置， $(u)_- = \min\{u, 0\}$ ， $(u)_+ = \max\{u, 0\}$ 。而 \mathbf{z}_i 为包含了常数项的 p 维向量形式。从模型形式中可以明显发现，解释变量和响应变量在折点两侧存在不同的线性关系，这是该模型的一大特点。

该模型的回归函数关于解释变量 x_i 和 \mathbf{z}_i 是连续的，但是在折点位置 $x_i = \delta$ 处不连续。在此处指定了模型除去折点的回归部分是线性形式，也可以使用其他的形式进行替代（例如多项式）。在传统的单折点回归模型中，折点的位置是已知的，而在此处需要将其认为是一个未知的待估参数。事实上，该模型只有在折点位于解释变量 x_i 的支撑集的内部时才有意义，因此不妨假定 $\delta \in \Delta$ ，其中 Δ 是一个紧集且严格嵌入在 x_i 的支撑集内部。

值得一提的是，折点回归模型隶属于不连续门限模型。因此，可以考虑在门限模型类之内检验连续性的假定^[21]。但这是一个很困难的问题，因此在这里不再讨论而是简单地假设回归函数是连续的。

2.1.2 损失函数与估计方法

众所周知，若模型反映的是与条件均值相关的问题，我们通常使用最小二乘法估计未知参数。在此处也不例外，应用带有未知折点位置的最小二乘法解决估计问题。本文将解释变量 x_i ， \mathbf{z}_i 以及折点位置 δ 整合成如下向量的形式：

$$\mathbf{x}_i(\delta) = [(x_i - \delta)_-, (x_i - \delta)_+, \mathbf{z}_i]^T,$$

同时令 $\beta = (\beta_1, \beta_2, \beta_3^T)^T$ 。模型(2-1)式可以改写为以下形式：

$$y_i = \beta^T \mathbf{x}_i(\delta) + e_i. \quad (2-2)$$

该模型的最小二乘损失函数为

$$L(\boldsymbol{\beta}, \delta) = \frac{1}{n} \sum_{i=1}^n (y_i - \boldsymbol{\beta}^T \mathbf{x}_i(\delta))^2.$$

对应的最小二乘估计为

$$(\hat{\boldsymbol{\beta}}, \hat{\delta}) = \underset{\boldsymbol{\beta} \in \mathbb{R}^{2+p}, \delta \in \Delta}{\operatorname{argmin}} L(\boldsymbol{\beta}, \delta). \quad (2-3)$$

若是对于一般的最小二乘估计，只需求解(2-3)式即可。但是对单折点回归模型的最小二乘，折点位置 δ 作为未知待估计参数，增加了这一问题的复杂性。(2-3)式相对于待估参数 δ 并非是一个凸函数。因此在估计折点位置 δ 时，需要运用 Lerman^[29] 提出的网格搜索算法。

Hansen^[21] 将原有的待估计参数进行了整合，(2-3)式写成如下形式：

$$\hat{\delta} = \underset{\delta \in \Delta}{\operatorname{argmin}} \min_{\boldsymbol{\beta} \in \mathbb{R}^{2+p}} L(\boldsymbol{\beta}, \delta) = \underset{\delta \in \Delta}{\operatorname{argmin}} L^*(\delta), \quad (2-4)$$

整合参数后的损失函数 $L^*(\delta)$ 如下所示：

$$L^*(\delta) = L(\hat{\boldsymbol{\beta}}(\delta), \delta) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\boldsymbol{\beta}}(\delta)^T \mathbf{x}_i(\delta))^2, \quad (2-5)$$

在此处假定 $\hat{\boldsymbol{\beta}}(\delta)$ 是给定折点 δ 情况下 $\boldsymbol{\beta}$ 的最小二乘估计系数。因此，我们可以先框定一个折点 δ 所存在的大致范围，对该范围做一个较为精细的格点划分，再将折点位置依次固定为该范围中的格点划分代入(2-5)式计算其损失函数的值。取(2-5)式最小位置处的 $\hat{\delta}$ 为折点位置 δ 的估计。再求解出 $\hat{\delta}$ 位置处的其他系数的估计，即为最终估计结果。

而对于随机误差项的方差，可用下式估计：

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \hat{e}_i^2 = L^*(\hat{\delta}). \quad (2-6)$$

显然，求得 $\hat{\delta}$ 后代入整合后的损失函数即可。

2.1.3 折点效应存在性检验

所有的模型都具有一定的使用范围。单折点回归模型适用于确实存在折点的数据集，对于(2-1)式 $\beta_1 = \beta_2$ 的情形并不适用。这就需要对于数据集是否存在折点效应进行检验：

$$\begin{cases} H_0 : \beta_1 = \beta_2 \\ H_1 : \beta_1 \neq \beta_2 \end{cases} \quad (2-7)$$

对于(2-1)中 $\beta_1 = \beta_2$ 的情形即为一般的均值回归，设 $\tilde{\beta}_1$ 和 $\tilde{\beta}_3$ 是 β_1 和 β_3 的最小二乘估计，则残差为

$$\tilde{e}_i = y_i - \tilde{\beta}_1 - \tilde{\beta}_3 z_i, \quad (2-8)$$

残项的方差估计为

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \tilde{e}_i^2.$$

受到一般线性回归中检验问题的启发，对于(2-7)的检验也可以通过一个 F 统计量来完成。F 统计量形式为

$$F_n = \frac{n(\tilde{\sigma}^2 - \hat{\sigma}^2)}{\hat{\sigma}^2}.$$

一般而言，对于大于临界值的 F_n ，我们需要拒绝原假设，进而选择模型(2-1)式进行回归。要进行检验，要求出统计量的极限分布。**Hansen**^[21] 对于该统计量的极限分布做了较为详细的理论解释与证明。在推导统计量的极限分布之前，需要满足以下正则性假设2.1。

假设 2.1 ^[21] 当 $k > 1$ 时，

(C1). 对于某些 $L > k/(k-1)$ ，带有混合系数 $\eta(a) = O(a^{-L})$ 的变量 (y_i, z_i, x_i) 是严平稳的，遍历性的且绝对正则的。

(C2). $E|y_i|^{4k} < \infty$, $E|x_i|^{4k} < \infty$ 且 $E||z_i||^{4k} < \infty$ 。

(C3). $\inf_{\delta \in \Delta} \det[Q(\delta)] > 0$ ，其中 $Q(\delta) = E[x_i(\delta)x_i(\delta)^T]$ 。

(C4). x_i 有密度函数 $f(x)$ 满足 $f(x) \leq \bar{f} < \infty$ 。

(C5). $\delta \in \Delta$ 其中 Δ 是紧集。□

在假设2.1满足的前提下，可以利用类似 Wald 统计量构造的方法，得到以下定理2.1。

定理 2.1 ^[21] 满足假定2.1且随机误差项 e_1, \dots, e_n 是一个鞅差序列。在原假设 $H_0 : \beta_1 = \beta_2$ 下

$$F_n \xrightarrow{d} \sup_{\delta \in \Delta} G(\delta)^T Q(\delta)^{-1} G(\delta) / \sigma^2,$$

其中 $G(\delta)$ 是一个均值为 0，且协方差满足

$$E[G(\delta_1)G(\delta_2)] = E[x_i(\delta_1)x_i(\delta_2)^T e_i^2].$$

的 Gauss 过程。□

如同定理2.1所示，可以构造统计量 F_n 的极限分布进而完成检验。但是受制于样本量等一系列因素，检验效果会受到一定影响。由此可以通过 **Hansen**^[21] 中的自助法完成这一检验。具体步骤如算法2-1所示。利用该算法即可完成对于折点均值回归模型的折

点效应的检验。

- 1: 从正态分布 $\mathbf{N}(0, 1)$ 生成 n 个独立同分布的变量 $u_i (i = 1, 2, \dots, n)$ 。
- 2: 设定 $y_i^* = \tilde{e}_i u_i$ 其中 \tilde{e}_i 是 $y_i = \tilde{\beta}_1 + \tilde{\beta}_3 z_i + \tilde{e}_i$ 的残差。
- 3: 使用观测量 (y_i^*, z_i, x_i) 估计上述线性模型与折点回归模型(2-1)并计算误差方差估计 $\tilde{\sigma}^{*2}$ 与 $\hat{\sigma}^{*2}$ 与 F 统计量 $F_n = \frac{n(\tilde{\sigma}^{*2} - \hat{\sigma}^{*2})}{\hat{\sigma}^{*2}}$ 。
- 4: 重复这一过程 B 次, 故可以获得模拟的 F 统计量样本序列 $F_n(1), \dots, F_n(B)$ 。
- 5: 计算 p 值, 即模拟的 F 统计量大于真实值的百分比: $p_n = \frac{1}{B} \sum_{b=1}^B I(F_n(b) \geq F_n)$ 。
- 6: 如果需要的话, 计算模拟的 F 统计量 $F_n(1), \dots, F_n(B)$ 的经验 $1 - \alpha$ 分位数作为置信水平 α 的临界值 c_α 。
- 7: 对于显著水平 α 若 $p_n < \alpha$ 或者等价地 $F_n > c_\alpha$ 时拒绝 H_0 接受 H_1 。

算法 2-1 对带有未知折点的回归模型折点效应的检验算法

2.1.4 模型统计推断

在考虑完折点均值回归模型的折点效应检验后, 自然需要考虑对于系数的统计推断问题。在理想情况下的均值回归模型常使用最小二乘算法完成。假定在数据集具有折点效应的前提下, 已经通过(2-4)与(2-5)式的网格搜索算法得到了折点的位置, 由此考虑 $\beta(\delta_0)$ 是固定 δ_0 时关于系数 β 的 $L(\beta(\delta_0))$ 的极小化变量, 得到的整合后的损失函数为

$$L(\hat{\beta}(\delta_0)) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\beta}(\delta_0)^T \mathbf{x}_i(\delta_0))^2.$$

事实上, 需要 $\beta(\delta_0)$ 满足唯一解假设2.2。

假设 2.2 ^[21] 对于式(2-5)而言:

(C1). $\delta_0 = \underset{\delta \in \Delta}{\operatorname{argmin}} L^*(\delta)$ 是唯一解。

(C2). $\beta \in B \subset \mathbb{R}^{2+p}$, 其中 B 为紧集。□

依据最小二乘的理论, 得到的极小化变量具有如下形式:

$$\beta(\delta_0) = E[x_i(\delta_0)x_i(\delta_0)^T]^{-1}E[x_i(\delta_0)y_i].$$

Hansen^[21] 将 Chan 和 Tsay^[14] 的连续门限自回归模型中斜率与门限系数的最小二乘估计进行了拓展, 得到了参数估计的渐近分布。预先设定 $\theta = (\beta, \delta)$, $\theta_0 = (\beta_0, \delta_0)$,

$$h_i(\theta) = -\frac{\partial(y_i - \beta^T x_i(\delta))}{\partial \theta} = \begin{pmatrix} x_i(\delta) \\ -\beta_1 I(x_i < \delta) - \beta_2 I(x_i > \delta) \end{pmatrix},$$

则有 Hansen^[21] 的定理2.2。

定理 2.2 ^[21] 满足假设2.1时, 有 $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathbf{N}(0, \Sigma)$, 其中 $\Sigma = D^{-1}GD^{-1}$, $G =$

$$\sum_{j=-\infty}^{\infty} E(h_i h_{i+j}^T e_i e_{i+j}),$$

$$D = E(h_i h_i^T) + E \left(\begin{bmatrix} 0 & 0 & 0 & e_i I(x_i < \delta_0) \\ 0 & 0 & 0 & e_i I(x_i > \delta_0) \\ 0 & 0 & 0 & 0 \\ e_i I(x_i < \delta_0) & e_i I(x_i > \delta_0) & 0 & 0 \end{bmatrix} \right),$$

$$h_i = h_i(\theta_0). \quad \square$$

当折点模型被正确构建时显然有 $E(e_i|x_i) = 0$ 。在有了以上定理2.2的保证后，可以得到协方差矩阵的渐近估计。具体见以下命题2.1所示。

命题 2.1 [21] 令 $\hat{h}_i = \begin{pmatrix} x_i(\hat{\delta}) \\ -\hat{\beta}_1 I(x_i < \hat{\delta}) - \hat{\beta}_2 I(x_i > \hat{\delta}) \end{pmatrix}$, $\hat{\Sigma} = \hat{D}^{-1} \hat{G} \hat{D}^{-1}$, 其中 $\hat{G} = \frac{1}{n-k} \sum_{i=1}^n \hat{h}_i \hat{h}_i^T \hat{e}_i^2$ 且

$$\hat{D} = \frac{1}{n} \sum_{i=1}^n \left(\hat{h}_i \hat{h}_i^T + \begin{bmatrix} 0 & 0 & 0 & \hat{e}_i I(x_i < \hat{\delta}) \\ 0 & 0 & 0 & \hat{e}_i I(x_i > \hat{\delta}) \\ 0 & 0 & 0 & 0 \\ \hat{e}_i I(x_i < \hat{\delta}) & \hat{e}_i I(x_i > \hat{\delta}) & 0 & 0 \end{bmatrix} \right),$$

则 $\hat{\Sigma}$ 可以作为 θ_0 的渐近方差。□

事实上，在得出模型系数与折点位置的方差估计后，可以构造出一般的 Wald 置信区间。但是对于折点的估计 $\hat{\delta}$ 而言，考虑到 Wald 置信区间存在着覆盖率不足的一系列问题，Hansen^[21] 采用反转检验置信集可以获得更好的覆盖率。具体的做法是构造关于折点估计检验

$$\begin{cases} H_0 : \delta = \delta_0 \\ H_1 : \delta \neq \delta_0 \end{cases}$$

的检验统计量 $F_n(\delta) = n(\hat{\sigma}^2(\delta) - \hat{\sigma}^2)/\hat{\sigma}^2$ 。该检验会拒绝大于临界值的 F 统计量 $F_n(\delta)$ 。 $\hat{\sigma}^2$ 的形式如(2-6)式所示。由渐近正态性可以知道，该统计量在 H_0 下渐近服从 $\chi^2(1)$ 分布。故对于一个名义上显著水平为 α 的检验可以从 $\chi^2(1)$ 分布获取临界值 $q_{1-\alpha}$ 。故通过这一“反转检验”可得 δ 的 $1 - \alpha$ 置信区间为 $C_\delta = \{\delta : F_n(\delta) \leq q_{1-\alpha}\}$ 。

基于以上的理论，为了进一步提高置信区间覆盖率的精度以及减少样本量之间存在的序列依赖性，Hansen^[21] 提出了一个基于样本的粗糙自助方法。具体的计算置信区间的步骤如下算法2-2所示。

Hansen^[21] 证明了该折点位置的置信区间构造方法相较于 Wald 置信区间构造方法

- 1: 从正态分布 $\mathbf{N}(0, 1)$ 生成 n 个独立同分布的 u_1, \dots, u_n 。
- 2: 设定 $e_i^* = \hat{e}_i u_i$ 其中 \hat{e}_i 是来自单折点回归模型(2-2)的最小二乘的估计的残差。
- 3: 设定 $y_i^* = \hat{\beta}^T x_i(\hat{\delta}) + e_i^*$, 其中 x_i 是样本观测值, 且 $(\hat{\beta}, \hat{\delta})$ 是最小二乘估计。
- 4: 使用 (y_i^*, z_i, x_i) 估计单折点回归模型(2-2)中的参数, 得到参数估计 $(\hat{\beta}^*, \hat{\delta}^*)$, 并且 $\hat{\sigma}^{*2} = n^{-1} \sum_{i=1}^n \hat{e}_i^{*2}$, 其中 $\hat{e}_i^* = y_i^* - \hat{\beta}^{*T} x_i(\hat{\delta}^*)$ 。
- 5: 计算 δ 的 F 统计量 $F_n = \frac{n(\hat{\sigma}^{*2}(\hat{\delta}) - \hat{\sigma}^{*2})}{\hat{\sigma}^{*2}}$, 式中 $\hat{\sigma}^{*2} = n^{-1} \sum_{i=1}^n \hat{e}_i^*(\hat{\delta})^2$ 且 $\hat{e}_i^*(\hat{\delta}) = y_i^* - \hat{\beta}^{*T} x_i(\hat{\delta})$ 。
- 6: 重复 B 次, 以至于得到模拟系数估计 $(\hat{\beta}^*, \hat{\delta}^*)$ 的样本与 F 统计量 F_n 。
- 7: 通过对称分位数的方法构建系数 $\beta_1, \beta_2, \beta_3$ 的 $1 - \alpha$ 自助置信区间: 区间左右端点分别为系数估计加减绝对中心自助估计的 $1 - \alpha$ 分位数。例如, 对于 β_1 而言区间为 $\hat{\beta}_1 \pm q_{1-\alpha}^*$, 其中 $q_{1-\alpha}^*$ 为 $|\hat{\beta}_1^* - \hat{\beta}_1|$ 的 $1 - \alpha$ 分位数。
- 8: 计算自助法得到的 F 统计量 F_n 的 $1 - \alpha$ 分位数 $c_{1-\alpha}^*$ 。
- 9: 以满足 δ 的经验 F 统计量 $F_n(\delta)$ 小于自助临界值 $c_{1-\alpha}^*$ 的 δ 的集合作为 δ 的 $1 - \alpha$ 自助置信区间 $C_\delta^* = \{\delta : F_n(\delta) \leq c_{1-\alpha}^*\}$ 。

算法 2-2 参数的粗糙自助式置信区间构造算法

而言具有更为优秀的覆盖率。尽管置信区间构造时的统计量以 \sqrt{n} 速度收敛到极限分布, 且结合自助法的特性, 构造过程需要大量的时间, 但是鉴于其工作性能良好的优势, 仍旧是值得推广的方法。后续对于实际数据集进行了相关的模拟试验, 相关结果分析在数值模拟部分给出。

Hansen^[21] 不仅给出了关于折点的置信区间构造方法, 还详细给出了回归方程 $g(\theta) = \beta^T x(\delta)$ 的置信区间构造方法。在此处对回归方程的置信区间构造方法做简单介绍。

有关回归方程的置信区间, 最需要关注的是折点回归函数在 $x = \delta$ 位置不可微的问题。但是在折点位置上, 其方向导数存在。因而从方向导数入手考虑问题。对于函数 $\phi(x) : \mathbb{R}^k \rightarrow \mathbb{R}$ 在方向 $t \in \mathbb{R}^k$ 上的方向导数为^[21]

$$\phi_x(t) = \lim_{\varepsilon \rightarrow 0} \frac{\phi(x + t\varepsilon) - \phi(x)}{\varepsilon}, \quad (2-9)$$

而对于回归方程而言在方向 $t = (t_\beta, t_\delta)$ 上的方向导数为^[21]

$$g_\theta(t) = x(\delta)^T t_\beta + g_\delta(t_\delta), \quad (2-10)$$

其中

$$g_\delta(t_\delta) = \begin{cases} -\beta_1 t_\delta, & x < \delta \\ -\beta_1 t_\delta I(t_\delta < 0) - \beta_2 t_\delta I(t_\delta \geq 0), & x = \delta \\ -\beta_2 t_\delta, & x > \delta \end{cases} \quad (2-11)$$

方向导数 $g_\delta(t_\delta)$ 在 $x \neq \delta$ 处线性但在 $x = \delta$ 处非线性, 在左侧具有斜率 $-\beta_1$ 而在右侧具

有斜率 $-\beta_2$ 。

Hansen^[21] 对 Shapiro^[30], Fang 与 Santos^[31] 的已有定理进行了一定程度的推广。他得到如下定理2.3。

定理 2.3 ^[21] 当 $Z \sim \mathbf{N}(0, \Sigma)$ 且 $g_\theta(t)$ 为(2-10)与(2-11)式中的定义形式时, $\sqrt{n}(g(\hat{\theta}) - g(\theta_0)) \xrightarrow{d} g_\theta(Z)$ 。□

Hansen^[21] 出于置信区间的精度考虑, 参考了许多其他的方法, 在他的算法中将定理2.3与自助法相结合。Fang 和 Santos^[31] 建设性地提出了一种与渐近分布相一致的自助法。他们通过 $\hat{g}_\theta(\sqrt{n}(\hat{\theta}^* - \hat{\theta}))$ 估计 $\sqrt{n}(g(\hat{\theta}) - g(\theta_0))$ 的分布, 其中 $\hat{\theta}^*$ 是 $\hat{\theta}$ 的自助法样本且 $\hat{g}_\theta(h)$ 是 $g_\theta(h)$ 的估计。置信区间可以通过这种备选的自助法样本构建出来。

而方法中核心的方向导数的估计问题已经在 Hong 和 Li^[32] 的一篇文章中得到解决。对于(2-9)式他们采用离散模拟估计的方法。对于该问题而言, Hong 和 Li^[32] 的方向导数估计为

$$\hat{g}_\delta(t_\delta) = \frac{g(\hat{\beta}, \hat{\delta} + t_\delta \varepsilon_n) - g(\hat{\beta}, \hat{\delta})}{\varepsilon_n},$$

因此, 我们的方向全导数的估计为

$$\hat{g}_\theta(t) = x(\hat{\delta})' t_\beta + \frac{g(\hat{\beta}, \hat{\delta} + t_\delta \varepsilon_n) - g(\hat{\beta}, \hat{\delta})}{\varepsilon_n}.$$

在 $h = (t_\beta, t_\delta) = (\sqrt{n}(\hat{\beta}^* - \hat{\beta}), \sqrt{n}(\hat{\delta}^* - \hat{\delta}))$ 处进行分析, $g(\hat{\theta}) - g(\theta_0)$ 的分布的自助法估计为

$$r^* = x(\hat{\delta})^T (\hat{\beta}^* - \hat{\beta}) + \frac{g(\hat{\beta}, \hat{\delta} + \sqrt{n} \varepsilon_n (\hat{\delta}^* - \hat{\delta})) - g(\hat{\beta}, \hat{\delta})}{\sqrt{n} \varepsilon_n}.$$

Hansen^[21] 称这种方法为基于 Delta 方法的自助算法, 并指出该方法对于所有 x 值都有渐近保守的覆盖率。其流程具体如下算法2-3所示。该算法对于估计方程的置信区间构造具有较好的覆盖率。Hansen^[21] 通过数值试验证明了这一点。

2.2 单折点分位数回归模型

此前对于单折点均值回归模型, 在损失函数、折点效应检验、参数估计等方面进行系统的介绍。但是均值回归模型存在着较为明显的缺陷: 被分析的对象主要是变量的条件均值, 尽管对于变量的均值有着较好的分析效果, 但是对于距离均值较大的极端数据往往会选择性忽略, 无法提供更多的信息, 进而导致分析上的一些误区; 很多时候研究者对于数据的均值并不感兴趣, 对于极端异常值的研究才是重中之重, 此时的

-
- 1: 遵照算法2-2的粗糙自助方法的步骤 1-3。
 - 2: 设定 $x(\hat{\delta}) = ((x - \hat{\delta})_-, (x - \hat{\delta})_+, z^T)^T$ 。
 - 3: 设定 $\varepsilon_n = cn^{-1/2}$ 。
 - 4: 设定 $r^* = x(\hat{\delta})^T(\hat{\beta}^* - \hat{\beta}) + \frac{g(\hat{\beta}, \hat{\delta} + \sqrt{n}\varepsilon_n(\hat{\delta}^* - \hat{\delta})) - g(\hat{\beta}, \hat{\delta})}{\sqrt{n}\varepsilon_n}$ 。
 - 5: 重复 B 次，因而获得模拟估计量 r^* 的一组样本。
 - 6: 计算 $|r^*|$ 的 $1 - \alpha$ 分位数 $q_{1-\alpha}^*$ 。
 - 7: 设定 g_{θ_0} 的 $1 - \alpha$ 置信区间为 $[g(\hat{\theta}) - q_{1-\alpha}^*, g(\hat{\theta}) + q_{1-\alpha}^*]$ 。
-

算法 2-3 在一 (x, z) 的固定值处的折点回归函数的数值 Delta 方法的自助算法的置信区间

均值回归便没有了用武之地；均值回归常用的最小二乘法对于随机误差项独立同分布的情形有着较好的估计效果，但是对于异方差模型的效果不够显著。

我们希望对于数据集的研究不仅仅局限于均值方面。例如研究社会收入情况，社会收入的均值往往难以反映社会的贫富差距过大问题。这时对于贫困群体与富裕群体的收入研究才是值得关心的问题。在此类问题的背景下，分位数回归应运而生。分位数回归由 Koenker 与 Bassett^[23] 提出，其优势恰好与均值回归的缺陷相对应，分位数回归可以提供更多关于数据集的信息，例如基于解释变量的各分位数等等，这有利于数据集极值的分析；而且，分位数回归对于异常值更为稳健；分位数回归可以更好地解决异方差模型的估计问题。在此基础上，Li 等人^[22] 将单折点均值回归模型推广到了单折点分位数回归模型。该模型主要是针对线性分位数回归的情形，解决了一般线性分位数回归模型无法解决的对于响应变量分段线性但是解释变量连续的回归问题。此部分主要对这一模型进行相关介绍。

2.2.1 模型形式与假设

假定分位数为 $\tau \in (0, 1)$ 。单折点分位数回归模型具有以下形式：

$$Q_Y(\tau|x_i, \mathbf{z}_i, \delta) = \alpha_0 + \alpha_1 x_i + \beta(x_i - \delta)I(x_i > \delta) + \gamma^T \mathbf{z}_i + e_i, \quad i = 1, 2, \dots, n, \quad (2-12)$$

其中 α_0 为截距项， γ 为解释变量 \mathbf{z}_i 的系数向量。可以发现位于折点 δ 左侧样本的线性分位数回归斜率为 α_1 ，而位于折点 δ 右侧样本的线性分位数回归斜率为 $\alpha_1 + \beta$ ，解释变量与响应变量在折点两侧存在着不同的线性关系。与单折点均值回归模型类似，只有折点位于解释变量 x_i 的支撑集内部时才有研究意义，因此 $\delta \in \Delta$ ，其中 Δ 是一个严格嵌入在 x_i 的支撑集内部的紧集。

2.2.2 损失函数与估计方法

令 $\boldsymbol{\eta} = (\alpha_0, \alpha_1, \beta, \gamma^T)^T$, $\mathbf{w}_i(\delta) = (1, x_i, (x_i - \delta)I(x_i > \delta), \mathbf{z}_i^T)^T$ ，故模型可写成

$$Q_Y(\tau|\mathbf{w}_i(\delta)) = \boldsymbol{\eta}^T \mathbf{w}_i(\delta) + e_i.$$

对于一般的分位数回归而言，其损失函数为 $\rho_\tau(y_i - Q_Y(\tau|x_i, \mathbf{z}_i))$ 的形式，其中 $\rho_\tau(u) = u[\tau - I(u < 0)]$ 。在已知折点位置 $\delta = \delta_0$ 的情况下，可以通过解决优化问题

$$\boldsymbol{\eta} = \underset{\boldsymbol{\eta} \in \mathbb{R}^{3+p}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n \rho_\tau[y_i - Q_Y(\tau|\mathbf{w}_i(\delta_0))] \quad (2-13)$$

求解系数向量的估计 $\hat{\boldsymbol{\eta}}$ ，但需要先估计出折点位置。

Li 等人^[22] 运用了 Koenker^[33] 中的轮廓估计技巧，将原有的损失函数(2-13)式转换为下式：

$$L_\tau(\boldsymbol{\eta}, \delta) = \frac{1}{n} \sum_{i=1}^n [\rho_\tau(y_i - Q_Y(\tau|\mathbf{w}_i(\delta))) - \rho_\tau(y_i)]. \quad (2-14)$$

对于固定位置的折点 δ_0 而言，可以得到 $\boldsymbol{\eta}$ 的轮廓估计为

$$\hat{\boldsymbol{\eta}}(\delta) = \underset{\boldsymbol{\eta} \in \mathbb{R}^{3+p}}{\operatorname{argmin}} L_\tau(\boldsymbol{\eta}, \delta), \quad (2-15)$$

而折点估计形式如下：

$$\hat{\delta} = \underset{\delta}{\operatorname{argmin}} L_\tau(\hat{\boldsymbol{\eta}}(\delta), \delta). \quad (2-16)$$

对于该式的求解，常用的方法仍旧是网格搜索方法^[29]，在折点 δ 的大致范围内做格点划分后依次将格点值代入(2-16)进行损失函数值的计算，取(2-16)最小位置处的 $\hat{\delta}$ 值即为折点位置的估计，之后再依照(2-15)式求解 $\boldsymbol{\eta}$ 的估计即可求出所有参数的估计。

2.2.3 折点效应存在性检验

对于折点效应存在性的检验，He 与 Zhu^[34] 提出的综合性的欠拟合检验方法和 Koenker 与 Bassett^[35] 提出的二项式显著性检验均是可行的。这两种方法形式上较为简单。Li 等人^[22] 着重介绍了一种基于自助法的折点检验方法，具有较好的检验效果。

设解释变量 x_1, \dots, x_n 的支撑集为 (a, b) 。对于单折点分位数回归模型，折点效应存在性的检验主要是检验如下假设：

$$\begin{cases} H_0 : \delta \notin (a, b) \\ H_1 : \delta \in (a, b) \end{cases}$$

相较于一般的检验，此处的检验是一个应用 Wald 检验统计量的非条件式检验。具体步骤如下算法2-4所示^[22]。该方法与 Hansen^[21] 检验单折点均值回归模型的折点效应有着一定的相似之处。结合自助法思想的检验具有更显著的检验效果。

- 1: 运用之前的估计方法拟合模型(2-12), 得到对应系数估计 $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}, \hat{\gamma}$ 以及残差项 $\hat{e}_i (i = 1, 2, \dots, n)$ 。
- 2: 通过第一步的估计系数, 利用 $y_i^0 = \hat{\alpha}_0 + \hat{\alpha}_1 x_i + \hat{\gamma}^T \mathbf{z}_i + \hat{e}_i$ 生成原数据样本 $(y_i^0, x_i, \mathbf{z}_i)$ 。
- 3: 从原数据样本 $(y_i^0, x_i, \mathbf{z}_i)$ 中抽样得到自助样本 $(y_i^*, x_i^*, \mathbf{z}_i^*)$ 。
- 4: 通过自助样本 $(y_i^*, x_i^*, \mathbf{z}_i^*)$ 得到折点估计为 $\hat{\delta}^*$ 。
- 5: 基于得到的自助折点估计 $\hat{\delta}^*$ 与原数据样本 $(y_i^0, x_i, \mathbf{z}_i)$ 得到(2-12)中 α_1 与 β 的估计为 $\hat{\alpha}_1^*, \hat{\beta}^*$ 。
- 6: 计算关于 β 的原假设 $H_{0,\beta} : \beta = \hat{\beta}^*$ 的 Wald 检验统计量 F_n 。
- 7: 重复这一过程 B 次, 故可以获得模拟的 F 统计量样本序列 $F_n(1), \dots, F_n(B)$ 。
- 8: 计算 p 值, 即模拟的 F 统计量大于真实值的百分比: $p_n = \frac{1}{B} \sum_{b=1}^B I(F_n(b) \geq F_n)$ 。
- 9: 如果需要的话, 计算模拟的 F 统计量 $F_n(1), \dots, F_n(B)$ 的经验 $1 - \alpha$ 分位数作为置信水平 α 的临界值 c_α 。
- 10: 对于显著水平 α 若 $p_n < \alpha$ 或者等价地 $F_n > c_\alpha$ 时拒绝 H_0 接受 H_1 , 认为存在折点效应。

算法 2-4 单折点分位数回归模型的折点效应自助法检验

2.2.4 模型统计推断

对于单折点分位数回归模型的统计推断部分, 鉴于篇幅原因, 主要介绍参数估计的方差估计理论。在此之前为了后续书写方便, 将前文的 $\boldsymbol{\eta}$ 与 δ 整合为 $\boldsymbol{\theta} = (\boldsymbol{\eta}^T, \delta)^T$ 。

可以注意到, 要最优化损失函数(2-13)事实上等价于求解下面方程:

$$\frac{1}{n} \sum_{i=1}^n \psi_\tau[y_i - Q_Y(\tau|x_i, \mathbf{z}_i, \delta)]h(\mathbf{w}_i) = 0,$$

式中 $\psi_\tau(u) = \tau - I(u < 0)$ 为 $\rho_\tau(u)$ 的导数。在给出渐近分布相关定理之前, 先需要引入以下正则化假设2.3。

假设 2.3 [22] 当折点效应存在, 即满足(2-12)中 $\beta \neq 0$ 时

(C1). 目标函数(2-14)的极限形式 $\lim_{n \rightarrow \infty} L_\tau(\boldsymbol{\eta}, \delta)$ 在参数 $\boldsymbol{\theta}_0$ 位置达到全局唯一最小值。

(C2). $0 < \inf f_{\tau, \mathbf{w}_i} < \sup f_{\tau, \mathbf{w}_i} < \infty$ 恒成立, 其中 f_{τ, \mathbf{w}_i} 为 e_{τ, \mathbf{w}_i} 给定 \mathbf{w}_i 下的条件密度函数。

(C3). $\max_{1 \leq i \leq n} \|\mathbf{z}_i\| = o_p(n^{1/2})$ 且 $E(\|\mathbf{z}_i\|^3) < \infty$ 。

(C4). 存在非负定矩阵 $C_{0,\tau}$ 满足 $C_\tau = n^{-1} \sum_{i=1}^n E[h(\mathbf{w}_i; \boldsymbol{\theta}_0)h(\mathbf{w}_i; \boldsymbol{\theta}_0)^T] \rightarrow C_{0,\tau}$, 其中 $h(\mathbf{w}_i; \boldsymbol{\theta}) = [I(x_i \leq \delta), x_i I(x_i \leq \delta), I(x_i > \delta), x_i I(x_i > \delta), \mathbf{z}_i^T]^T$, $\lim_{n \rightarrow \infty} C_\tau = C_{0,\tau}$ 。

(C5). 存在满秩矩阵 $D_{0,\tau}$ 满足 $D_\tau = n^{-1} \frac{\partial E \sum_{i=1}^n \psi_\tau[y_i - Q_Y(\tau|x_i, \mathbf{z}_i, \delta)]h(\mathbf{w}_i)}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \rightarrow D_{0,\tau}$, 即 $\lim_{n \rightarrow \infty} D_\tau = D_{0,\tau}$ 。□

此时 $\boldsymbol{\theta}$ 估计 $\hat{\boldsymbol{\theta}}$ 具有以下渐近定理[22]。该定理用到了统计学中著名的 Delta 方法[36]。

定理 2.4 ^[22] 当折点效应存在，即满足(2-12)中 $\beta \neq 0$ 以及假设2.3时，有

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathbf{N}(\mathbf{0}, \tau(1 - \tau)D_{0,\tau}^{-1}C_{0,\tau}D_{0,\tau}^{-T}). \quad \square$$

通过定理2.4即可完成参数估计 $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\eta}}^T, \hat{\delta})^T$ 的方差估计，进而构造出对应的 Wald 置信区间。此处不做过多赘述。

3 多折点分位数回归模型

单折点分位数回归模型可以刻画解释变量与响应变量之间不同分位数处的带有单折点的回归关系,折点两侧解释变量与响应变量存在的线性关系不同。但在现实问题中,单折点模型可能无法反映趋势变化多次的数据的真实情况。例如 Royston 与 Sauerbrei^[37]搜集研究的冈比亚女性肱三头肌皮褶厚度与年龄数据,数据大致呈现 0-10 岁阶段肱三头肌皮褶厚度减小,10-20 岁阶段肱三头肌皮褶厚度增大,20 岁以后趋近于稳定的趋势,此时采用单折点分位数回归模型进行分析显然是不合适的。这时就需要引入一个含有多折点的模型以刻画这种多变的趋势。Zhong 等人^[24]对多折点分位数回归模型进行了系统的研究。

多折点分位数回归模型可以较好地刻画解释变量与响应变量之间的变化趋势。与一般的线性模型相比,折点式的回归模型很好地放宽了变量之间的线性假设,因而可以更好地捕捉到变量之间的线性关系的变化;而非参数回归模型相比,折点线性回归模型又具有更好的可解释性,因为对于解释变量 x 的不同区域,其很好地保留了区域之中存在的线性关系,这使得变量之间的关系不会过度受制于数据本身而导致变量间关系的变化难以捕捉。

在这一部分将着重介绍多折点分位数回归模型的参数估计。主要以理论加算法的形式进行介绍。

3.1 模型形式与假设

对于多折点分位数回归模型,不妨假定折点个数为 K ,样本量为 n 。多折点分位数回归模型具有以下形式:

$$Q_Y(\tau; \boldsymbol{\eta}, \boldsymbol{\delta} | x_i, \mathbf{z}_i) = \alpha_0 + \alpha_1 x_i + \sum_{k=1}^K \beta_k (x_i - \delta_k) I(x_i > \delta_k) + \boldsymbol{\gamma}^T \mathbf{z}_i, \quad i = 1, 2, \dots, n, \quad (3-1)$$

式中 $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_K)^T$ 为折点位置向量。从模型形式上可以明显看出,所有的解释变量 x_1, \dots, x_n 被 K 个折点分割为 $K+1$ 个区域,在每个区域上解释变量与响应变量之间存在着不同的线性关系,例如第 m ($2 \leq m \leq K+1$) 个区域的二者之间的线性关系斜率为 $\alpha_1 + \sum_{k=1}^{m-1} \beta_k$ 。

不妨令 $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_K)^T$, 系数向量 $\boldsymbol{\eta} = (\alpha_0, \alpha_1, \boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$, 待估参数包含系数向量与折点位置向量,记为 $\boldsymbol{\theta} = (\boldsymbol{\eta}^T, \boldsymbol{\delta}^T)^T = (\alpha_0, \alpha_1, \beta_1, \beta_2, \dots, \beta_K, \boldsymbol{\gamma}^T, \delta_1, \delta_2, \dots, \delta_K)^T$ 。在完成参数整合后,显然对于折点位置而言有以下假设。

假设 3.1 ^[24] $\boldsymbol{\delta} \in \Delta \subseteq \Lambda^K$, 其中 Λ 表示 x_1, \dots, x_n 的支撑集,且 Δ 是包含于 Λ^K 一个紧集。 $\boldsymbol{\eta} \in H \subseteq \mathbb{R}^{2+K+p}$, 且 H 是一个包含于 \mathbb{R}^{2+K+p} 的一个紧集。□

这一假设保证所有的折点 $\delta_1, \dots, \delta_K$ 均位于解释变量的支撑集内部,折点回归模型

不会退化为一般的线性模型。在这一假设下我们对于折点位置的研究才有意义。

3.2 参数估计方法

本节主要介绍模型的参数估计方法。首先考虑已知折点数 K 情况下折点位置的估计，这是该模型估计的重中之重；对于实际数据而言，折点数 K 的个数也许不得而知，这就需要对折点数 K 进行估计；在得到折点数 K 的估计以及 K 个折点位置的估计后，就可以利用一般的线性分位数回归方法得到模型各个部分的系数估计；最后，对所有待估参数组成向量 $\theta = (\eta^T, \delta^T)^T$ 的方差进行估计。

3.2.1 损失函数

多折点分位数回归模型归根结底还是分位数回归模型，其损失函数的形式与单折点情形下类似，主要是将模型的折点数推广到 K 个的形式。基于待估参数向量 θ 而言，损失函数写成以下形式：

$$L_n(\theta) = n^{-1} \sum_{i=1}^n \rho_\tau[y_i - Q_Y(\tau; \theta | x_i, z_i)]. \quad (3-2)$$

那么对于 θ 而言的估计可写成：

$$\hat{\theta} = (\hat{\eta}^T, \hat{\delta}^T)^T = \underset{\eta \in H, \delta \in \Delta}{\operatorname{argmin}} L_n(\theta), \quad (3-3)$$

最终的参数估计问题就转化为一个优化问题。

3.2.2 折点位置估计

对于单折点的模型，无论是均值回归还是分位数回归模型，为了得到折点位置而求解类似于(3-3)的优化问题，均是应用网格搜索算法。网格搜寻算法的优势在于将折点所在的 Δ 域做相关划分后，通过对每一格点的计算得到折点位置估计，这和线性规划的思想有些许类似，该方法具有较好的可解释性。但是该方法的计算量较大，且主要是应对损失函数 $\rho_\tau(\cdot)$ 不可导情况下所采用的权宜之计。而对于此处的多折点分位数回归模型，采用网格搜索算法求解 K 个折点位置的估计的做法是不可行的。主要原因在于对于多个折点的情况，相较于原有的单折点情况，网格搜索算法的计算复杂度呈 K 的指数级增长。而且损失函数(3-3)对于折点位置 δ 而言也是一个非凸的函数，这使得一般的凸优化算法在此处没有用武之地。因此，Zhong 等人^[24]将 Muggeo^[25]中为了解决基于折线线性最小二乘回归问题的参数化估计所使用的线性化技巧与 Efron^[26]的自助法的思想相结合，给出了用于折点位置估计与系数估计的自助式迭代分段分位数回归算法。Zhong 等人^[24]将这一算法简称为 BRISQ 算法。

首先介绍 Muggeo^[25]中用于估计折点位置而给出的线性化技巧。这种估计折点位置的算法是一种迭代算法，此处用上标的 (t) 表示第 t 次迭代的结果。假定已知模型

的折点数为 K 。对于 $(x_i - \delta_k)I(x_i > \delta_k)$ 而言, δ_k 位置处是不可微的。不妨假定 K 维折点位置向量的初值为 $\boldsymbol{\delta}^{(0)} = (\delta_1^{(0)}, \delta_2^{(0)}, \dots, \delta_K^{(0)})^T$, 那么可以得到在初值 $\delta_k^{(0)}$ 处的 $(x_i - \delta_k)I(x_i > \delta_k)$ 的一阶 Taylor 展开为

$$(x_i - \delta_k)I(x_i > \delta_k) \approx (x_i - \delta_k^{(0)})I(x_i > \delta_k^{(0)}) - (\delta_k - \delta_k^{(0)})I(x_i > \delta_k^{(0)}), \quad (3-4)$$

那么在(3-4)的基础上, 可以得到新的协变量 $\phi_k = \beta_k(\delta_k - \delta_k^{(0)})$, $u_{ki} = (x_i - \delta_k^{(0)})I(x_i > \delta_k^{(0)})$, $v_{ki} = -I(x_i > \delta_k^{(0)})$, 因此模型表达式(3-1)就有以下近似式:

$$Q_Y(\tau; \boldsymbol{\theta} | x_i, \mathbf{z}_i) \approx \tilde{Q}_Y(\tau; \boldsymbol{\theta} | x_i, \mathbf{z}_i) = \alpha_0 + \alpha_1 x_i + \sum_{k=1}^K \beta_k u_{ki} + \sum_{k=1}^K \phi_k v_{ki} + \boldsymbol{\gamma}^T \mathbf{z}_i. \quad (3-5)$$

由此得到了新的待估系数向量 $\boldsymbol{\xi} = (\alpha_0, \alpha_1, \boldsymbol{\beta}^T, \boldsymbol{\phi}^T, \boldsymbol{\gamma}^T)^T$ 。而损失函数(3-2)可近似转化为

$$\tilde{L}_n(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^n \rho_\tau[y_i - \tilde{Q}_Y(\tau; \boldsymbol{\theta} | x_i, \mathbf{z}_i)].$$

通过一般的线性分位数回归方法可以得到(3-5)中的所有系数估计。折点位置 δ_k 可以通过以下式子进行迭代估计:

$$\delta_k^{(1)} = \delta_k^{(0)} + \frac{\hat{\phi}_k^{(1)}}{\hat{\beta}_k^{(1)}}, \quad (3-6)$$

其中 $\hat{\phi}_k^{(1)}$ 与 $\hat{\beta}_k^{(1)}$ 来自基于折点初值位置 $\boldsymbol{\delta}^{(0)}$ 下的(3-5)式中的第一步系数估计 $\boldsymbol{\xi}^{(1)} = (\alpha_0^{(1)}, \alpha_1^{(1)}, \boldsymbol{\beta}^{(1)T}, \boldsymbol{\phi}^{(1)T}, \boldsymbol{\gamma}^{(1)T})$ 。因此, 可以通过第 t 次得到的折点位置 $\boldsymbol{\delta}^{(t)}$ 用一般线性分位数回归的方法得到(3-5)中的第 $t+1$ 步系数估计 $\boldsymbol{\xi}^{(t+1)} = (\alpha_0^{(t+1)}, \alpha_1^{(t+1)}, \boldsymbol{\beta}^{(t+1)T}, \boldsymbol{\phi}^{(t+1)T}, \boldsymbol{\gamma}^{(t+1)T})$, 通过(3-6)的形式得到第 $t+1$ 步的折点位置估计 $\boldsymbol{\delta}^{(t+1)}$ 。不断迭代直到算法收敛为止。

值得一提的是, 对于多折点的情形, Zhong 等人^[24]进行了更为全面的考虑。他们考虑到对于多折点分位数回归模型, 如果 K 个折点位置的初值选取不够恰当, 很有可能导致第 t 次折点位置的迭代值 $\boldsymbol{\delta}^{(t)}$ 会跳出假设3.1中的紧集 Δ 以及解释变量 x_1, \dots, x_n 的支撑集; 也很可能会导致两个折点位置的估计过近, 从而导致两个折点难以分辨, 以至于两个折点合并为一个折点。由此 Zhong 等人^[24]定义了一个关于折点位置估计 $\hat{\boldsymbol{\delta}}$ 的不许可集^[24], 即定义3.1。

定义 3.1 ^[24] $\hat{\boldsymbol{\delta}}$ 的不许可集为

$$\Upsilon = \left\{ \hat{\delta}_k : (\hat{\delta}_k \notin \Lambda, \Lambda \text{ 表示 } x \text{ 的支撑集}) \cup (|\hat{\delta}_k - \hat{\delta}_l| < \varepsilon, \text{ 其中 } \varepsilon \text{ 是一小值且 } k \neq l) \right\}. \quad \square$$

在迭代过程中位于不许可集中的折点需要直接舍弃掉, 以防止折点位置估计合并所带来的影响。与此同时, 为了防止因为局部线性估计对于初值 $\boldsymbol{\delta}^{(0)}$ 敏感而使得算法受困于局部最优而无法达到全局最优的问题发生, Zhong 等人^[24]在算法中结合了自助

抽样法的思想，提出了 BRISQ 算法。

首先要对折点位置估计给定一个初值。一般而言，折点位置的初值选取与 K 有一定关系，一般会选择解释变量 x_1, \dots, x_n 的 $\frac{1}{K+1}, \frac{2}{K+1}, \dots, \frac{K}{K+1}$ 分位数位置作为折点位置估计的初值，进而通过(3-6)的迭代估计得到收敛后的 $\hat{\delta}^{(0)}$ 作为自助法中的首次折点位置估计。由此可通过线性分位数回归的系数估计方法得到(3-5)式中的系数估计 $\hat{\eta}^{(0)}$ 进而得到待估参数的首次估计 $\hat{\theta}^{(0)} = (\hat{\eta}^{(0)T}, \hat{\delta}^{(0)T})^T$ 。接着，在原数据集 $\Xi = (x_i, y_i, z_i)$ 通过随机抽样的方式，得到自助样本 Ξ^* 。在自助样本上利用此前作用于原数据集已收敛的折点位置估计 $\hat{\delta}^{(0)}$ 进行此前的迭代算法从而得到首次的自助辅助估计为 $\tilde{\theta}_*^{(1)} = (\tilde{\eta}_*^{(1)T}, \tilde{\delta}_*^{(1)T})^T$ 。此后在原数据集 Ξ 上使用自助辅助折点估计 $\tilde{\delta}_*^{(1)}$ 作为折点迭代初值进而获得一个全新的自助参数估计为 $\tilde{\theta}^{(1)} = (\tilde{\eta}^{(1)T}, \tilde{\delta}^{(1)T})^T$ 。得到该自助参数估计后，计算自助参数估计与原有的参数估计的损失函数值 $L_n(\tilde{\theta}^{(1)})$ 与 $L_n(\hat{\theta}^{(0)})$ 并进行比较，若 $L_n(\tilde{\theta}^{(1)}) < L_n(\hat{\theta}^{(0)})$ 则更新 $\hat{\theta}^{(1)} = \tilde{\theta}^{(1)}$ ，否则更新 $\hat{\theta}^{(1)} = \hat{\theta}^{(0)}$ 。如此反复以上过程直到算法收敛。可以事先设定自助法抽样的次数 B ，进行多次自助法抽样，最后通过 Monte-Carlo 估计的方式得到最终的参数估计。

具体的 BRISQ 算法的步骤如下算法3-1所示。Zhong 等人^[24] 证明了该算法的收敛性。该算法对于参数估计具有较高的准确性。

3.2.3 折点数估计

以上折点位置估计算法对于折点数 K 已知情况下具有较好的效果。但是在实际数据的分析过程中，折点数 K 往往是未知的。这就需要对存在折点效应的数据集进行折点数估计。

关于折点数估计的理论方法为从一个较大的折点数初值开始，通过反复应用 BRISQ 算法3-1拟合模型(3-1)，在每次拟合中舍弃掉落入定义3.1的不许可集的 $\hat{\delta}_k$ ，直到得到的 $\hat{\delta}_k$ 不属于不许可集，得到折点数的最终估计为 \hat{K} 。但是该方法存在一定的缺陷：得到的 \hat{K} 总是比 K 的真值 K_0 大，即 $\hat{K} > K_0$ 。因此 Zhong 等人^[24] 决定通过最小化信息准则的方法提升模型选择与估计的准确性。他们所主推的信息准则为 sBIC 信息准则^[24]：

$$\text{sBIC}(K) = \log(L_n(\hat{\theta}_K)) + (2 + p + 2K) \frac{\log n}{2n} G_n, \quad (3-7)$$

式中的 $\hat{\theta}_K$ 表示折点数为 K 时采用 BRISQ 算法得到的所有待估参数的估计， $(2+p+2K)$ 对应所有待估参数向量的维数，而 G_n 为一正常数，且满足 $\lim_{n \rightarrow \infty} G_n = \infty$ 。

sBIC 准则可以看成是各类 BIC 信息准则的整合^[24]。取 $G_n = 1$ ，即为标准分位数 BIC 准则^[38]；当 $G_n > 1$ ，其形式又类似于修正式的 BIC 信息准则。BIC 及一系列由 BIC 推广的信息准则被广泛应用于模型选择的研究。此处对于折点数的估计实质上也可以看成是一类模型选择问题。

sBIC 信息准则进行折点数 K 估计的思路大致如下所示：首先给定一个较大的折点

- 1: 初始化折点位置参数的初值 $\delta^{(0)}$ ，一般通过均匀选取初值的方式进行折点位置初始化（选择解释变量 x_1, \dots, x_n 的 $1/(K+1), 2/(K+1), \dots, K/(K+1)$ 分位数位置作为折点位置估计的初值）。
- 2: 基于折点位置初值 $\delta^{(0)}$ ，使用标准的线性分位数回归拟合(3-5)式，进而得到 β 与 ϕ 的第一次估计为 $\beta^{(1)}$ 与 $\phi^{(1)}$ ，从而通过(3-6)更新折点位置估计，得到 $\delta^{(1)}$ 。
- 3: 重复第二步直到折点位置估计收敛或折点落入定义3.1的不许可集，得到的折点位置估计记为 $\hat{\delta}^{(0)}$ 。此时得到的参数估计记作 $\hat{\theta}^{(0)} = (\hat{\eta}^{(0)T}, \hat{\delta}^{(0)T})^T$ 。
- 4: 设定自助法抽样样本量 n^* 以及抽样次数为 B 。
- 5: 对于 $b = 1 : B$ ，完成以下步骤：
- 6: 依据自助法抽样样本量 n^* 生成自助样本 Ξ^* 。
- 7: 对于自助样本 Ξ^* ，利用步骤 2 与 3 的迭代算法得到此次自助法抽样的折点位置估计 $\hat{\delta}^{(b-1)}$ 。接着在自助样本 Ξ^* 上基于 $\hat{\delta}^{(b-1)}$ 利用标准线性分位数回归的估计方法得到自助法辅助参数估计向量 $\tilde{\theta}_*^{(b)} = (\tilde{\eta}_*^{(b)T}, \tilde{\delta}_*^{(b)T})^T$ 。
- 8: 设置 $\tilde{\delta}_*^{(b)}$ 为折点位置估计的迭代初值，利用步骤 2 与 3 的迭代算法得到基于原数据集 Ξ 的折点位置估计 $\hat{\delta}^{(b)}$ 。进而在原数据集 Ξ 上基于 $\hat{\delta}^{(b)}$ 利用标准线性分位数回归的估计方法得到自助参数估计向量 $\hat{\theta}^{(b)} = (\hat{\eta}^{(b)T}, \hat{\delta}^{(b)T})^T$ 。
- 9: 比较两个损失函数值 $L_n(\tilde{\theta}^{(b)})$ 与 $L_n(\hat{\theta}^{(b-1)})$ 的大小。若 $L_n(\tilde{\theta}^{(b)}) < L_n(\hat{\theta}^{(b-1)})$ 则更新 $\hat{\theta}^{(b)} = \tilde{\theta}^{(b)}$ ，否则更新 $\hat{\theta}^{(b)} = \hat{\theta}^{(b-1)}$ 。
- 10: 结束自助抽样算法的流程。
- 11: 基于此前的自助法抽样，得到 B 次抽样中损失函数最小的参数估计为 $\hat{\theta}^{min} = \underset{\hat{\theta}^{(b)}, 1 \leq b \leq B}{\operatorname{argmin}} L_n(\hat{\theta}^{(b)})$ ， B 次抽样中满足抽样标准的次数为

$$B_\varepsilon = \# \left\{ b_\varepsilon : \left| \frac{L_n(\hat{\theta}^{(b_\varepsilon)}) - L_n(\hat{\theta}^{min})}{L_n(\hat{\theta}^{min})} \right| \leq \varepsilon, \varepsilon \text{ 为人为设定的小量} \right\},$$

进而得到最终的参数估计为 $\hat{\theta} = \frac{1}{B_\varepsilon} \sum_{b_\varepsilon} \hat{\theta}^{(b_\varepsilon)}$ 。

算法 3-1 自助式迭代分段分位数算法 (BRISQ)

数 K 的估计值 K_{\max} ，该值需要明显大于 K 的真值 K_0 。接着对于 $K = K_{\max}$ 采用 BRISQ 算法，排除所有位于不许可集中的 $\hat{\delta}_k$ ，直到算法收敛。此时再删除掉一部分进入不许可集的 $\hat{\delta}_k$ 后，可以得到此时的 K 的估计值记为 K^* 。接着对 $K = K^*$ 和 $K = K^* - 1$ 计算 $\text{sBIC}(K)$ 的值，且对 $\text{sBIC}(K^*)$ 与 $\text{sBIC}(K^* - 1)$ 。若 $\text{sBIC}(K^*) < \text{sBIC}(K^* - 1)$ ，则取 $K = K^*$ ，终止算法；反之取 $K = K^* - 1$ ，继续之前的步骤，直到得到最小的 $\text{sBIC}(K)$ 为止。

该算法的思想与模型选择中的后项消除思想类似。与此同时，该方法具有相关的极限理论作为支撑，可以说明该方法的合理性。在此之前，需要满足以下假设3.2。

假设 3.2 ^[24] 设分布函数为 $F_i = F(u|x_i, \mathbf{z}_i)$ ，条件分位数为 $F_i^{-1}(\tau|(x_i, \mathbf{z}_i)) = \inf\{u : F(u|(x_i, \mathbf{z}_i)) \geq \tau\}$ 。定义损失函数对参数一阶导数向量为

$$h(x_i, \mathbf{z}_i|\theta) = [1, x_i, (x_i - \delta_1)_+, \dots, (x_i - \delta_K)_+, \mathbf{z}_i^T, -\beta_1 I(x_i > \delta_1), \dots, -\beta_K I(x_i > \delta_K)]^T。$$

(C1). 函数 F_i 具有连续密度函数 $f_i(u|x_i, \mathbf{z}_i)$ 且在 (x_i, \mathbf{z}_i) 的分位点 $F^{-1}(\tau|x_i, \mathbf{z}_i)$ 满足 $0 < \inf_i f_i < \sup_i f_i < \infty$ 。

(C2). 矩阵 $E[h(x_i, \mathbf{z}_i|\boldsymbol{\theta})h(x_i, \mathbf{z}_i|\boldsymbol{\theta})^T]$ 有限且正定。

(C3). $\lim_{n \rightarrow \infty} \frac{G_n \log(n)}{n} = 0$ 。□

假设3.2中的第三条是为了防止 G_n 发散太快以至于模型欠拟合的情况发生。在以上模型假设成立的条件下，有以下定理3.1。

定理 3.1 [24] 在假设3.2成立的条件下，令 $\hat{K} = \underset{K=0, \dots, K^*}{\operatorname{argmin}} \text{sBIC}(K)$ ，则当 $n \rightarrow \infty$ 时， $P(\hat{K} = K_0) \rightarrow 1$ 成立。□

该定理可以保证该折点数估计方法的合理性与稳健性，由此可以得到关于折点数估计 \hat{K} 的后项消除算法。具体步骤如下算法3-2所示。该算法可以和模型的折点效应检验配合使用。折点效应检验后续会介绍。

- 1: 设定折点数的初值为 K_{\max} ，要求 K_{\max} 远大于模型折点数真值 K_0 。重复算法(3-1)的第二步，不断移除位于不许可集中的折点位置估计 $\hat{\delta}_k$ ，直到算法收敛。由此可以移除所有位于不许可集中的折点位置估计，此时的折点数记为 K^* 。
- 2: 利用算法(3-1)估计 $K = K^*$ 时的待估参数向量估计为 $\hat{\boldsymbol{\theta}}_{K^*}$ 以及 $K = K^* - 1$ 时的待估参数向量估计为 $\hat{\boldsymbol{\theta}}_{K^*-1}$ 。接着计算 $\text{sBIC}(K^*)$ 与 $\text{sBIC}(K^* - 1)$ 。
- 3: 将 $\text{sBIC}(K^*)$ 与 $\text{sBIC}(K^* - 1)$ 进行比较。若 $\text{sBIC}(K^* - 1) < \text{sBIC}(K^*)$ 则更新 $K^* = K^* - 1$ 并重复第二步；若 $\text{sBIC}(K^* - 1) \geq \text{sBIC}(K^*)$ 则终止算法， K 的估计为 $\hat{K} = K^*$ 。若 $K^* = 0$ 则直接终止算法，得出结论：数据集中不存在折点。

算法 3-2 折点数估计的后项消除算法

3.2.4 方差估计的极限分布

接下来研究参数估计的方差估计。这实质上是一个极限分布问题。在方差估计的极限分布推导过程中，常常使用 White^[36] 给出的 Delta 方法。在这里使用 Delta 方法构造极限分布时，与以往的一些估计情形一样，需要考虑和损失函数一阶导以及二阶导有关的统计量。在此处，做以下定义3.2。

定义 3.2 [24] 定义 Delta 方法相关矩阵：

$$C_n = E \left\{ \left(\frac{\partial L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \right\} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = n^{-1} \sum_{i=1}^n \tau(1-\tau) E[h(x_i, \mathbf{z}_i; \boldsymbol{\theta}_0)h^T(x_i, \mathbf{z}_i; \boldsymbol{\theta}_0)],$$

$$D_n = E \left\{ \frac{\partial^2 L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right\} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = n^{-1} \sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\theta}} E\{\psi_\tau[y_i - Q_Y(\tau; \boldsymbol{\theta}|x_i, \mathbf{z}_i)]\} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},$$

其中 $\boldsymbol{\theta}_0 = (\boldsymbol{\eta}_0^T, \boldsymbol{\delta}_0^T)^T = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} L(\boldsymbol{\theta})$ ，即为损失函数最小值位置处的参数估计。□

此处运用 Delta 方法的极限分布定理还要求模型及其变量满足以下假设3.3。

假设 3.3 ^[24] 对于模型(3-1)及定义3.2,

(C1). 分布函数 $F_i = F(u|x_i, \mathbf{z}_i)$ 具有连续密度函数 $f_i(u|x_i, \mathbf{z}_i)$ 且在 (x_i, \mathbf{z}_i) 的任意序列值的分位点 $F^{-1}(\tau|x_i, \mathbf{z}_i)$ 满足 $0 < \inf_i f_i < \sup_i f_i < \infty$ 。

(C2). 损失函数 $L_n(\boldsymbol{\theta})$ 在 $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ 处有全局唯一最优解。

(C3). 解释变量 x_i 在一紧集 $[-M, M]$ 上有连续密度函数, 其中 M 为一正常数。

(C4). $\max_{1 \leq i \leq n} \|\mathbf{z}_i\| = o_p(\sqrt{n})$ 且 $E(\|\mathbf{z}\|)$ 有界。

(C5). 当给定折点数 K 且 $\boldsymbol{\beta} \neq \mathbf{0}$ 时, 存在一个非负定矩阵 C 与一个满秩矩阵 D 满足 $\lim_{n \rightarrow \infty} C_n = C, \lim_{n \rightarrow \infty} D_n = D$ 。□

在以上假设中, 假设 (C1) 是所有分位数回归的总假设, 假设 (C2) 确保了估计的可识别性, 假设 (C5) 是为了满足参数估计的渐近正态性。而假设 (C1)-(C4) 可用于证明参数估计的一致性。在假设3.3成立的基础上, 就可得到以下与参数估计相关的极限定理3.2。

定理 3.2 ^[24] 在假设3.3成立且折点数的真值 K_0 给定的情况下, 对于模型(3-1)的估计为 $\hat{\boldsymbol{\theta}}$, 则有

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathbf{N}(0, \boldsymbol{\Sigma}),$$

其中 $\boldsymbol{\Sigma} = D^{-1}CD^{-1}$ 。□

在得到以上关于参数估计向量的极限定理后, 自然而然可得到以下与方差估计相关的命题3.1。

命题 3.1 ^[24] 参照定理3.2的形式, 可得到 $\boldsymbol{\Sigma}$ 的估计为 $\hat{\boldsymbol{\Sigma}}_n = \hat{D}_n^{-1}\hat{C}_n\hat{D}_n^{-1}$, 其中 $\hat{C}_n = n^{-1} \sum_{i=1}^n \tau(1-\tau)h(x_i, \mathbf{z}_i|\hat{\boldsymbol{\theta}})h^T(x_i, \mathbf{z}_i|\hat{\boldsymbol{\theta}})$, $\hat{D}_n = n^{-1} \sum_{i=1}^n \hat{f}_i(\hat{e}_i)h(x_i, \mathbf{z}_i|\hat{\boldsymbol{\theta}})h^T(x_i, \mathbf{z}_i|\hat{\boldsymbol{\theta}})$, 而 $\hat{e}_i = y_i - Q_Y(\tau; \hat{\boldsymbol{\theta}}|x_i, \mathbf{z}_i)$ 。□

在此处还遗留的问题在于 \hat{D}_n 依赖于随机误差项 e_i 的条件密度函数 f_i 的一致估计 \hat{f}_i 。这是关于随机误差项的估计 \hat{e}_i 的密度函数估计的问题, 该问题可以采用 Hendricks 与 Koenker^[39] 提出的 Hendricks-Koenker 三明治密度估计公式解决。该密度估计公式是一种离散导数的形式, 其具体形式为

$$\hat{f}_i[Q_Y(\tau; \boldsymbol{\theta}|x_i, \mathbf{z}_i)] = \frac{2h_n}{Q_Y(\tau + h_n; \hat{\boldsymbol{\theta}}|x_i, \mathbf{z}_i) - Q_Y(\tau - h_n; \hat{\boldsymbol{\theta}}|x_i, \mathbf{z}_i)}, \quad (3-8)$$

其中 h_n 为一带宽光滑参数, 满足 $\lim_{n \rightarrow \infty} h_n = 0$ 。 $Q_Y(\tau \pm h_n; \hat{\boldsymbol{\theta}}|x_i, \mathbf{z}_i)$ 表示给定分位数水平为 $\tau + h_n$ 与 $\tau - h_n$ 时的(3-1)式的模型函数估计。关于此处的带宽 h_n , 一般的选择有两种: Hall-Sheather 带宽^[40] 与 Bofinger^[41] 带宽。Hall-Sheather 带宽形式如下:

$$h_n = 1.57n^{-1/3} \left\{ \frac{1.5\phi^2[\Phi^{-1}(\tau)]}{2[\Phi^{-1}(\tau)]^2 + 1} \right\}^{1/3},$$

式中 ϕ 和 Φ 分别表示标准正态分布的密度函数与分布函数。Bofinger 带宽形式如下：

$$h_n = n^{-1/5} \left\{ \frac{4.5\phi^4[\Phi^{-1}(\tau)]}{(2[\Phi^{-1}(\tau)]^2 + 1)^2} \right\}^{1/5}.$$

对于不同的数据集，估计 \hat{f}_i 时可以基于不同的带宽进行比较，进而选择更为合适的带宽进行方差估计。

3.3 模型折点效应检验

一个显然的事实是，当且仅当模型存在折点效应的情况下，使用多折点分位数回归模型才是合理的。因此在本节中对多折点分位数回归模型的折点效应进行检验。此时检验的原假设和备择假设为：

$$\begin{cases} H_0 : \text{对所有 } k = 1, 2, \dots, K, \beta_k = 0 \\ H_1 : \text{对某个 } k = 1, 2, \dots, K, \beta_k \neq 0 \end{cases} \quad (3-9)$$

当原假设成立时，用线性分位数回归模型即可分析；而备择假设成立时，在 τ 分位数位置存在折点。与此同时， β_k 依赖于 $\tau \in (0, 1)$ 。此时需要考虑一般线性分位数回归相关的统计量作为检验统计量。

由此可以考虑和一般线性分位数回归相关的统计量^[23]

$$\Omega_n(\tau) = \sup_{\delta \in \Delta} |R_n(\delta)|, \quad (3-10)$$

其中 $R_n(\delta) = n^{-1/2} \sum_{i=1}^n \psi_\tau(y_i - \hat{\zeta}^T \mathbf{w}_i)(x_i - \delta)I(x_i \leq \delta)$, $\mathbf{w}_i = (1, x_i, \mathbf{z}_i^T)^T$, $\zeta = (\alpha_0, \alpha_1, \gamma^T)^T$ 。可以发现向量内积 $\zeta^T \mathbf{w}_i$ 是一般线性分位数回归的模型形式，而 $\hat{\zeta} = \underset{\zeta}{\operatorname{argmin}} \sum_{i=1}^n \rho_\tau(y_i - \zeta^T \mathbf{w}_i)$ 表示一般线性分位数回归模型中 ζ 的估计形式。此时统计量 $\Omega_n(\tau)$ 可以看作一般线性分位数回归中基于分位数回归残差的符号形成的累积式检验统计量。在原假设下，即数据服从一般线性分位数回归的情形，残差 $y_i - \hat{\zeta}^T \mathbf{w}_i$ 应该位于零点附近，由此导致检验统计量 $\Omega_n(\tau)$ 取一个较小值；反之在备择假设下，即数据服从带折点的分位数回归的情形，模型被识别错误，因而导致残差变大，从而导致检验统计量 $\Omega_n(\tau)$ 偏大。

要进行检验，势必要给出检验统计量的渐近分布。在此处考虑一个简单的单折点模型即可^[22]：

$$Q_y(\tau|\mathbf{w}_i) = \alpha_0 + \alpha_1 x_i + n^{-1/2} \beta(x_i - \delta)I(x_i > \delta) + \gamma^T \mathbf{z}_i. \quad (3-11)$$

令 $H_n = n^{-1} \sum_{i=1}^n E[\mathbf{w}_i \mathbf{w}_i^T f_i(e_i)]$, $H_{1n}(\delta) = n^{-1} \sum_{i=1}^n E[\mathbf{w}_i(x_i - \delta)I(x_i < \delta)f_i(e_i)]$,

$H_{2n}(\delta, \beta) = n^{-1} \sum_{i=1}^n E[\mathbf{w}_i \beta(x_i - \delta) I(x_i > \delta) f_i(e_i)]$ 。这得到相关的检验统计量的渐近分布之前，需要满足以下假设3.4。

假设 3.4 [24] 对于此前定义的矩阵而言，

(C1). 存在一个正定矩阵 H 满足 $\lim_{n \rightarrow \infty} H_n = H$ ，同时存在两个一致约束于 δ, β 的矩阵 $H_1(\delta)$ 与 $H_2(\delta, \beta)$ 满足 $\lim_{n \rightarrow \infty} H_{1n}(\delta) = H_1(\delta)$, $\lim_{n \rightarrow \infty} H_{2n}(\delta, \beta) = H_2(\delta, \beta)$ 。

(C2). 对于随机误差项 e_i 的密度函数 $f_i(i = 1, 2, \dots, n)$ 而言，其有界且一阶可导。

□

在满足假设3.4的以及假设3.3的 (C1) 的条件下，以下渐近定理3.3成立。

定理 3.3 [24] 在(3-9)的原假设 H_0 成立时，对于单折点模型(3-11)，有

$$\Omega_n(\tau) \Rightarrow \sup_{\delta \in \Delta} |R(\delta) + v(\delta, \beta)|,$$

其中 \Rightarrow 表示弱收敛， $R(\delta)$ 为一均值为 0，协方差函数为

$$\text{Cov}[R(\delta), R(\delta^*)] = \tau(1 - \tau) E\{[(x_i - \delta)I(x_i < \delta) - U(\delta)][(x_i - \delta^*)I(x_i < \delta^*) - U(\delta^*)]\}$$

的 Gauss 过程，其中 $U(\delta) = H_1^T(\delta)H^{-1}\mathbf{w}_i$ ， $v(\delta, \beta) = -H_1(\delta)H^{-1}H_2(\delta, \beta)$ 。□

依据定理3.3，在原假设 H_0 成立的条件下，有 $v(\delta, \beta) = 0$ 且 $R_n(\delta)$ 收敛于一个零均值 Gauss 过程。而当备择假设 H_1 成立的条件下， $v(\delta, \beta) \neq 0$ 且检验统计量 $\Omega_n(\tau)$ 为一个正值。只需要对(3-11)式推广到多个折点的形式，就可以对多折点模型的每一个折点效应都进行检验。

Zhong 等人[24] 将这一 Gauss 过程弱收敛定理与粗糙自助法相结合[21]，给出了用于检验折点效应的算法。流程如下算法3-3所示。

-
- 1: 依据(3-10)式计算检验统计量 $\Omega_n(\tau)$ 。
 - 2: 设定自助法抽样次数 B 。
 - 3: 对于 $b = 1 : B$ ，完成以下步骤：
 - 4: 从分布式 $\mathbf{N}(0, 1) - \Phi^{-1}(\tau)$ 生成独立同分布的自助样本 $\{w_1, \dots, w_n\}$ 。
 - 5: 从两点累积分布 $P(m_i = 1) = P(m_i = -1) = 0.5$ 生成独立于 w_i 的独立同分布自助样本 $\{m_1, \dots, m_n\}$ 。
 - 6: 计算修正后的检验统计量 $\Omega_{nb}^*(\tau) = \sup_{\delta} |R_n^*(\delta)|$ ，其中

$$R_n^*(\delta) = n^{-1/2} \sum_{i=1}^n m_i \psi_{\tau}(w_i) [(x_i - \delta)I(x_i < \delta) - H_{1n}(\delta)H_n^{-1}\mathbf{w}_i]。$$
 - 7: 结束 B 次循环。
 - 8: 计算 $\{\Omega_{n1}^*(\tau), \dots, \Omega_{nB}^*(\tau)\}$ 中大于第一步计算得到的检验统计量 $\Omega_n(\tau)$ 的比例作为最后 p 值。
-

算法 3-3 检验折点效应的粗糙自助算法

值得注意的是，第四步中需要减去标准正态分布的对应 τ 分位数，这是为了消除基于分位数回归模型的抽样过程中分位数产生的影响。当统计量 $\Omega_n(\tau)$ 的取值大于临界

值时，需要拒绝原假设，因此将通过 B 次自助抽样计算得到的修正统计量大于原统计量的比例作为检验 p 值。不难看出，第六步中的修正后的检验统计量实际上为定理3.3中右侧弱收敛统计量去掉 $v(\delta, \beta)$ 仅剩 Gauss 过程的形式，这与(3-10)的统计量有着相似性，因此可以用于是否拒绝原假设的检验。

3.4 折点位置置信区间构造

在多折点分位数回归模型的统计推断方面，主要考虑折点位置的置信区间构造。Zhong 等人^[24] 对于此问题构造了一个基于光滑秩得分检验的反转检验式置信区间。具体是对于 $\tau \in (0, 1)$ ，给定 x_i 上的 K 个折点位置向量 $\tilde{\delta}$ ，做如下的假设检验：

$$\begin{cases} H_0 : \delta = \tilde{\delta} \\ H_1 : \delta \neq \tilde{\delta} \end{cases} \quad (3-12)$$

该置信区间构造的具体思想为：假定检验水平为 α ，对于此前得到的 δ 的位置估计 $\hat{\delta}$ 做(3-12)式的检验，若不拒绝原假设，则在折点位置估计 $\hat{\delta}$ 上加一小量 ε ，做如下的假设检验：

$$\begin{cases} H_0 : \delta = \hat{\delta} + \varepsilon \\ H_1 : \delta \neq \hat{\delta} + \varepsilon \end{cases} \quad (3-13)$$

若仍旧不拒绝(3-13)式的检验，将检验的目标再加上一小量 ε ，对 $\delta = \hat{\delta} + 2\varepsilon$ 做相关检验，直到对假设 $\delta = \hat{\delta} + u\varepsilon$ 做检验时拒绝原假设为止，则可得到折点位置置信区间的右端点为 $\hat{\delta} + u\varepsilon$ 。对于置信区间的左端点而言，步骤和右端点一样，只需将加一个小量 ε 改为依次减去一个小量即可。若直到对于假设 $\delta = \hat{\delta} - l\varepsilon$ 做检验时拒绝了原假设，最后可得到折点位置的 $(1 - \alpha)$ 置信区间为 $[\hat{\delta} - l\varepsilon, \hat{\delta} + u\varepsilon]$ 。

在得知通过假设检验构造置信区间的思想后，关键在于如何进行这一光滑秩得分检验。Muggeo^[42] 指出，对于门限模型（折点模型）而言，由于其某处不光滑与不可微的特性，原始得分统计量会降低检验的势。为了应对模型中示性函数部分 $I(x_i > \delta_k)$ 的不光滑问题，Zhong 等人^[24] 考虑核函数光滑化之后的模型(3-1)，即用光滑的标准正态分布的分布函数 $\Phi(\frac{x_i - \delta_k}{h_k})$ 对示性函数 $I(x_i > \delta_k)$ 进行近似。其中 h_k 表示核函数的带宽。在进行这样的操作后，原模型(3-1)转化为以下形式：

$$\tilde{Q}_y(\tau; \theta | x_i, z_i) = \alpha_0 + \alpha_1 x_i + \sum_{k=1}^K \beta_k (x_i - \delta_k) \Phi\left(\frac{x_i - \delta_k}{h_k}\right) + \gamma^T z_i, \quad i = 1, 2, \dots, n, \quad (3-14)$$

式中 $\theta = (\eta^T, \delta^T)^T$, $\eta = (\alpha_0, \alpha_1, \beta^T, \gamma)$ 。

下面介绍光滑秩得分检验的检验统计量构造。

对于(3-14)而言，在假设检验(3-12)式的待检验折点位置 $\tilde{\delta}$ 处可以得到对应的系数估计为 $\hat{\eta}(\tilde{\delta})$ 。在得到待检验位置的系数估计后，就可以考虑(3-14)式中关于待检验位置

的一阶偏导数 $A_i(\tau; \hat{\eta}(\tilde{\delta}), \tilde{\delta}) = (a_{i1}, a_{i2}, \dots, a_{iK})^T$, 其中 a_{ik} 表示第 i 个样本在第 k 个折点 δ_k 处的关于其待检验对象 $\tilde{\delta}_k$ 的偏导数

$$a_{ik} = -\hat{\beta}_k \Phi\left(\frac{x_i - \tilde{\delta}_k}{h_k}\right) - \frac{\hat{\beta}_k}{h_k}(x_i - \tilde{\delta}_k)\phi\left(\frac{x_i - \tilde{\delta}_k}{h_k}\right).$$

在得到对(3-14)式中对于待检验位置 $\tilde{\delta}$ 处的系数估计 $\hat{\eta}(\tilde{\delta})$ 后, 就可以得到基于原模型(3-1)在假设检验(3-12)中原假设成立条件下的对应残差为 $\tilde{e}_i = y_i - Q_y(\tau; \theta|x_i, z_i)$ 。接下来需要基于这个残差 \tilde{e}_i 对于此前得到的关于待检验位置的一阶偏导数 $A_i(\tau; \hat{\eta}(\tilde{\delta}), \tilde{\delta})$ 进行适当的修正。设矩阵 $A = (A_1(\tau; \hat{\eta}(\tilde{\delta}), \tilde{\delta}), \dots, A_n(\tau; \hat{\eta}(\tilde{\delta}), \tilde{\delta}))$, 与样本及待检验折点位置 $\tilde{\delta}$ 有关的变量 $G_i(\tilde{\delta}) = (1, x_i, (x_i - \tilde{\delta}_1)_+, \dots, (x_i - \tilde{\delta}_K)_+, z_i^T)^T$, 矩阵 $G(\tilde{\delta}) = [G_1(\tilde{\delta}), \dots, G_n(\tilde{\delta})]^T$ 。由此可得以下修正后的一阶偏导数矩阵:

$$A^* = (I_n - \Lambda)A, \quad (3-15)$$

其中 I_n 为 n 阶单位阵, $n \times n$ 阶矩阵 Λ 定义为

$$\Lambda = G(\tilde{\delta})[G^T(\tilde{\delta})\Psi G(\tilde{\delta})]^{-1}G^T(\tilde{\delta})\Psi, \quad (3-16)$$

式中 $\Psi = \text{diag}(\hat{f}_1(\tilde{e}_1), \dots, \hat{f}_n(\tilde{e}_n))$, 即为假设检验(3-12)中原假设成立条件下 \tilde{e}_i 的密度函数估计形成的 $n \times n$ 对角阵形式。而(3-15)式中的 $n \times K$ 矩阵 A^* 此时可以被认为是将偏导数向量 A 投影在 $G(\tilde{\delta})$ 上的残差形式, 而 $A_i^*(\tau; \hat{\eta}(\tilde{\delta}), \tilde{\delta})$ 是矩阵 A^* 的第 i 行。

由此可以得到构造光滑秩得分检验统计量的中间量:

$$S_n^* = n^{-1/2} \sum_{i=1}^n A_i^*(\tau; \hat{\eta}(\tilde{\delta}), \tilde{\delta}) \psi_\tau(\tilde{e}_i), V_n = n^{-1} \sum_{i=1}^n \tau(1 - \tau) A_i^*(\tau; \hat{\eta}(\tilde{\delta}), \tilde{\delta}) A_i^*(\tau; \hat{\eta}(\tilde{\delta}), \tilde{\delta})^T,$$

最后得到光滑秩得分的 SRS 检验统计量^[24]:

$$\text{SRS}_n(\tau) = S_n^{*T} V_n^{-1} S_n^*. \quad (3-17)$$

直觉上看出, 当(3-12)的原假设成立, 会使得 \tilde{e}_i 趋近于 0, 进而导致(3-17)的 SRS 检验统计量呈现一个小值; 反之, 当 SRS 检验统计量呈现一个较大的值, 往往为拒绝原假设提供了有力证据。与此同时, 关于 SRS 检验统计量的极限分布, 有以下命题 3.2 成立。

命题 3.2 ^[24] 在假设 3.3 的 1、4、5 与假设 3.4 均成立与(3-12)原假设成立条件下, 当 $n \rightarrow \infty$, 有 $\text{SRS}_n(\tau) \xrightarrow{d} \chi^2(K)$ 。□

由此可以利用该检验统计量极限分布为 χ^2 分布的命题, 进行光滑秩得分检验, 进而得到折点位置估计的置信区间。具体的置信区间计算算法流程^[24] 如下算法 3-4 所示。通过以上算法可以依次求得每个折点位置估计的置信区间。Zhong 等人^[24] 证明了该方法比 Wald 置信区间构造方法有着更好的真值覆盖率。

-
- 1: 从算法3-2得到折点数估计 \hat{K} ，算法3-1得到折点位置估计 $\hat{\delta}$ 。设检验水平为 α 。
 - 2: 对于 $k = 1 : \hat{K}$ 而言：
 - 3: 通过光滑秩得分检验得到 δ_k 的上界 $\hat{\delta}_k^u$ ：
 - 4: 检验 $H_{0k} : \delta_k = \tilde{\delta}_k^u$ 其中 $\tilde{\delta}_k^u = \hat{\delta} + \varepsilon$ ， ε 为一大于 0 的小量。
 - 5: 若假设 H_{0k} 未被拒绝，则令 $\tilde{\delta}_k^u = \tilde{\delta}_k^u + \varepsilon$ ，继续重复第 4 步，直到假设 H_{0k} 被拒绝，得到最后的 $1 - \alpha$ 置信区间右端点为 $\hat{\delta}_k^u = \tilde{\delta}_k^u$ 。
 - 6: 通过光滑秩得分检验得到 δ_k 的下界 $\hat{\delta}_k^l$ ：
 - 7: 检验 $H_{0k} : \delta_k = \tilde{\delta}_k^l$ 其中 $\tilde{\delta}_k^l = \hat{\delta} - \varepsilon$ ， ε 为一大于 0 的小量。
 - 8: 若假设 H_{0k} 未被拒绝，则令 $\tilde{\delta}_k^l = \tilde{\delta}_k^l - \varepsilon$ ，继续重复第 4 步，直到假设 H_{0k} 被拒绝，得到最后的 $1 - \alpha$ 置信区间左端点为 $\hat{\delta}_k^l = \tilde{\delta}_k^l$ 。
 - 9: 最后得到的 $\hat{\delta}_k$ 的 $1 - \alpha$ 置信区间为 $[\hat{\delta}_k^l, \hat{\delta}_k^u]$ ， $k = 1, 2, \dots, K$ 。
-

算法 3-4 光滑秩得分检验计算折点位置置信区间

4 算法改进

本部分主要介绍对多折点分位数回归模型中的折点位置估计算法进行的一些改进。本部分侧重于算法改进的原理及算法推导，具体的改进后的模拟仿真试验结果将在下一章节中展示。

4.1 算法改进 1: 线性部分核光滑化

4.1.1 改进原理

对于多折点分位数回归模型，此前在估计折点位置时，采用了线性化的技巧^[25]。该技巧可以有效弥补网格搜索法的缺陷——计算成本较大、折点的估计局限在离散的网格格点上等等。这是为了应对多折点分位数回归模型中折点位置处不可微的特性而动用的技巧。但是对于带折点的分位数回归模型而言，因为回归模型附带了折点，这使得对这一模型的折点估计本身就是一件较为困难的事情。其中问题的最具困难之处还是在于损失函数是一个非光滑、非处处可微的函数，这极大地增加了估计的挑战性。而此前的线性化技巧进行折点估计的算法也存在一定的缺陷，对变点估计低估的情况时有发生。

对于损失函数的光滑化技巧，首先能想到的是从模型入手。此前由于模型中存在示性函数，因而导致了模型(3-1)本身并非是光滑且处处可微的。同时受到此前 3.4 节多折点分位数回归的统计推断中 SRS 检验统计量构造时对模型示性函数部分光滑化的启发，首先对模型(3-1)的示性函数部分进行光滑化。关于核函数，Aizerman^[43] 最早在势函数方法的研究中就将这一技术引入到机器学习领域，但是该方法首先由 Horowitz^[44] 提出。与此同时，Zhou 和 Zhang^[27] 证明了该方法的极限性质以及相关的可行性。

在将模型(3-1)的示性函数部分采用标准正态分布的核函数光滑化后，可以得到如下的模型：

$$\tilde{Q}_y(\tau; \boldsymbol{\eta}, \boldsymbol{\delta} | x_i, \mathbf{z}_i) = \alpha_0 + \alpha_1 x_i + \sum_{k=1}^K \beta_k (x_i - \delta_k) \Phi\left(\frac{x_i - \delta_k}{h_k}\right) + \boldsymbol{\gamma}^T \mathbf{z}_i, \quad i = 1, 2, \dots, n, \quad (4-1)$$

其中的带宽必须满足假设4.1。

假设 4.1 ^[27] 当样本量 $n \rightarrow \infty$ 时，有带宽 $h_k \rightarrow 0 (k = 1, 2, \dots, K)$ 成立。□

在假设4.1成立的情况下，自然有以下定理4.1成立。

定理 4.1 ^[27] 当假设4.1成立，且当 $h \rightarrow 0$ 时有

$$\Phi\left(\frac{x-t}{h}\right) = I(x > t) + o(h)$$

成立。□

在定理4.1的条件下，就可以将模型(4-1)看作是模型(3-1)的近似。由此估计(3-1)式中的参数问题就转化为估计(4-1)式中的参数问题。此时的模型(4-1)式相对于折点位置 δ 而言，是光滑且处处可微的。

不妨像前文一样，将各待估参数做以下整合：令 $\boldsymbol{\eta} = (\alpha_0, \alpha_1, \boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$ 为待估系数向量， $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_K)^T$ 为待估折点位置向量，而所有待估参数向量为 $\boldsymbol{\theta} = (\boldsymbol{\eta}^T, \boldsymbol{\delta}^T)^T$ 。此时的模型的损失函数形式如下所示：

$$L_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \rho_\tau[y_i - \tilde{Q}_y(\tau; \boldsymbol{\theta}|x_i, \mathbf{z}_i)]. \quad (4-2)$$

对于损失函数式(4-2)中的 $\boldsymbol{\theta}$ 求导，可得导数形式为

$$l_n(\boldsymbol{\theta}) = \frac{\partial L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{n} \sum_{i=1}^n \psi_\tau[y_i - \tilde{Q}_y(\tau; \boldsymbol{\theta}|x_i, \mathbf{z}_i)]q(x_i, \mathbf{z}_i; \boldsymbol{\theta}), \quad (4-3)$$

式中 $q(x_i, \mathbf{z}_i; \boldsymbol{\theta})$ 为模型(4-1)式对待估参数向量 $\boldsymbol{\theta}$ 求导之后的向量形式，即

$$q(x_i, \mathbf{z}_i; \boldsymbol{\theta}) = \frac{\partial \tilde{Q}_y(\tau; \boldsymbol{\theta}|x_i, \mathbf{z}_i)}{\partial \boldsymbol{\theta}} = [1, x_i, pz_{i1}, \dots, pz_{iK}, \mathbf{z}_i^T, pd_{i1}, \dots, pd_{iK}],$$

式中 $pz_{ik} = (x_i - \delta_k)\Phi\left(\frac{x_i - \delta_k}{h_k}\right)$, $pd_{ik} = -\beta_k\Phi\left(\frac{x_i - \delta_k}{h_k}\right) - \frac{\beta_k}{h_k}(x_i - \delta_k)\phi\left(\frac{x_i - \delta_k}{h_k}\right)$, 且 $k = 1, 2, \dots, K$ 。

由此，该估计问题被转化为一个求解(4-3)式的方程问题：

$$l_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \psi_\tau[y_i - \tilde{Q}_y(\tau; \boldsymbol{\theta}|x_i, \mathbf{z}_i)]q(x_i, \mathbf{z}_i; \boldsymbol{\theta}) = 0$$

对于该问题的求解，常常采用梯度下降算法。梯度下降算法对于参数向量 $\boldsymbol{\theta}$ 的损失函数 $g(\boldsymbol{\theta})$ 而言，梯度下降的定义式为：

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - r \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \quad (4-4)$$

式中的 $\boldsymbol{\theta}^{(t)}$ 表示待估参数向量 $\boldsymbol{\theta}$ 第 t 次迭代的参数估计值， r 表示每次迭代的学习率。在这里依照梯度下降法的通式(4-4)，可以得到待估参数向量 $\boldsymbol{\theta}$ 的每个参数的更新迭代式如下所示：

$$\begin{cases} \alpha_0^{(t+1)} = \alpha_0^{(t)} + r \frac{1}{n} \sum_{i=1}^n \psi_\tau[y_i - \tilde{Q}_y(\tau; \boldsymbol{\theta}|x_i, \mathbf{z}_i)] \\ \alpha_1^{(t+1)} = \alpha_1^{(t)} + r \frac{1}{n} \sum_{i=1}^n \psi_\tau[y_i - \tilde{Q}_y(\tau; \boldsymbol{\theta}|x_i, \mathbf{z}_i)]x_i \\ \beta_k^{(t+1)} = \beta_k^{(t)} + r \frac{1}{n} \sum_{i=1}^n \psi_\tau[y_i - \tilde{Q}_y(\tau; \boldsymbol{\theta}|x_i, \mathbf{z}_i)](x_i - \delta_k) \Phi\left(\frac{x_i - \delta_k}{h_k}\right) \\ \boldsymbol{\gamma}^{(t+1)} = \boldsymbol{\gamma}^{(t)} + r \frac{1}{n} \sum_{i=1}^n \psi_\tau[y_i - \tilde{Q}_y(\tau; \boldsymbol{\theta}|x_i, \mathbf{z}_i)]\mathbf{z}_i^T \\ \delta_k^{(t+1)} = \delta_k^{(t)} - r \frac{1}{n} \sum_{i=1}^n \psi_\tau[y_i - \tilde{Q}_y(\tau; \boldsymbol{\theta}|x_i, \mathbf{z}_i)]\left[\beta_k \Phi\left(\frac{x_i - \delta_k}{h_k}\right) + \frac{\beta_k}{h_k}(x_i - \delta_k)\phi\left(\frac{x_i - \delta_k}{h_k}\right)\right]. \end{cases} \quad (4-5)$$

式中 $k = 1, 2, \dots, K$ 。在 $\boldsymbol{\theta}$ 的初值 $\boldsymbol{\theta}^{(0)}$ 主要通过以下方式设置：对于折点位置的迭代初值 $\boldsymbol{\delta}^{(0)}$ ，与 3.2 节中的一样，选择解释变量 x_1, \dots, x_n 的 $\frac{1}{K+1}, \frac{2}{K+1}, \dots, \frac{K}{K+1}$ 分位数位置作为折点位置估计的初值。而对于其他系数 $\boldsymbol{\eta}$ 而言，就选择 $\boldsymbol{\delta}^{(0)}$ 作为折点位置估计情况下拟合模型式(3-1)的系数作为系数估计的迭代初值 $\boldsymbol{\eta}^{(0)}$ 。

关于学习率 r ，设置初始的算法学习率为 r_0 ，第 t 步时的迭代学习率为 r_t 。在每次迭代计算完成后，计算对应的损失函数的值。若第 $t+1$ 步迭代的损失函数值 $L_n^{(t+1)}(\boldsymbol{\theta})$ 小于第 t 步迭代的损失函数值 $L_n^{(t)}(\boldsymbol{\theta})$ ，则说明函数正在往正确的方向上进行梯度下降，此时就增大步长（学习率）至 $r_{t+1} = ar_t$ ，且 $a > 1$ 。反之，若 $L_n^{(t+1)}(\boldsymbol{\theta}) \geq L_n^{(t)}(\boldsymbol{\theta})$ ，则说明函数并没有在正确的方向上进行梯度下降，此时就需要减小步长（学习率）至 $r_{t+1} = br_t$ ，且 $b < 1$ 。当前后两步迭代之间的损失函数值几乎不发生变化时，即对于算法容忍度 ε 而言，有 $|L_n^{(t+1)}(\boldsymbol{\theta}) - L_n^{(t)}(\boldsymbol{\theta})| < \varepsilon$ ，则可认为迭代终止，由此得到模型(4-1)式的参数估计 $\hat{\boldsymbol{\theta}}$ 。

而对于估计 $\hat{\boldsymbol{\theta}}$ 的方差估计，在这里与此前 3.2 节中的方差估计思想类似，同样采用 Delta 方法。在此处还是主要介绍相关的定理。先介绍 Delta 方法所用到的与损失函数一阶、二阶导有关的矩阵。此处的矩阵与定义3.2中有着几乎一样的形式：

$$C_n = E \left\{ \left(\frac{\partial L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \right\} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = n^{-1} \sum_{i=1}^n \tau(1-\tau)q(x_i, \mathbf{z}_i; \boldsymbol{\theta})q^T(x_i, \mathbf{z}_i; \boldsymbol{\theta}),$$

$$D_n = E \left\{ \frac{\partial^2 L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right\} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = -\frac{1}{n} \sum_{i=1}^n \hat{f}_i[\tilde{Q}_y(\tau; \boldsymbol{\theta}|x_i, \mathbf{z}_i)]q(x_i, \mathbf{z}_i; \boldsymbol{\theta})q^T(x_i, \mathbf{z}_i; \boldsymbol{\theta}).$$

此处的条件密度函数 \hat{f}_i 在后续仿真模拟试验中采用的还是(3-8)式的 Hendricks-KoenKer 离散导数的形式。具体的带宽选择满足假设4.1的条件即可。

对于参数估计的方差估计的渐近定理，需要满足以下假设4.2。

假设 4.2 [27] 对于模型(4-1)，需要满足以下假设：

(C1). $\boldsymbol{\theta}$ 的唯一真值 $\boldsymbol{\theta}_0 \in \Theta$ ，其中 Θ 是位于参数所在定义域内的一个紧集。

(C2). 对所有的解释变量 x_i ， $\sup_{\boldsymbol{\theta} \in \Theta} |\beta_k(x_i - \delta_k)|, k = 1, 2, \dots, K$ 均有界。

(C3). 对于任意的解释变量 x_i, \mathbf{z}_i ，均有 $\sup_{\boldsymbol{\theta} \in \Theta} \|q(x_i, \mathbf{z}_i; \boldsymbol{\theta})\|^3 < \infty$ 。

(C4). 关于 $\tilde{Q}_y(\tau; \boldsymbol{\theta}|x_i, \mathbf{z}_i)$ 的密度函数 \hat{f}_i 有界且连续。

(C5). 矩阵 C_n 与 D_n 均有界。□

此时有以下关于估计 $\hat{\theta}$ 的渐近定理4.2。

定理 4.2 ^[27] 设 θ_0 为参数唯一真值， $\hat{\theta}$ 为参数的最终估计向量。则在4.1与4.2两个假设均成立的基础上，有 $\hat{\theta} - \theta_0 = O_p(n^{-1/2})$ ，且

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma)$$

其中 $\Sigma = D_n^{-1}C_nD_n^{-T}$ 。□

通过该渐近分布定理即可得到该算法中的参数估计的方差估计。

4.1.2 算法流程

依照 4.1 节的内容，关于模型(4-1)式中的参数估计算法流程如下算法4-1所示。该算法很好地将原有的模型(3-1)进行了光滑化，可以完成对于该模型的参数估计。与此同时，因为结合了梯度下降这一算法，使得算法本身的运算效率很高。但是改进之后确实改变了原有模型的线性部分，这使得当数据的样本量不大时，难以达成标准正态核函数对示性函数的完美替代，因而或多或少地改变了原有模型的本质。因此对于样本量较小的情况，若要采用多折点分位数回归模型进行分析，不太建议使用这一改进算法。

-
- 1: 按照 4.1 节中提及步骤初始化待估参数的迭代初值 $\theta^{(0)} = (\eta^{(0)T}, \delta^{(0)T})^T$ 。
 - 2: 计算迭代初值 $\theta^{(0)}$ 位置处的损失函数值 $L_n(\theta^{(0)})$ 。
 - 3: 通过(4-5)式计算第一步迭代得到的参数结果 $\theta^{(1)}$ 。比较 $L_n(\theta^{(0)})$ 与 $L_n(\theta^{(1)})$ 大小，按照 4.1 节中的学习率更新方法更新学习率。
 - 4: 持续进行第二步与第三步，直到满足 4.1 节中提及的算法迭代终止条件，终止迭代，输出最终的参数估计结果 $\hat{\theta}$ 。
 - 5: 得到最终的参数估计 $\hat{\theta}$ 后，利用渐近定理4.2计算参数估计的方差估计。
-

算法 4-1 折点位置改进算法 1 与参数估计的方差估计计算

4.2 算法改进 2: Bent-Cable 平滑化方法

4.2.1 改进原理

在 4.1 节的改进中，我们将改进的首要问题转移到模型的光滑化上，将示性函数光滑化之后，对于已经处处可微的损失函数，自然可通过求导和梯度下降的方式得到模型的参数估计。第二种改进同样也是采用光滑化的技巧，只是和之前的示性函数的光滑化不同，此处在一定程度上保留了原有模型的线性函数形式，而是着眼于模型(3-1)中导致模型无法处处可微的折点位置。由此引出了 Bent-Cable 平滑化技术。

这一技术由 Chiu^[17] 提出, Das 等人^[20] 对这一方法进行了推广。其思想在于在不可导的折点位置取一带宽, 在该带宽所覆盖的位置上用另一光滑函数做逼近, 从而起到替代折点位置的效果。此时在折点及其邻域上, 模型函数光滑且可导。

在多折点分位数回归模型中, 在折点位置处用一个二次函数进行逼近。此时的模型表达式为:

$$\tilde{Q}_y(\tau; \boldsymbol{\eta}, \boldsymbol{\delta} | x_i, \mathbf{z}_i) = \alpha_0 + \alpha_1 x_i + \sum_{k=1}^K \beta_k S_n(\delta_k, x_i) + \boldsymbol{\gamma}^T \mathbf{z}_i, i = 1, 2, \dots, n, \quad (4-6)$$

其中 $S_n(\delta_k, x_i)$ 的表达式为

$$S_n(\delta_k, x_i) = \begin{cases} 0, & x_i < \delta_k - h_k \\ (x_i - \delta_k + h_k)^2 / 4h_k, & \delta_k - h_k \leq x_i \leq \delta_k + h_k \\ x_i - \delta_k, & x_i > \delta_k + h_k. \end{cases} \quad (4-7)$$

这样, 模型(3-1)式中的 $(x_i - \delta_k)_+$ 也可以看成用(4-7)式进行近似。此处的 h_k 表示函数中所需要的带宽, 同样需要满足假设4.1对于带宽的要求。

不难看出此时的模型(4-7)式光滑且处处可微, 而且在离折点较远的位置模型保留了原有的线性形式。此时的模型的目标损失函数为

$$\begin{aligned} P_n(\boldsymbol{\theta}) &= \frac{1}{n} \sum_{i=1}^n \rho_\tau[y_i - \tilde{Q}_y(\tau; \boldsymbol{\theta} | x_i, \mathbf{z}_i)] \\ &= \frac{1}{n} \sum_{i=1}^n \rho_\tau[y_i - \alpha_0 - \alpha_1 x_i - \sum_{k=1}^K \beta_k S_n(\delta_k, x_i) - \boldsymbol{\gamma}^T \mathbf{z}_i]. \end{aligned}$$

对损失函数式(4-2)中的 $\boldsymbol{\theta}$ 求导并使其为 0, 可得方程

$$p_n(\boldsymbol{\theta}) = \frac{\partial P_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{n} \sum_{i=1}^n \psi_\tau[y_i - \tilde{Q}_y(\tau; \boldsymbol{\theta} | x_i, \mathbf{z}_i)] [-t(x_i, \mathbf{z}_i; \boldsymbol{\theta})] = 0, \quad (4-8)$$

式中 $t(x_i, \mathbf{z}_i; \boldsymbol{\theta})$ 为模型(4-6)式对待估参数向量 $\boldsymbol{\theta}$ 求导之后的向量形式, 即

$$t(x_i, \mathbf{z}_i; \boldsymbol{\theta}) = [1, x_i, S_n(\delta_1, x_i), \dots, S_n(\delta_K, x_i), \mathbf{z}_i^T, \beta_1 s_n(\delta_1, x_i), \dots, \beta_K s_n(\delta_K, x_i)]^T,$$

式中 $s_n(\delta_k, x_i)$ 为 Bent-Cable 结函数 $S_n(\delta_k, x_i)$ 对折点 δ_k 而言的一阶导数形式, 即为

$$s_n(\delta_k, x_i) = \begin{cases} 0, & x_i < \delta_k - h_k \\ -(x_i - \delta_k + h_k) / 2h_k, & \delta_k - h_k \leq x_i \leq \delta_k + h_k \\ -1, & x_i > \delta_k + h_k. \end{cases} \quad (4-9)$$

因此, 该估计问题被转化为一个求解(4-8)式方程的问题。

在这里, 为了方便起见, 仍旧使用梯度下降算法求解该方程。结合(4-4)式的梯度下降的定义式, 可以得到待估参数向量 $\boldsymbol{\theta}$ 的每个参数估计的更新迭代式如下所示:

$$\begin{cases} \alpha_0^{(t+1)} = \alpha_0^{(t)} + r \frac{1}{n} \sum_{i=1}^n \psi_\tau[y_i - \tilde{Q}_y(\tau; \boldsymbol{\theta}|x_i, \mathbf{z}_i)] \\ \alpha_1^{(t+1)} = \alpha_1^{(t)} + r \frac{1}{n} \sum_{i=1}^n \psi_\tau[y_i - \tilde{Q}_y(\tau; \boldsymbol{\theta}|x_i, \mathbf{z}_i)]x_i \\ \beta_k^{(t+1)} = \beta_k^{(t)} + r \frac{1}{n} \sum_{i=1}^n \psi_\tau[y_i - \tilde{Q}_y(\tau; \boldsymbol{\theta}|x_i, \mathbf{z}_i)]S_n(\delta_k, x_i) \\ \boldsymbol{\gamma}^{(t+1)} = \boldsymbol{\gamma}^{(t)} + r \frac{1}{n} \sum_{i=1}^n \psi_\tau[y_i - \tilde{Q}_y(\tau; \boldsymbol{\theta}|x_i, \mathbf{z}_i)]\mathbf{z}_i^T \\ \delta_k^{(t+1)} = \delta_k^{(t)} + r \frac{1}{n} \sum_{i=1}^n \psi_\tau[y_i - \tilde{Q}_y(\tau; \boldsymbol{\theta}|x_i, \mathbf{z}_i)][\beta_k S_n(\delta_k, x_i)] \\ h_k^{(t+1)} = h_k^{(t)} + r \frac{1}{n} \sum_{i=1}^n \psi_\tau[y_i - \tilde{Q}_y(\tau; \boldsymbol{\theta}|x_i, \mathbf{z}_i)] \frac{h_k^{(t)^2 - (x_i - \delta_k)^2}{4h_k^{(t)^2}}, \end{cases} \quad (4-10)$$

式中 $k = 1, 2, \dots, K$ 。需要注意的是，此处的带宽 h_k 也将其看作是参数向量的一部分，因此在迭代过程中每一步的带宽也都需要进行迭代。这是从估计稳健性出发而进行的考虑。在后续的仿真模拟试验中，带宽 h_k 的初值为 $h_k^{(0)} = \hat{\sigma} \times 30^{-0.26}$ 。与此同时，迭代过程中的 $\boldsymbol{\theta}$ 的初值 $\boldsymbol{\theta}^{(0)}$ 以及学习率的相关设置与 4.1 节中的一致。

至于此时的方差估计，与 4.1 节的有很多相似之处，同样是采用 Delta 方法^[36]，“三明治”公式构造关于参数估计 $\hat{\boldsymbol{\theta}}$ 和真值 $\boldsymbol{\theta}_0$ 之间的渐近定理。此时的矩阵 C_n 、 D_n 形式如下：

$$C_n(\boldsymbol{\theta}_0) = E \left\{ \left(\frac{\partial P_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial P_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \right\} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = n^{-1} \sum_{i=1}^n \tau(1-\tau) t(x_i, \mathbf{z}_i; \boldsymbol{\theta}) t^T(x_i, \mathbf{z}_i; \boldsymbol{\theta}),$$

$$D_n(\boldsymbol{\theta}_0) = E \left\{ \frac{\partial^2 P_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right\} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = -\frac{1}{n} \sum_{i=1}^n \hat{f}_i[\tilde{Q}_y(\tau; \boldsymbol{\theta}|x_i, \mathbf{z}_i)] t(x_i, \mathbf{z}_i; \boldsymbol{\theta}) t^T(x_i, \mathbf{z}_i; \boldsymbol{\theta}).$$

条件密度函数 \hat{f}_i 来自于(3-8)式的离散导数形式。此时的渐近定理要满足假设4.3。

假设 4.3 ^[27] 对于模型(4-6)，需要满足以下假设：

(C1). $\boldsymbol{\theta}$ 的唯一真值 $\boldsymbol{\theta}_0 \in \Theta$ ，其中 $\Theta \in \mathbb{R}^{2+2K+p}$ 是一个紧集。

(C2). $E \sup_{\boldsymbol{\theta} \in \Theta} \|S_n(\boldsymbol{\delta}, \mathbf{x})\|$ 与 $E \sup_{\boldsymbol{\theta} \in \Theta} \|\beta_k I(x_i > \delta_k)\|$ 均有限。

(C3). 对于任意的解释变量 x_i 及 \mathbf{z}_i ，均有 $E \sup_{\boldsymbol{\theta} \in \Theta} \|t(x_i, \mathbf{z}_i; \boldsymbol{\theta})\| < \infty$ 。

(C4). 关于 $\tilde{Q}_y(\tau; \boldsymbol{\theta}|x_i, \mathbf{z}_i)$ 的密度函数 \hat{f}_i 有界且连续。

(C5). 矩阵 $C_n(\boldsymbol{\theta}_0)$ 与 $D_n(\boldsymbol{\theta}_0)$ 均为有限的正定矩阵。□

此时有以下关于估计 $\hat{\boldsymbol{\theta}}$ 的渐近定理4.3。

定理 4.3 ^[27] 设 $\boldsymbol{\theta}_0$ 为参数唯一真值， $\hat{\boldsymbol{\theta}}$ 为参数的最终估计向量。则在4.1与4.3两个假设均成立的基础上，有

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}), \quad (4-11)$$

其中 $\boldsymbol{\Sigma} = D_n(\boldsymbol{\theta}_0)^{-1} C_n(\boldsymbol{\theta}_0) D_n(\boldsymbol{\theta}_0)^{-T}$ 。□

通过该渐近分布定理即可得到该算法中得到的参数估计的方差估计。具体算法步骤见算法4-2所示。该算法很好地解决了原多折点分位数回归模型(3-1)存在的折点位置处无法求导的问题，用函数逼近的思想完成了对于参数的估计。但美中不足的是，该方法因为在迭代过程中涉及到 **Bent-Cable** 结函数 $S_n(\delta_k, x_i)$ 以及其导数 $s_n(\delta_k, x_i)$ 的计算，同时需要对带宽也进行迭代，导致算法的运算效率并不高，该算法仍有待改进的地方。

4.2.2 算法流程

- 1: 按照 4.1 节中提及步骤初始化待估参数的迭代初值 $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\eta}^{(0)\top}, \boldsymbol{\delta}^{(0)\top})^\top$ ，同时按照 $h_k^{(0)} = \hat{\sigma} \times 30^{-0.26}$ 选择带宽的初值 $h_k^{(0)}$ 。
- 2: 计算(4-8)与(4-9)式的 **Bent-Cable** 结函数及其导数，同时计算迭代初值 $\boldsymbol{\theta}^{(0)}$ 位置处的损失函数值 $P_n(\boldsymbol{\theta}^{(0)})$ 。
- 3: 通过(4-10)式计算第一步迭代得到的参数结果 $\boldsymbol{\theta}^{(1)}$ 与 $h_k^{(1)}$ 。比较 $P_n(\boldsymbol{\theta}^{(0)})$ 与 $P_n(\boldsymbol{\theta}^{(1)})$ 大小，按照 4.1 节中的学习率更新方法更新学习率。
- 4: 持续进行第二步与第三步，直到满足 4.1 节中提及的算法迭代终止条件，终止迭代，输出最终的参数估计结果 $\hat{\boldsymbol{\theta}}$ 以及最终的模型带宽 h_k 。
- 5: 在得到最终的参数估计结果 $\hat{\boldsymbol{\theta}}$ 和最终带宽 h_k 的前提下，利用渐近定理4.3计算参数的方差估计。

算法 4-2 折点位置改进算法 2 与参数估计的方差估计计算

4.3 算法改进 3: 分位数回归损失函数的替换

4.3.1 改进原理

前面两种算法的改进方法主要是关于模型形式的改进，将模型形式光滑化进而起到光滑损失函数的效果，使得参数估计问题可以通过求导以及梯度下降的方法得到解决。

对于光滑化这一手段而言，也不仅仅是局限于光滑化模型这一个方向。也可以对损失函数本身进行光滑化。因为对于分位数回归模型而言，其损失函数 $\rho_\tau(u) = u[\tau - I(u < 0)]$ 也不是一个处处可微的函数，其在 $u = 0$ 位置处显然是不可导的。因此，自然而然地考虑对于分位数回归的损失函数的光滑化。

Zheng^[45] 提出了一种对于损失函数而言的替换方法，主要研究了光滑函数

$$S_{\tau, \kappa}(u) = \tau u + \kappa \log \left(1 + \exp\left(-\frac{u}{\kappa}\right) \right), \quad (4-12)$$

替代 $\rho_\tau(u)$ 函数情况下的分位数回归。式(4-12)中的 κ 被称为光滑参数。

对于这一光滑函数而言，Zheng^[45] 给出了以下两个重要引理4.1与4.2。

引理 4.1 ^[45] 对于任意 $\kappa > 0$ ， $S_{\tau, \kappa}(u)$ 是关于 u 的凸函数。□

引理 4.2 ^[45] 对于 $\forall u \in \mathbb{R}$ 而言, 有 $0 < S_{\tau, \kappa}(u) - \rho_{\tau}(u) \leq \kappa \log 2$, 因而 $\lim_{\kappa \rightarrow 0^+} S_{\tau, \kappa}(u) = \rho_{\tau}(u)$ 。□

引理4.1表明函数 $S_{\tau, \kappa}(u)$ 是一个凸函数, 因而可以找到全局最优解。引理4.2则是说明了函数 $S_{\tau, \kappa}(u)$ 与 $\rho_{\tau}(u)$ 之间的逼近关系。值得注意的是, 这两个函数之间存在的逼近关系并不依赖于分位数 τ 。Zheng^[45] 建议选择(4-12)式中的 κ 为 0.1 左右。

由引理4.2可知在两个不同的损失函数下得到的系数估计之间存在逼近关系。若假定一般分位数回归的模型形式为 $Q_y(\tau; \boldsymbol{\eta} | x_i, \mathbf{z}_i)$ 。对于损失函数 $\rho_{\tau}(u)$ 得到的分位数回归系数为 $\hat{\boldsymbol{\eta}} = \underset{\boldsymbol{\eta}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n \rho_{\tau}(y_i - Q_y(\tau; \boldsymbol{\eta} | x_i, \mathbf{z}_i))$, 而对于损失函数 $S_{\tau, \kappa}(u)$ 而言得到的分位数回归系数为 $\hat{\boldsymbol{\eta}}_{\kappa} = \underset{\boldsymbol{\eta}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n S_{\tau, \kappa}(y_i - Q_y(\tau; \boldsymbol{\eta} | x_i, \mathbf{z}_i))$ 。则有以下定理4.4。

定理 4.4 ^[45] 在引理(4.1)与(4.2)的基础上

$$\lim_{\kappa \rightarrow 0^+} \hat{\boldsymbol{\eta}}_{\kappa} = \hat{\boldsymbol{\eta}}. \quad \square$$

定理4.4说明了二者的系数估计之间的极限关系。因而对于分位数回归而言, 采用 $S_{\tau, \kappa}(u)$ 替换 $\rho_{\tau}(u)$ 进行系数估计的做法是可行的。

因为带折点的分位数回归在 Zhong 等人^[24] 的理论中系数 $\boldsymbol{\eta}$ 的估计和线性分位数回归的估计几乎一致, 因而对于多折点分位数回归模型的系数估计, 使用光滑函数替换原损失函数的方法是完全可行的。在此处的算法大致思路如下: 利用此前的算法3-1中的第 2、3 步线性化技巧求解模型的折点位置估计, 接着在得到折点位置估计的基础上, 利用 $S_{\tau, \kappa}(u)$ 替代原有的损失函数计算系数估计。此处由于损失函数是处处可微的, 因此可以采用梯度下降算法得到模型的系数估计。

此时的损失函数形式如下:

$$Ls_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n S_{\tau, \kappa} \left[y_i - \alpha_0 - \alpha_1 x_i - \sum_{k=1}^K \beta_k (x_i - \delta_k) I(x_i > \delta_k) - \boldsymbol{\gamma}^T \mathbf{z}_i \right]. \quad (4-13)$$

损失函数替换后对自变量求导的导函数为

$$s_{\tau, \kappa}(u) = \frac{\partial S_{\tau, \kappa}(u)}{\partial u} = \tau - \frac{1}{1 + \exp(\frac{u}{\kappa})}. \quad (4-14)$$

利用(4-4)式的梯度下降的定义式, 可以得到这里的对于系数向量 $\boldsymbol{\eta}$ 而言的每个参数的迭代式为:

$$\begin{cases} \alpha_0^{(t+1)} = \alpha_0^{(t)} + r \frac{1}{n} \sum_{i=1}^n s_{\tau, \kappa}[y_i - \tilde{Q}_y(\tau; \boldsymbol{\theta} | x_i, \mathbf{z}_i)] \\ \alpha_1^{(t+1)} = \alpha_1^{(t)} + r \frac{1}{n} \sum_{i=1}^n s_{\tau, \kappa}[y_i - \tilde{Q}_y(\tau; \boldsymbol{\theta} | x_i, \mathbf{z}_i)] x_i \\ \beta_k^{(t+1)} = \beta_k^{(t)} + r \frac{1}{n} \sum_{i=1}^n s_{\tau, \kappa}[y_i - \tilde{Q}_y(\tau; \boldsymbol{\theta} | x_i, \mathbf{z}_i)] (x_i - \delta_k) I(x_i > \delta_k) \\ \boldsymbol{\gamma}^{(t+1)} = \boldsymbol{\gamma}^{(t)} + r \frac{1}{n} \sum_{i=1}^n s_{\tau, \kappa}[y_i - \tilde{Q}_y(\tau; \boldsymbol{\theta} | x_i, \mathbf{z}_i)] \mathbf{z}_i^T \\ \delta_k^{(t+1)} = \delta_k^{(t)} - r \frac{1}{n} \sum_{i=1}^n s_{\tau, \kappa}[y_i - \tilde{Q}_y(\tau; \boldsymbol{\theta} | x_i, \mathbf{z}_i)] [\beta_k \Phi\left(\frac{x_i - \delta_k}{h_k}\right) + \frac{\beta_k}{h_k} (x_i - \delta_k) \phi\left(\frac{x_i - \delta_k}{h_k}\right)], \end{cases}$$

式中 $k = 1, 2, \dots, K$, 且对于折点位置 $\boldsymbol{\delta}$ 而言, 采用的是此前 3.4 节构造 SRS 统计量时

核函数近似示性函数后的(3-14)式的形式来推导迭代算法。

同时, 对于迭代初值 $\theta^{(0)}$ 以及学习率 r 的相关设置与 4.1 节中一致。方差估计仍旧采用 Delta 方法, 具体方法与 4.1 节、4.2 节中一致, 此处不做赘述。

4.3.2 算法流程

具体算法流程如下算法4-3所示。该改进算法的主要优势在于计算速度快。光滑函数(4-12)式在解释变量的全局凸性明显, 可以快速迭代得到目标损失函数的最优解。但是该方法不利于高维情形下数据的处理^[45], 在高维情况下得到的估计结果十分糟糕。

-
- 1: 按照 4.1 节中提及步骤初始化待估参数的迭代初值 $\theta^{(0)} = (\eta^{(0)\top}, \delta^{(0)\top})^\top$ 。
 - 2: 计算(4-11)与(4-13)式的光滑损失函数及其导数, 同时计算迭代初值 $\theta^{(0)}$ 位置处的损失函数值 $L_{S_n}(\theta^{(0)})$ 。
 - 3: 利用算法3-1的第 2-3 步计算折点位置估计 $\hat{\delta}$ 。
 - 4: 通过(4-14)式计算第一步迭代得到的系数结果 $\eta^{(1)}$ 。比较 $L_{S_n}(\theta^{(0)})$ 与 $L_{S_n}(\theta^{(1)})$ 大小, 按照 4.1 节中的学习率更新方法更新学习率。
 - 5: 持续进行第二步与第三步, 直到满足 4.1 节中提及的算法迭代终止条件, 终止迭代, 输出最终的参数估计结果 $\hat{\theta}$ 。
-

算法 4-3 折点位置改进算法 3

4.4 算法改进 4: 基于损失函数替换的 Boosting 方法

4.4.1 改进原理

鉴于一般的基于将损失函数 $\rho_\tau(\cdot)$ 替换为 $S_{\tau, \kappa}(\cdot)$ 对于高维情况下的处理十分糟糕, Zheng^[45] 提出了基于损失函数替换的 Boosting 方法。该方法类似于高维下的坐标梯度下降方法, 是基于 Boosting 算法框架得到的方法。这一算法框架由 Friedman^[46] 与 Mason 等人^[47] 提出。在这里对这一方法做简单介绍。

Boosting 算法是基于弱学习机得到的算法, 可以认为是泛函式的梯度下降算法。假定 \mathcal{K} 为弱学习机的集合, $\mathcal{K} = \{k_1(x), \dots, k_m(x)\}$, 其中 m 为弱学习机的个数。通过将弱学习机看成是空间中一组基的方法, 可以将弱学习机张成的空间表示为

$$\mathcal{L}(\mathcal{K}) = \left\{ \sum_{t=1}^m c_t k_t(x) \mid c_t \in \mathbb{R}, t = 1, 2, \dots, m \right\}.$$

而对于多折点分位数回归模型(3-1)式而言, 可以将

$$\mathcal{K}(x_i, \mathbf{z}_i, \boldsymbol{\delta}) = [\mathbf{1}, \mathbf{x}, (\mathbf{x} - \delta_1)I(\mathbf{x} > \delta_1), \dots, (\mathbf{x} - \delta_K)I(\mathbf{x} > \delta_K), \mathbf{Z}]$$

看成是弱学习机张成的空间 $\mathcal{L}(\mathcal{K})$ 的一组基。由此依据模型(3-1)式的形式, 需要估计的 $\boldsymbol{\eta}$ 可以看成是 $\mathcal{K}(x_i, \mathbf{z}_i, \boldsymbol{\delta})$ 在这组基下的线性表示的坐标。

因此，原有的求解模型表达式的问题被转化为以下问题的形式：

$$Q_Y(\tau; \boldsymbol{\eta}, \boldsymbol{\delta} | x_i, \mathbf{z}_i) = \underset{Q_Y(\tau; \boldsymbol{\eta}, \boldsymbol{\delta} | x_i, \mathbf{z}_i) \in \mathcal{L}(\mathcal{K})}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n \rho_\tau[y_i - Q_Y(\tau; \boldsymbol{\eta}, \boldsymbol{\delta} | x_i, \mathbf{z}_i)].$$

但是鉴于分位数回归的损失函数 $\rho_\tau(u)$ 在 $u = 0$ 位置处不可微的缺陷，将原有损失函数光滑化后才可以采用 Boosting 算法的框架进行梯度下降算法。因此采用(4-12)式代替损失函数。这样求解模型的问题被转化为以下形式：

$$Q_Y(\tau; \boldsymbol{\eta}, \boldsymbol{\delta} | x_i, \mathbf{z}_i) = \underset{Q_Y(\tau; \boldsymbol{\eta}, \boldsymbol{\delta} | x_i, \mathbf{z}_i) \in \mathcal{L}(\mathcal{K})}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n S_{\tau, \kappa}[y_i - Q_Y(\tau; \boldsymbol{\eta}, \boldsymbol{\delta} | x_i, \mathbf{z}_i)].$$

由此该问题被转化为一个基于坐标的梯度下降问题。基于 Zheng^[45] 中对于一般分位数回归方法的 Boosting 算法，给出应用于多折点分位数回归的系数 Boosting 算法。该算法主要还是需要利用原模型基于系数的负梯度进行计算，不过此处需要将梯度与原解释变量之间的函数关系作为模型表达式的迭代增量。方差估计主要采用 Delta 方法，具体方法与 4.1 节、4.2 节中一致。

4.4.2 算法流程

- 1: 按照 4.1 节中提及步骤初始化待估参数的迭代初值 $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\eta}^{(0)\top}, \boldsymbol{\delta}^{(0)\top})^\top$ 。同时设定算法的迭代容忍度为 ε ，算法最大迭代次数为 T 。
- 2: 利用算法3-1的第 2-3 步计算折点位置估计 $\hat{\boldsymbol{\delta}}$ 。
- 3: 对于算法迭代次数 $t = 1 : T$ 而言：
- 4: 计算第 t 次迭代计算中的对于损失函数的梯度为

$$s_i = \tau - \left[1 + \exp\left(\frac{y_i - Q_Y(\tau; \boldsymbol{\eta}^{(t)}, \boldsymbol{\delta} | x_i, \mathbf{z}_i)}{\kappa}\right) \right]^{-1}, \quad i = 1, 2, \dots, n$$

- 5: 拟合 s_i 与 x_i 之间的函数关系式为 $d^{(t)}(x_i)$ 。并将其作为迭代增量。
- 6: 通过 $Q_Y(\tau; \boldsymbol{\eta}^{(t)}, \boldsymbol{\delta} | x_i, \mathbf{z}_i) = Q_Y(\tau; \boldsymbol{\eta}^{(t+1)}, \boldsymbol{\delta} | x_i, \mathbf{z}_i) + r^{(t)} d^{(t)}(x_i)$ 的方式迭代模型表达式。其中 $r^{(t)}$ 的迭代与 4.1 节中的学习率变化方法一致。
- 7: 判断收敛条件，即对于损失函数(4-13)式而言，若 $|Ls_n^{(t+1)}(\boldsymbol{\theta}) - Ls_n^{(t)}(\boldsymbol{\theta})| < \varepsilon$ ，则终止迭代。
- 8: 结束循环。输出模型表达式。

算法 4-4 折点位置改进算法 4

算法4-4给出基于损失函数替换的 Boosting 改进算法的流程。该算法的计算效率不如算法4-3，但是结合了 Boosting 算法后对于高维数据有较好效果。与此同时，Zheng^[45]通过试验说明该方法具有较好的变量选择功能，可以用于符合特征选择机制的数据集。

4.5 算法改进 5: 诱导光滑化方法

4.5.1 改进原理

本文尝试的最后一种改进算法是基于估计函数的诱导光滑方法的改进^[48]。该方法由 Wang 和 Brown^[48] 于 2005 年提出。Wang 和 Brown^[48] 认为该方法可以应用于任何基于参数的凸目标函数的估计问题，他们在文章中对位置参数的秩估计、分位数估计以及多重秩回归估计进行了光滑化。后续很多专家学者对这一方法进行了推广。例如 Wang 和 Fu^[49] 将这一方法推广到面板数据回归模型，Pang 等人^[50] 将这一方法推广到删失数据的分位数回归等等。本文将这一算法推广到多折点分位数回归模型，以完成对算法的改进。下面简单介绍该方法的原理。

对于一般的估计问题而言，设 $S(\theta)$ 是样本和参数 θ 的函数，通过 $S(\theta) = 0$ 可得到 θ 的估计量 $\hat{\theta}$ 。若 $S(\theta)$ 是一个光滑函数，则自然可进行求导，且 $S(\hat{\theta})$ 可以在真值 θ_0 处展开，即

$$0 = S(\hat{\theta}) = S(\theta_0) + S'(\theta_0)(\hat{\theta} - \theta_0) + o_p(\theta_0).$$

同时 $S'(\theta_0)$ 通常近似为一常数， $S(\theta_0)$ 的方差表示为 $\text{Cov}[S(\theta_0)]$ ，且随着样本量 $n \rightarrow \infty$ ， $\text{Cov}(S(\theta_0))$ 收敛到一半正定矩阵。在此基础上，可以推导出估计 $\hat{\theta}$ 的极限定理，同时利用 Delta 方法构造出估计 $\hat{\theta}$ 的渐近协方差矩阵为

$$\text{Cov}(\hat{\theta}) = D^{-1}CD^{-1},$$

其中 $C = \text{Cov}[S(\theta_0)]$ ， $D = E[S'(\theta_0)]$ 。

但是以上成立的前提条件是 $S(\theta)$ 是一个足够光滑的函数，否则估计的精确度以及估计方差无法被很好地定义。

因此，Wang 和 Brown^[48] 基于“伪贝叶斯”的相关理论，提出了诱导光滑方法，可以很好地得到参数估计 $\hat{\theta}$ 及其协方差 $\text{Cov}[\hat{\theta}]$ 。其核心思想为：参数的真值 θ 与估计 $\hat{\theta}$ 之间的差值通常满足渐近的正态分布

$$\hat{\theta} - \theta \xrightarrow{d} \mathbf{N}(0, \Sigma).$$

因此，将问题的重心放到这个渐近的正态分布上来。具体地，若 θ 是一个向量的形式 $\boldsymbol{\theta}$ ，令

$$\boldsymbol{\theta} = \hat{\boldsymbol{\theta}} + H^{1/2}Z, \quad (4-15)$$

其中 $Z \sim \mathbf{N}(\mathbf{0}, I_p)$ ， I_p 为 p 阶单位阵。对于单参数问题而言， $H = s^2$ ，其中标准差 s 起到了类似于核函数光滑化中带宽 h 的作用。因此，此时的 θ 可以被看成是估计 $\hat{\theta}$ 的随机扰动的形式，二者之差服从一个零均值的正态分布。此时估计函数的诱导光滑函数为

$$\tilde{S}(\boldsymbol{\theta}) = E_Z[S(\boldsymbol{\theta} + H^{1/2}Z)], \quad (4-16)$$

解估计方程 $\tilde{S}(\boldsymbol{\theta}) = \mathbf{0}$ 即可得到估计 $\hat{\boldsymbol{\theta}}$ 。与此同时，因为 $\tilde{S}(\cdot)$ 函数的光滑性质，也可以用“三明治公式”估计方差 H ，即

$$H = \{E[\tilde{S}'(\hat{\boldsymbol{\theta}})]\}^{-1} \text{Cov}[S(\boldsymbol{\theta}_0)] \{E[\tilde{S}'(\hat{\boldsymbol{\theta}})]\}^{-1}. \quad (4-17)$$

(4-16)与(4-17)式给出了参数估计 $\hat{\boldsymbol{\theta}}$ 及其方差 H 的诱导光滑化估计方法。

下面对算法的迭代更新式进行推导。此处的迭代算法主要针对模型(3-1)的系数 $\boldsymbol{\eta}$ 而言，对于模型的折点 $\boldsymbol{\delta}$ ，主要还是采用线性化技巧进行估计。在得到折点估计 $\boldsymbol{\delta}$ 后利用诱导光滑算法计算模型系数估计 $\hat{\boldsymbol{\eta}}$ 。

首先令 $p(\mathbf{w}_i) = p(x_i, \mathbf{z}_i) = [1, x_i, (x_i - \delta_1)I(x_i > \delta_1), \dots, (x_i - \delta_K)I(x_i > \delta_K), \mathbf{z}_i^T]^T$ 。对于多折点分位数回归模型，其损失函数形式为(3-2)式。在将模型的变量整合为 $p(\mathbf{w}_i)$ 的基础上，目标损失函数可以写成

$$L_n(\boldsymbol{\eta}) = n^{-1} \sum_{i=1}^n \rho_\tau[y_i - \boldsymbol{\eta}^T p(\mathbf{w}_i)],$$

此时对损失函数(4-17)式求导，可得损失函数的导数形式为

$$U_n(\boldsymbol{\eta}) = n^{-1} \sum_{i=1}^n \psi_\tau[y_i - \boldsymbol{\eta}^T p(\mathbf{w}_i)] p(\mathbf{w}_i), \quad (4-18)$$

此处对(4-18)式中的 $U_n(\boldsymbol{\eta})$ 进行诱导光滑化即可。其中关于示性函数的期望部分计算如下：

$$E_V \{I[y_i - (\boldsymbol{\eta} + H^{1/2}V)^T p(\mathbf{w}_i) < 0]\} = P[y_i - \boldsymbol{\eta}^T p(\mathbf{w}_i) < p(\mathbf{w}_i)^T H^{1/2}V].$$

令 $\tilde{V} = p(\mathbf{w}_i)H^{1/2}V$ ，则 $\tilde{V} \sim \mathbf{N}(\mathbf{0}, p^T(\mathbf{w}_i)Hp(\mathbf{w}_i))$ 。故

$$E_V \{I[y_i - (\boldsymbol{\eta} + H^{1/2}V)^T p(\mathbf{w}_i) < 0]\} = P[\tilde{V} > y_i - \boldsymbol{\eta}^T p(\mathbf{w}_i)] = \Phi \left(\frac{\boldsymbol{\eta}^T p(\mathbf{w}_i) - y_i}{\sqrt{p^T(\mathbf{w}_i)Hp(\mathbf{w}_i)}} \right).$$

对于 $V \sim \mathbf{N}(\mathbf{0}, I_{(K+2+p)})$ ，可以得到估计函数的诱导光滑函数

$$\begin{aligned} \tilde{U}(\boldsymbol{\eta}, H) &= E_V[U(\boldsymbol{\eta} + H^{1/2}V)] \\ &= E_V[n^{-1} \sum_{i=1}^n p(\mathbf{w}_i) \{\tau - I[y_i - (\boldsymbol{\eta} + H^{1/2}V)^T p(\mathbf{w}_i)]\}] \\ &= n^{-1} \sum_{i=1}^n p(\mathbf{w}_i) \left[\tau - \Phi \left(\frac{\boldsymbol{\eta}^T p(\mathbf{w}_i) - y_i}{\sqrt{p^T(\mathbf{w}_i)Hp(\mathbf{w}_i)}} \right) \right]. \end{aligned} \quad (4-19)$$

与此同时，要对 $\boldsymbol{\eta}$ 使用梯度下降算法，就需要对 $\tilde{U}(\boldsymbol{\eta}, H)$ 进行求导，得到的导数

形式为

$$\tilde{A}(\boldsymbol{\eta}, H) = \frac{\partial \tilde{U}(\boldsymbol{\eta}, H)}{\partial \boldsymbol{\eta}} = n^{-1} \sum_{i=1}^n \left[-\phi \left(\frac{\boldsymbol{\eta}^T p(\mathbf{w}_i) - y_i}{\sqrt{p^T(\mathbf{w}_i) H p(\mathbf{w}_i)}} \right) \frac{p(\mathbf{w}_i) p^T(\mathbf{w}_i)}{\sqrt{p(\mathbf{w}_i)^T H p(\mathbf{w}_i)}} \right].$$

由此可得对于系数向量 $\boldsymbol{\eta}$ 而言的迭代公式为

$$\boldsymbol{\eta}^{(t+1)} = \boldsymbol{\eta}^{(t)} + [-\tilde{A}(\boldsymbol{\eta}^{(t)}, H^{(t)})]^{-1} \tilde{U}(\boldsymbol{\eta}^{(t)}, H^{(t)}), \quad (4-20)$$

其中 $H^{(t)}$ 的每一步迭代的计算式可以遵照(4-17)式的计算方法，具体计算式如下：

$$H^{(t+1)} = n^{-1} \tilde{A}^{-1}(\boldsymbol{\eta}^{(t)}) \text{Cov}[\tilde{U}(\boldsymbol{\eta}^{(t)}, H^{(t)})] \tilde{A}^{-1}(\boldsymbol{\eta}^{(t)}). \quad (4-21)$$

事实上，迭代算法最终得到的 H 即为参数估计的方差估计。(4-21)式中的协方差 $\text{Cov}[\tilde{U}(\boldsymbol{\eta}^{(t)}, H^{(t)})]$ 同样可以采用 4.1 节中的三明治公式进行计算。与此同时也可以证明结论： $n \rightarrow \infty$ 时， $\text{Cov}[\tilde{U}(\boldsymbol{\eta}, H)]$ 与 $\text{Cov}[U(\boldsymbol{\eta}, H)]$ 渐近一致。

4.5.2 算法流程

-
- 1: 按照 4.1 节中提及步骤初始化待估参数的迭代初值 $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\eta}^{(0)T}, \boldsymbol{\delta}^{(0)T})^T$ 。初始化矩阵 $H^{(0)} = n^{-1} I_n$ 。同时设定算法迭代容忍度为 ε ，算法最大迭代次数为 T 。
 - 2: 利用算法3-1的第 2-3 步计算折点位置估计 $\hat{\boldsymbol{\delta}}$ 。
 - 3: 对于算法迭代次数 $t = 0 : T$ 而言：
 - 4: 计算第 t 次迭代计算中的矩阵 $H^{(t)}$ 以及迭代过程的中间量 $\tilde{U}(\boldsymbol{\eta}^{(t)}, H^{(t)})$ 和 $\tilde{A}(\boldsymbol{\eta}^{(t)}, H^{(t)})$ 。通过(4-20)式计算 $\boldsymbol{\eta}^{(t+1)}$ 。利用 Delta 方法和(4-21)式计算 H^{t+1} 。
 - 5: 判断收敛条件，即对于损失函数(3-2)式而言，若 $|L_n^{(t+1)}(\boldsymbol{\theta}) - L_n^{(t)}(\boldsymbol{\theta})| < \varepsilon$ ，则终止迭代。
 - 6: 结束循环。输出 $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\eta}}^T, \hat{\boldsymbol{\delta}}^T)^T$ 为最终的参数估计，输出对角阵 \hat{H} 为最终的参数估计的方差估计。
-

算法 4-5 折点位置估计的诱导光滑化算法

算法4-5给出诱导光滑化方法的具体计算流程。该算法具有如下较为明显的优势：算法过程中直接得到最终的参数估计以及方差估计，无需另外求解方差估计；算法具有很强的理论保证；算法迭代几步就可以收敛，具有较快的收敛速度等等。

5 数值模拟试验

该部分展示与折点模型相关的模拟试验结果。通过这些试验对模型及算法的各方面进行相关评估。所有的数值模拟试验基于 R 语言完成。

5.1 单折点情况下的数值模拟

5.1.1 试验设置

对于单折点模型，主要对单折点均值回归模型的参数估计、单折点分位数回归的参数估计与折点效应检验进行研究。本文通过以下模型生成数据：

$$y_i = \alpha_0 + \alpha_1 x_i + \beta(x_i - \delta)I(x_i > \delta) + \gamma^T \mathbf{z}_i + e_i, \quad i = 1, 2, \dots, 1000 \quad (5-1)$$

其中 $\alpha_0 = 1$, $\alpha_1 = 1$, $\gamma = 1$, $\beta = -3$, $\delta = 0.5$ 。而 $x_i \sim \mathbf{U}(-5, 5)$, $\mathbf{z}_i \sim \mathbf{N}(1, 1)$ 。

而对于(5-1)式中的随机误差项 e_i ，本文模拟了以下 6 种不同情况下的数据生成：对于同方差情形， $e_i \sim \mathbf{N}(0, 1)$, $e_i \sim \mathbf{t}(3)$, $e_i \sim \mathbf{LA}(0, 1)$ ；对于异方差情形，其随机误差项与解释变量 x_i 有关，本文设置为 $e_i = (1 + 0.2x_i)e_i^*$ ，其中 $e_i^* \sim \mathbf{N}(0, 1)$, $e_i^* \sim \mathbf{t}(3)$, $e_i^* \sim \mathbf{LA}(0, 1)$ 。

在此处设计随机误差项 e_i 服从标准 Laplace 分布是出于分位数回归的损失函数进行考虑的。分位数回归的损失函数在 $\tau = 0.5$ 位置处的损失函数与拉普拉斯分布的密度函数存在相似之处。这与概率模型中最小二乘与正态分布密度函数的相似有着异曲同工之妙。

通过(5-1)式设置抽样种子数为 1 - 1000 生成了 $N = 1000$ 组数据。对于这 1000 组数据，分别利用单折点均值回归模型、单折点分位数回归模型以及多折点分位数回归模型（ $K = 1$ 情形）进行拟合，得到 1000 次模型的参数估计与方差估计的均值作为最终的估计结果。进而可以计算得到参数估计的偏（Bias），标准偏差（SD），估计的标准差（SE）以及估计的均方误差（MSE）。通过比对试验结果对各方法进行评估。

对于折点效应的检验，试验设计的思路与 Zhong 等人^[24]有着相似之处。对于(5-1)式而言，设定 $\beta = n^{-1/2}c$ 。此处主要对于 $c = 0, 2, 4, 6, 8, 10$ 情形进行折点效应检验，进而计算检验的势。通过绘制 c 与检验的势之间的图像关系对检验进行评估。在这里，本文主要采用 2 种不同的检验方式，分别对应算法 2-1 中 Hansen^[21]提出的折点检验方法与 3-3 中 Zhong 等人^[24]提出的折点检验方法，将两种方法的 c 与检验势之间的图像进行比较。He 与 Zhu^[34]也曾提出分位数回归的 Lack-of-Fit 检验方法，但是 Zhong^[24]等人已经通过试验证明了该方法对于折点回归模型而言检验效果并不好，在此处不再赘述。此处的随机误差项还是服从此前提到的 6 种情况，对于这 6 种情况，本文都进行了试验及结果比较。

对于置信区间而言,也是通过(5-1)式设置 $\delta = 0.5$ 生成样本。在这里,采用多种方法构造折点位置 δ 的置信区间^[21]:传统 Wald 置信区间构造方法;分位数自助方法;倒转分位数自助方法;对称分位数自助方法;算法2-2的粗糙自助置信区间构造方法;倒转式的粗糙自助方法。这些方法在 Hansen^[21] 中均有体现。

通过以上 6 种置信区间构造方法构造关于折点位置 δ 的 95% 置信区间,比较这些方法的效果从而评估各方法的好坏,其中方法 1 需要用到定理2.2的方差估计方法,通过 $\hat{\delta} \pm 1.96\hat{\sigma}_{\delta}$ 构造置信区间 (± 1.96 分别对应标准正态分布的 0.025 与 0.975 分位数),其中 $\hat{\sigma}_{\delta}$ 表示折点位置 δ 的标准差估计。后续的方法都用到了自助法的思想,本文选用的自助抽样样本量为 500。此处同样通过之前的 6 种情形生成数据,从而计算 6 种不同置信区间构造方法的覆盖率,比较出这些方法的好坏。

5.1.2 数值模拟结果分析

首先给出单折点情形下对应 3 种不同模型的模拟结果对比。表5-1中 SKLS 表示单折点均值回归模型,SKQR 表示单折点分位数回归模型,MKQR 表示多折点分位数回归模型 ($K = 1$ 情形)。对于不同情形下不同参数的 MSE 最小的值,本人用红色字体标出。

从表5-1可以看出,对于随机误差项为标准正态分布的情形,单折点均值回归模型估计均方误差较小,基本上都控制在 0.05 以内;而对于随机误差项为异方差的情形,多折点分位数回归模型的效果总体而言稍好一些。而在方差估计方面,可以看到单折点均值回归模型存在着明显的缺陷,估计的标准差 (SE) 与标准偏差 (SD) 不够接近;而对于多折点分位数回归模型,这二者的值十分接近,这充分说明了 Zhong 等人^[24] 方法在方差估计方面的优势较为明显。

对于折点效应检验的结果见表5-2和5-3。表5-2给出单折点均值回归模型的折点效应检验结果,表5-3给出多折点均值分位数模型的折点效应检验结果。将结果可视化,具体如图5-1所示。图像横坐标表示(5-1)式中的 $\beta = n^{-1/2}c$ 的 c 的值,纵坐标表示 1000 次试验后得到的检验势函数的均值,其中红线表示单折点均值回归的折点效应检验算法2-1,蓝线表示多折点分位数回归的折点效应检验算法3-3。六幅图中第一行对应随机误差项 $e_i \sim \mathbf{N}(0, 1)$,第二行对应随机误差项 $e_i \sim \mathbf{t}(3)$,第三行对应随机误差项 $e_i \sim \mathbf{LA}(0, 1)$ 。左侧三幅图对应随机误差项同方差情况,而右侧对应异方差情况。

从表5-2以及图像5-1可以看出,对于同方差的情形,算法2-1和3-3都可以很好地控制第一类错误。但是对于异方差情形下,当随机误差项 e_i 不再服从标准正态分布时,单折点均值回归的折点效应检验算法2-1并不能很好地阻止第一类错误的发生,但是算法3-3的势要明显高于算法2-1,无论在什么情形下,都可以很好地实现折点效应检验的任务,是更值得推荐的方法。

对于置信区间的构造,具体的模拟结果如表5-3所示。从表中可以明显看出 Wald 置信区间构造方法对于随机误差项 $e_i \sim \mathbf{N}(0, 1)$ 的情形真值覆盖率偏低;倒转分位数自助

表 5-1 三种模型效果比对

			同方差情形					异方差情形				
Method			α_0	α_1	β_1	γ	δ	α_0	α_1	β_1	γ	δ
N(0, 1)	SKLS	Bias	-0.003	0.001	-0.002	-0.002	0.002	0.007	0.001	0.003	-0.002	-0.004
		SD	0.122	0.039	0.062	0.044	0.066	0.107	0.027	0.090	0.053	0.075
		SE	0.038	0.051	0.045	0.103	0.060	0.027	0.082	0.052	0.106	0.070
	SKQR	MSE	0.002	0.003	0.002	0.011	0.004	0.001	0.007	0.003	0.011	0.005
		Bias	0.011	0.003	-0.005	-0.003	-0.004	0.006	0.001	-0.005	0.000	-0.001
		SD	0.142	0.048	0.076	0.056	0.085	0.099	0.022	0.101	0.025	0.089
	MKQR	SE	0.144	0.048	0.081	0.056	0.076	0.111	0.027	0.103	0.037	0.083
		MSE	0.021	0.002	0.007	0.003	0.006	0.012	0.001	0.011	0.001	0.007
		Bias	0.003	-0.002	0.000	-0.004	0.002	0.004	0.000	0.001	-0.000	-0.009
		SD	0.146	0.059	0.050	0.081	0.081	0.099	0.026	0.022	0.106	0.090
		SE	0.144	0.056	0.048	0.081	0.077	0.099	0.023	0.022	0.102	0.082
		MSE	0.021	0.003	0.002	0.007	0.006	0.010	0.001	0.010	0.001	0.007
t(3)	SKLS	Bias	0.004	0.002	-0.006	0.001	-0.003	0.006	-0.002	-0.014	-0.012	0.010
		SD	0.198	0.068	0.106	0.071	0.105	0.194	0.048	0.141	0.087	0.130
		SE	0.064	0.084	0.074	0.171	0.100	0.044	0.134	0.086	0.180	0.116
	SKQR	MSE	0.004	0.007	0.006	0.029	0.001	0.002	0.018	0.007	0.032	0.014
		Bias	0.007	0.004	-0.006	-0.003	0.001	0.009	0.002	-0.001	-0.001	-0.008
		SD	0.164	0.055	0.088	0.065	0.093	0.114	0.026	0.114	0.026	0.093
	MKQR	SE	0.162	0.054	0.091	0.064	0.085	0.131	0.034	0.116	0.047	0.095
		MSE	0.026	0.003	0.008	0.004	0.007	0.017	0.001	0.013	0.002	0.009
		Bias	-0.001	0.000	0.001	-0.005	0.002	0.000	0.000	0.000	-0.001	0.001
		SD	0.157	0.061	0.053	0.084	0.087	0.103	0.027	0.023	0.111	0.091
		SE	0.159	0.061	0.053	0.090	0.084	0.109	0.026	0.024	0.114	0.091
		MSE	0.025	0.004	0.003	0.008	0.007	0.012	0.001	0.001	0.013	0.008
LA(0, 1)	SKLS	Bias	0.010	0.001	-0.002	-0.001	-0.001	0.001	-0.001	-0.011	-0.003	0.011
		SD	0.172	0.057	0.093	0.065	0.095	0.155	0.038	0.128	0.076	0.106
		SE	0.053	0.073	0.063	0.144	0.085	0.037	0.116	0.073	0.151	0.099
	SKQR	MSE	0.003	0.005	0.004	0.021	0.007	0.001	0.014	0.005	0.023	0.010
		Bias	0.006	0.002	0.002	0.001	-0.005	0.000	0.000	0.000	0.000	0.001
		SD	0.131	0.047	0.070	0.051	0.076	0.094	0.021	0.090	0.021	0.079
	MKQR	SE	0.135	0.045	0.077	0.054	0.072	0.112	0.028	0.094	0.041	0.078
		MSE	0.018	0.002	0.006	0.003	0.005	0.012	0.001	0.009	0.002	0.006
		Bias	0.007	0.001	0.002	0.001	0.001	0.000	0.001	0.002	0.004	0.001
		SD	0.131	0.071	0.061	0.080	0.057	0.070	0.066	0.082	0.023	0.031
		SE	0.136	0.075	0.063	0.088	0.052	0.074	0.065	0.079	0.022	0.033
		MSE	0.017	0.002	0.010	0.004	0.006	0.013	0.001	0.010	0.003	0.005

表 5-2 不同 β 真值下折点效应检验的势比较

		c					
Method		0	2	4	6	8	10
N(0, 1) 同方差	SKLS	0.058	0.221	0.701	0.961	1.000	1.000
	MKQR	0.067	0.215	0.607	0.910	0.993	1.000
N(0, 1) 异方差	SKLS	0.050	0.091	0.286	0.671	0.929	0.994
	MKQR	0.145	0.393	0.803	0.974	0.997	1.000
t(3) 同方差	SKLS	0.050	0.115	0.323	0.612	0.832	0.950
	MKQR	0.063	0.209	0.590	0.873	0.978	0.995
t(3) 异方差	SKLS	0.059	0.082	0.149	0.291	0.482	0.692
	MKQR	0.153	0.382	0.766	0.957	0.995	1.000
LA(0, 1) 同方差	SKLS	0.051	0.136	0.394	0.746	0.940	0.989
	MKQR	0.061	0.307	0.742	0.964	0.994	0.999
LA(0, 1) 异方差	SKLS	0.059	0.080	0.168	0.358	0.599	0.828
	MKQR	0.186	0.502	0.871	0.986	0.998	1.000

方法对于所有的情况都存在着欠覆盖的情况，覆盖率均小于理论上设定的 95%。而对于传统分位数自助自助方法和对称分位数自助方法，覆盖率又过高了，此时的置信区间并不具有显著的构造价值；如此比较下来，算法2-2的粗糙自助方法以及其倒转形式

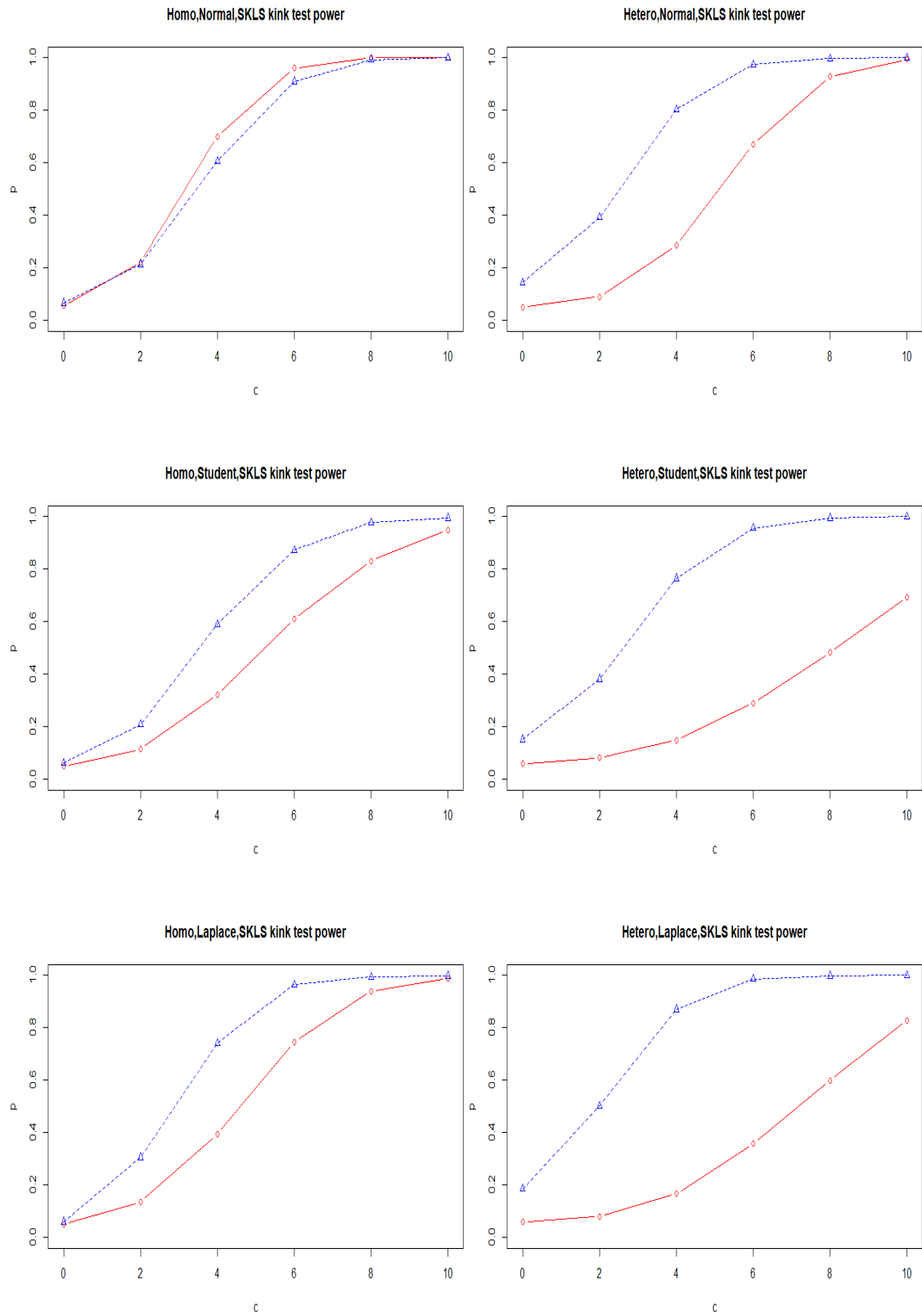


图 5-1 折点效应势检验结果比较

的覆盖率始终在 95% 附近浮动，很好地达到了覆盖率的理论值要求，是值得推荐的方法。

表 5-3 不同置信区间构造方法置信区间真值覆盖率

	e_i					
	同方差			异方差		
	N(0, 1)	t(3)	LA(0, 1)	N(0, 1)	t(3)	LA(0, 1)
Wald	75.3%	94.6%	98.1%	74.9%	92.0%	95.8%
分位数自助法	99.9%	97.4%	98.2%	99.6%	98.6%	96.7%
倒转分位数自助法	88.7%	80.8%	79.3%	85.7%	91.9%	80.9%
对称分位数自助法	99.9%	97.3%	98.2%	99.6%	97.5%	96.2%
粗糙自助法	95.4%	95.1%	95.2%	94.5%	94.6%	95.4%
倒转粗糙自助法	95.3%	95.2%	94.6%	94.7%	94.0%	94.6%

5.2 多折点情况下的数值模拟

5.2.1 试验设置

对于多折点情况下的数值模拟，本文将核心放在 Zhong 等人^[24]的参数估计以及各种改进方法的参数估计上，主要针对算法3-1、4-1、4-2、4-3、4-4和4-5进行比较。本文主要通过以下模型式生成样本：

$$y_i = \alpha_0 + \alpha_1 x_i + \sum_{k=1}^K \beta_k (x_i - \delta_k) I(x_i > \delta_k) + \gamma^T z_i + e_i, \quad i = 1, 2, \dots, 1000 \quad (5-2)$$

在上式中，对 $K = 1, 2, 3$ 的情况，本文均设置 $\alpha_0 = 1$, $\alpha_1 = 1$, $\gamma = 1$ 。对于 $K = 1$ 的情形，设置 $\beta_1 = -3$, $\delta = 0.5$ ；对于 $K = 2$ 的情形，设置 $\beta_1 = -3$, $\beta_2 = 4$ 且 $\delta_1 = -1$, $\delta_2 = 2$ ；对于 $K = 3$ 的情形，设置 $\beta_1 = -3$, $\beta_2 = 4$, $\beta_3 = -4$ 且 $\delta_1 = -3$, $\delta_2 = 0$, $\delta_3 = 3$ 。而对于随机误差项 e_i ，还是基于方差性与服从的分布，分成之前的 6 种不同的情况设置种子数为 1 – 1000 进行数据的生成，进行 1000 次试验，参数估计取 1000 次试验的均值。对于分位数回归模型的分位数，本人主要选用 $\tau = 0.25, 0.5, 0.75$ 进行试验，这是因为这几个分位数位置是统计学中的重要分位数位置，对这些分位数位置进行研究具有很强的代表性。模型与算法评估的指标还是参数估计的偏 (Bias)，1000 次估计结果的标准偏差 (SD)，估计的标准差 (SE) 以及估计的均方误差 (MSE)。

对于多折点分位数回归模型的折点效应的检验的模拟试验，本人将其归类于 5.1 节单折点情况下的数值模拟试验，将算法2-1与算法3-3放在一起进行了比较。除此之外，本人针对 $K = 1, 2, 3$ 情形下的折点数估计算法3-2进行了模拟试验。具体方法为基于折点数 $K = 1, 2, 3$ 以及 6 种不同的随机误差项情况生成 $n = 1000$ 的样本，针对该样本估计模型的折点数，进行 $N = 1000$ 次试验，计算算法3-2的折点数估计 \hat{K} 的正确率。对于(3-7)式中的 G_n 项，分别采用 $G_n = 1$, $G_n = \log(\log(n))$, $G_n = \log(n)$ 这 3 种不同的情况，其中对于 G_n 的选择均是满足3.2的要求的。这样对于不同的 G_n 项下的 sBIC

准则，可以起到试验对比的效果。

与此同时对于算法4-1和算法4-2，本文还研究了带宽 h 的选择对于这两种算法估计效果的影响。具体方法为设定带宽为 $h = n^{-d}$ 对 $K = 1$ 和 $K = 2$ 的情形用(5-2)式依据种子数 1 – 1000 生成样本量 $n = 1000$ 的 1000 组样本，计算估计结果的所有参数的 MSE 和，记为 TMSE，研究带宽变动参数 d 和 TMSE 之间的关系。本文选择 d 为 0 – 5 之间按照 0.5 的差值依次增加进行试验，随机误差项 e_i 主要服从标准正态分布和 t 分布，基于同方差、异方差两种不同情形都进行试验，绘制出了参数总均方误差 TMSE 随带宽 h 的变动图像，详见图5-2与5-3。

对于算法3-4的光滑秩得分检验算法计算折点位置的置信区间，本文主要将该方法与传统的 Wald 置信区间构造方法进行比较^[24]。通过计算两种不同方法的 95% 置信区间的覆盖率从而对两种方法进行对比。此时本文同样在 $\tau = 0.25, 0.5, 0.75$ 三个不同的分位数标准下对于 $K = 2$ 进行了试验，主要将真值 δ 的覆盖率以及置信区间长度作为评估指标。

为了方便起见，后文中 MKQR 指代 Zhong 等人^[24] 提出的 BRISQ 估计算法即算法3-1，算法改进 1 指代算法4-1，算法改进 2 指代算法4-2，算法改进 3 指代算法4-3，算法改进 4 指代算法4-4，算法改进 5 指代算法4-5。

5.2.2 数值模拟结果分析

先给出关于折点数估计算法3-2的模拟结果。在不同 K ，随机误差项服从 5.1 节中 6 种不同情形不同分位数下的 1000 次试验折点数估计正确率如下表5-4所示。

表 5-4 后项消除算法对折点数 K 估计正确率

		$K = 1$			$K = 2$			$K = 3$		
e_i	τ	1	$\log(\log(n))$	$\log(n)$	1	$\log(\log(n))$	$\log(n)$	1	$\log(\log(n))$	$\log(n)$
同方差										
N(0, 1)	0.25	89.1%	95.4%	100.0%	81.4%	95.4%	100.0%	91.5%	94.9%	98.8%
	0.5	81.4%	98.7%	99.8%	82.2%	97.6%	98.6%	90.7%	95.8%	98.6%
	0.75	91.6%	99.8%	100.0%	82.2%	97.6%	98.6%	91.0%	96.9%	98.5%
t(3)	0.25	80.3%	96.4%	100.0%	83.7%	97.2%	99.5%	91.3%	96.1%	98.8%
	0.5	83.1%	94.6%	99.7%	84.8%	92.3%	96.9%	90.1%	93.9%	95.2%
	0.75	89.8%	92.2%	99.1%	84.8%	97.5%	98.7%	91.5%	93.4%	97.2%
LA(0, 1)	0.25	91.1%	98.8%	100.0%	91.5%	99.6%	99.6%	93.6%	98.1%	99.4%
	0.5	95.2%	99.3%	100.0%	93.3%	98.9%	99.4%	95.4%	97.5%	99.0%
	0.75	94.6%	98.3%	100.0%	95.4%	98.4%	98.8%	95.9%	98.7%	99.3%
异方差										
N(0, 1)	0.25	86.7%	93.3%	99.7%	91.8%	97.4%	99.6%	90.3%	98.0%	98.3%
	0.5	90.6%	97.9%	100.0%	92.4%	99.5%	99.4%	89.6%	95.5%	98.2%
	0.75	86.5%	97.5%	99.9%	90.7%	98.9%	100.0%	89.3%	97.6%	98.6%
t(3)	0.25	88.6%	97.8%	99.3%	92.0%	98.3%	98.7%	90.3%	97.2%	98.0%
	0.5	91.9%	97.3%	100.0%	88.8%	98.0%	99.2%	91.0%	94.4%	98.4%
	0.75	88.6%	98.0%	99.5%	90.7%	96.5%	97.2%	90.5%	94.7%	95.8%
LA(0, 1)	0.25	89.1%	96.7%	98.7%	88.5%	97.6%	99.4%	91.6%	96.1%	98.4%
	0.5	90.2%	96.3%	99.0%	91.3%	94.9%	98.4%	89.9%	97.5%	97.2%
	0.75	90.8%	97.3%	99.8%	92.4%	94.7%	98.8%	93.7%	96.7%	97.3%

从表中可以看出，当 G_n 取常数 1 时，算法3-2的效果并不好，这和 G_n 取常数项时无法依据样本量调整 sBIC 准则有关；而当 $G_n = \log(n)$ 时，几乎正确率都趋近于 100% 甚至完全等于 100%。对于 $K > 1$ 情形， $G_n = \log(\log(n))$ 的估计正确率有时和 $G_n = \log(n)$ 的估计正确率接近。后续试验中主要用 $G_n = \log(n)$ 。

接下来给出 $K = 1, 2$ 时基于(5-2)式生成 $N = 1000$ 组完全一致的样本量 $n = 1000$ 的样本用算法3-1, 4-1, 4-2, 4-3, 4-4和4-5进行参数估计与方差估计试验的结果，如下表5-5至表5-40所示。

表 5-5 折点数 $K = 1$ ，同方差 $e_i \sim \mathbf{N}(0, 1)$ ， $\tau = 0.25$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	δ_1
MKQR	Bias	-0.003	0.002	0.004	0.002	-0.002
	SD	0.133	0.072	0.066	0.089	0.092
	SE	0.132	0.077	0.070	0.093	0.090
	MSE	0.017	0.006	0.005	0.009	0.008
算法改进 1	Bias	-0.030	0.006	-0.003	0.009	-0.025
	SD	0.109	0.036	0.044	0.061	0.060
	SE	0.252	0.085	0.141	0.098	0.394
	MSE	0.064	0.072	0.020	0.010	0.156
算法改进 2	Bias	0.049	0.012	-0.004	-0.001	-0.028
	SD	0.111	0.037	0.044	0.060	0.060
	SE	0.258	0.087	0.142	0.098	0.418
	MSE	0.069	0.008	0.020	0.010	0.176
算法改进 3	Bias	-0.016	-0.002	-0.000	0.007	-0.003
	SD	0.102	0.034	0.041	0.056	0.055
	SE	0.250	0.084	0.142	0.098	0.393
	MSE	0.063	0.007	0.020	0.010	0.154
算法改进 4	Bias	-0.021	-0.003	0.001	0.011	-0.003
	SD	0.106	0.037	0.041	0.057	0.060
	SE	0.251	0.084	0.142	0.098	0.392
	MSE	0.063	0.007	0.010	0.020	0.154
算法改进 5	Bias	0.010	0.002	0.005	0.031	-0.008
	SD	0.113	0.035	0.044	0.062	0.053
	SE	0.120	0.038	0.066	0.045	0.053
	MSE	0.014	0.001	0.004	0.003	0.003

着眼于两种不同方差性情况下的总体估计结果可以看到，对于同方差的情形，无论是采用哪种估计多折点分位数模型的方法，效果都是比较好的。基本上大部分估计的偏的绝对值都小于 0.010，很多参数估计的偏甚至小于 0.005，达到了 0.001 左右，只有一些算法的特殊情况下的估计的偏的绝对值达到了 0.05 左右。即使是最大的偏差也基本没有达到 0.1。而且不难发现，对于所有算法的同方差情况下的模拟结果，往往随机误差项 $e_i \sim \mathbf{N}(0, 1)$ 情况下的模拟结果的偏会更小一些，相对地估计效果更好。这是可以预料的，因为给原始模型(5-1)式与(5-2)式所加的噪声的大小往往会影响最终的估计结果，所加噪声越大，使得数据本身受到外部影响越大从而覆盖了数据真实信息，对于模型识别与估计存在很大的干扰，因而估计效果往往会越差。而对于这几种不同分布的随机误差噪声项，由于分布本身的重尾性，往往 $e_i \sim \mathbf{t}(3)$ 时被认为所加噪声最大，自然估计效果也就没有 $e_i \sim \mathbf{N}(0, 1)$ 时的好。而对于 $e_i \sim \mathbf{LA}(0, 1)$ 的情况，因为拉普拉斯分布的密度函数形式与分位数回归的损失函数存在一定的相似性，所以相对地对 $e_i \sim \mathbf{LA}(0, 1)$ 进行估计时的效果也会稍好一些。

表 5-6 折点数 $K = 1$, 同方差 $e_i \sim \mathbf{t}(3)$, $\tau = 0.25$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	δ_1
MKQR	Bias	0.004	0.012	-0.003	0.013	0.007
	SD	0.106	0.132	0.089	0.148	0.162
	SE	0.108	0.134	0.093	0.143	0.160
	MSE	0.012	0.018	0.009	0.021	0.026
算法改进 1	Bias	0.024	0.005	-0.006	0.025	-0.036
	SD	0.141	0.048	0.054	0.074	0.077
	SE	0.316	0.107	0.177	0.123	0.493
	MSE	0.101	0.011	0.031	0.016	0.244
算法改进 2	Bias	0.049	0.014	-0.002	-0.003	-0.029
	SD	0.144	0.047	0.052	0.077	0.076
	SE	0.330	0.110	0.179	0.123	0.544
	MSE	0.111	0.012	0.032	0.015	0.297
算法改进 3	Bias	-0.032	-0.008	-0.005	0.030	-0.011
	SD	0.143	0.070	0.075	0.177	0.077
	SE	0.313	0.106	0.177	0.123	0.491
	MSE	0.099	0.011	0.031	0.016	0.241
算法改进 4	Bias	-0.042	-0.011	-0.007	0.032	-0.006
	SD	0.137	0.070	0.073	0.178	0.080
	SE	0.307	0.104	0.175	0.121	0.483
	MSE	0.096	0.011	0.031	0.016	0.233
算法改进 5	Bias	0.001	0.001	-0.002	0.035	-0.017
	SD	0.144	0.049	0.045	0.083	0.079
	SE	0.145	0.047	0.045	0.080	0.072
	MSE	0.021	0.002	0.006	0.004	0.005

表 5-7 折点数 $K = 1$, 同方差 $e_i \sim \mathbf{LA}(0, 1)$, $\tau = 0.25$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	δ_1
MKQR	Bias	0.033	0.012	0.007	-0.008	-0.007
	SD	0.123	0.046	0.048	0.072	0.074
	SE	0.127	0.068	0.051	0.077	0.072
	MSE	0.017	0.005	0.003	0.006	0.005
算法改进 1	Bias	0.023	0.006	-0.003	0.020	-0.035
	SD	0.138	0.046	0.056	0.077	0.077
	SE	0.321	0.108	0.181	0.125	0.500
	MSE	0.104	0.012	0.033	0.016	0.251
算法改进 2	Bias	0.046	0.013	-0.001	0.000	-0.031
	SD	0.139	0.046	0.056	0.078	0.077
	SE	0.333	0.112	0.183	0.125	0.544
	MSE	0.113	0.013	0.033	0.016	0.296
算法改进 3	Bias	-0.030	-0.003	0.001	0.014	-0.008
	SD	0.131	0.043	0.052	0.071	0.072
	SE	0.319	0.108	0.182	0.125	0.498
	MSE	0.103	0.012	0.033	0.016	0.248
算法改进 4	Bias	-0.041	-0.006	0.003	0.018	-0.006
	SD	0.134	0.044	0.052	0.072	0.078
	SE	0.318	0.107	0.181	0.125	0.496
	MSE	0.103	0.012	0.033	0.016	0.246
算法改进 5	Bias	0.002	0.008	0.000	0.026	-0.010
	SD	0.140	0.050	0.052	0.080	0.073
	SE	0.141	0.046	0.079	0.054	0.072
	MSE	0.020	0.002	0.006	0.004	0.005

然而对于不同的分位数而言, 对于中位数回归 ($\tau = 0.5$) 的情况, 估计效果要稍好于 1/4 分位数回归与 3/4 分位数回归。以算法改进 5 的 $K = 2$ 时 $e_i \sim \mathbf{t}(3)$ 情况为例, 对于 $\tau = 0.5$ 分位数位置, 第一个折点系数 β_1 的估计的偏最大, 达到了 0.028, 但还是远

表 5-8 折点数 $K = 1$ ，同方差 $e_i \sim N(0, 1)$ ， $\tau = 0.5$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	δ_1
MKQR	Bias	0.003	-0.002	0.000	-0.004	0.002
	SD	0.146	0.059	0.050	0.081	0.081
	SE	0.144	0.056	0.048	0.081	0.077
	MSE	0.021	0.003	0.002	0.007	0.006
算法改进 1	Bias	0.033	0.008	-0.006	0.005	-0.022
	SD	0.107	0.035	0.041	0.055	0.057
	SE	0.203	0.068	0.114	0.079	0.018
	MSE	0.042	0.005	0.013	0.006	0.002
算法改进 2	Bias	0.054	0.014	-0.006	-0.007	-0.024
	SD	0.10	0.033	0.040	0.054	0.054
	SE	0.206	0.069	0.113	0.079	0.335
	MSE	0.043	0.005	0.013	0.006	0.113
算法改进 3	Bias	-0.001	-0.001	-0.003	0.005	-0.003
	SD	0.098	0.032	0.039	0.051	0.051
	SE	0.201	0.067	0.114	0.079	0.317
	MSE	0.041	0.005	0.013	0.006	0.100
算法改进 4	Bias	-0.009	-0.003	0.001	0.011	-0.005
	SD	0.098	0.033	0.036	0.054	0.059
	SE	0.202	0.068	0.114	0.079	0.317
	MSE	0.041	0.005	0.013	0.006	0.101
算法改进 5	Bias	-0.005	-0.002	0.001	0.032	0.001
	SD	0.101	0.030	0.040	0.054	0.054
	SE	0.112	0.036	0.061	0.042	0.050
	MSE	0.013	0.001	0.004	0.003	0.003

表 5-9 折点数 $K = 1$ ，同方差 $e_i \sim t(3)$ ， $\tau = 0.5$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	δ_1
MKQR	Bias	-0.001	0.000	0.001	-0.005	0.002
	SD	0.157	0.061	0.053	0.084	0.087
	SE	0.159	0.061	0.053	0.084	0.087
	MSE	0.025	0.004	0.003	0.008	0.007
算法改进 1	Bias	0.012	0.001	-0.004	0.023	-0.027
	SD	0.118	0.039	0.043	0.063	0.063
	SE	0.221	0.074	0.123	0.087	0.345
	MSE	0.049	0.006	0.015	0.008	0.120
算法改进 2	Bias	0.046	0.013	-0.004	-0.007	-0.022
	SD	0.108	0.036	0.043	0.060	0.059
	SE	0.229	0.076	0.124	0.086	0.378
	MSE	0.054	0.006	0.015	0.007	0.144
算法改进 3	Bias	-0.033	-0.013	-0.002	0.034	-0.006
	SD	0.123	0.065	0.068	0.174	0.065
	SE	0.219	0.074	0.123	0.086	0.342
	MSE	0.049	0.006	0.015	0.009	0.117
算法改进 4	Bias	-0.038	-0.010	0.001	0.027	-0.003
	SD	0.125	0.051	0.051	0.111	0.077
	SE	0.219	0.073	0.124	0.086	0.343
	MSE	0.049	0.005	0.015	0.008	0.118
算法改进 5	Bias	-0.001	0.002	-0.002	0.030	-0.007
	SD	0.127	0.038	0.039	0.069	0.058
	SE	0.124	0.039	0.068	0.047	0.050
	MSE	0.015	0.002	0.005	0.003	0.003

小于 $\tau = 0.75$ 情况下的最大偏值 0.057，这二者的偏差了一倍之多。因为从数据本身看来，处于中位数附近的数据应该是最密集的，往往可以更好地反映数据的本质信息，因而估计效果较好。对于极端位置的分位数回归模拟试验，经常通过加大模拟试验的样

表 5-10 折点数 $K = 1$, 同方差 $e_i \sim \mathbf{LA}(0, 1)$, $\tau = 0.5$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	δ_1
MKQR	Bias	0.007	0.001	0.002	0.001	0.001
	SD	0.131	0.071	0.061	0.080	0.057
	SE	0.136	0.075	0.063	0.088	0.052
	MSE	0.017	0.002	0.010	0.004	0.006
算法改进 1	Bias	0.020	0.004	-0.005	0.013	-0.023
	SD	0.089	0.029	0.033	0.049	0.047
	SE	0.178	0.060	0.100	0.070	0.278
	MSE	0.032	0.004	0.010	0.005	0.078
算法改进 2	Bias	0.045	0.012	-0.003	-0.008	-0.019
	SD	0.089	0.029	0.033	0.049	0.047
	SE	0.184	0.061	0.101	0.070	0.302
	MSE	0.036	0.004	0.010	0.005	0.092
算法改进 3	Bias	-0.013	-0.004	-0.000	0.015	-0.006
	SD	0.086	0.029	0.033	0.048	0.046
	SE	0.177	0.059	0.100	0.070	0.277
	MSE	0.031	0.004	0.010	0.005	0.076
算法改进 4	Bias	-0.023	-0.007	0.001	0.020	-0.006
	SD	0.111	0.036	0.033	0.050	0.078
	SE	0.180	0.060	0.101	0.070	0.281
	MSE	0.033	0.004	0.010	0.005	0.079
算法改进 5	Bias	0.006	0.001	-0.004	0.021	-0.004
	SD	0.091	0.030	0.028	0.053	0.049
	SE	0.117	0.035	0.041	0.060	0.049
	MSE	0.013	0.001	0.004	0.002	0.002

表 5-11 折点数 $K = 1$, 同方差 $e_i \sim \mathbf{N}(0, 1)$, $\tau = 0.75$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	δ_1
MKQR	Bias	0.001	0.004	-0.003	-0.004	0.002
	SD	0.151	0.070	0.062	0.084	0.095
	SE	0.147	0.073	0.066	0.082	0.092
	MSE	0.022	0.005	0.004	0.006	0.008
算法改进 1	Bias	0.037	0.010	-0.004	0.008	-0.030
	SD	0.117	0.038	0.044	0.062	0.062
	SE	0.250	0.084	0.141	0.098	0.392
	MSE	0.064	0.007	0.020	0.010	0.154
算法改进 2	Bias	0.052	0.014	-0.006	-0.005	-0.026
	SD	0.110	0.037	0.041	0.062	0.060
	SE	0.259	0.086	0.142	0.098	0.421
	MSE	0.070	0.008	0.020	0.010	0.178
算法改进 3	Bias	0.006	-0.001	-0.000	0.007	-0.005
	SD	0.107	0.035	0.041	0.058	0.058
	SE	0.249	0.084	0.142	0.098	0.491
	MSE	0.062	0.007	0.020	0.010	0.153
算法改进 4	Bias	-0.003	-0.002	0.002	0.011	-0.003
	SD	0.105	0.34	0.040	0.055	0.060
	SE	0.252	0.084	0.142	0.098	0.395
	MSE	0.064	0.007	0.020	0.010	0.156
算法改进 5	Bias	-0.007	-0.001	0.005	0.031	0.002
	SD	0.117	0.038	0.050	0.059	0.064
	SE	0.123	0.039	0.066	0.045	0.060
	MSE	0.015	0.002	0.004	0.003	0.004

本量以提高试验结果的估计精度, 通过数据的大样本统计性质达到估计的稳健性这一目的。在这里由于计算工具受限以及时间和文章篇幅的原因, 不做过多赘述。但是事实上, 对于以上的所有算法, 无论在何种情况下, 无论对应分位数 τ 取多少, 估计的偏的

表 5-12 折点数 $K = 1$ ，同方差 $e_i \sim t(3)$ ， $\tau = 0.75$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	δ_1
MKQR	Bias	0.013	0.010	-0.009	-0.011	0.014
	SD	0.090	0.070	0.062	0.084	0.075
	SE	0.093	0.077	0.066	0.088	0.073
	MSE	0.008	0.006	0.004	0.008	0.006
算法改进 1	Bias	0.031	0.006	-0.008	0.020	-0.035
	SD	0.313	0.106	0.055	0.079	0.077
	SE	0.125	0.169	0.176	0.122	0.492
	MSE	0.099	0.011	0.031	0.015	0.243
算法改进 2	Bias	0.065	0.018	-0.002	-0.008	-0.033
	SD	0.145	0.47	0.053	0.079	0.076
	SE	0.328	0.109	0.178	0.122	0.542
	MSE	0.112	0.012	0.032	0.015	0.295
算法改进 3	Bias	-0.005	-0.008	-0.005	-0.026	-0.007
	SD	0.142	0.64	0.068	0.153	0.076
	SE	0.310	0.105	0.176	0.122	0.489
	MSE	0.096	0.011	0.031	0.016	0.239
算法改进 4	Bias	-0.014	-0.009	-0.000	0.029	-0.008
	SD	0.135	0.055	0.059	0.119	0.076
	SE	0.314	0.105	0.177	0.122	0.492
	MSE	0.099	0.011	0.031	0.016	0.242
算法改进 5	Bias	0.005	-0.001	-0.008	0.037	-0.009
	SD	0.159	0.053	0.049	0.083	0.071
	SE	0.148	0.047	0.081	0.055	0.069
	MSE	0.022	0.002	0.007	0.004	0.005

表 5-13 折点数 $K = 1$ ，同方差 $e_i \sim \mathbf{LA}(0,1)$ ， $\tau = 0.75$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	δ_1
MKQR	Bias	0.017	-0.003	-0.004	0.056	0.063
	SD	0.102	0.033	0.030	0.097	0.081
	SE	0.111	0.041	0.042	0.099	0.080
	MSE	0.013	0.002	0.002	0.013	0.010
算法改进 1	Bias	0.034	0.007	-0.004	0.017	-0.035
	SD	0.141	0.047	0.053	0.076	0.074
	SE	0.322	0.108	0.181	0.125	0.502
	MSE	0.105	0.012	0.033	0.016	0.254
算法改进 2	Bias	0.056	0.014	-0.002	-0.003	-0.030
	SD	0.140	0.047	0.053	0.075	0.075
	SE	0.334	0.112	0.183	0.125	0.546
	MSE	0.115	0.013	0.033	0.016	0.299
算法改进 3	Bias	0.005	-0.004	0.000	0.015	-0.007
	SD	0.129	0.044	0.049	0.071	0.070
	SE	0.320	0.108	0.181	0.126	0.501
	MSE	0.102	0.012	0.033	0.016	0.251
算法改进 4	Bias	-0.005	-0.006	0.002	0.020	-0.006
	SD	0.135	0.045	0.049	0.071	0.078
	SE	0.323	0.108	0.182	0.126	0.506
	MSE	0.104	0.012	0.033	0.016	0.256
算法改进 5	Bias	0.012	0.001	-0.002	0.036	-0.015
	SD	0.131	0.052	0.048	0.080	0.082
	SE	0.143	0.048	0.055	0.080	0.080
	MSE	0.021	0.002	0.006	0.004	0.007

绝对值很少有大于 0.1 的情况，这说明以上的方法估计效果总体都是较好的。

而对于异方差情况下的估计结果也是较为准确的。此时的数据噪声和波动比同方差情形下更大，但是估计的偏仍旧较小。在所有的估计结果中，有偏的绝对值小于 0.001

表 5-14 折点数 $K = 1$, 异方差 $e_i \sim N(0, 1)$, $\tau = 0.25$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	δ_1
MKQR	Bias	0.005	0.001	0.002	-0.011	0.008
	SD	0.188	0.040	0.092	0.111	0.146
	SE	0.184	0.042	0.096	0.108	0.149
	MSE	0.034	0.002	0.008	0.012	0.022
算法改进 1	Bias	0.030	0.007	-0.002	0.030	-0.043
	SD	0.074	0.016	0.018	0.077	0.066
	SE	0.173	0.038	0.177	0.038	0.421
	MSE	0.031	0.001	0.031	0.002	0.179
算法改进 2	Bias	0.034	0.008	-0.002	0.023	-0.038
	SD	0.069	0.015	0.017	0.079	0.065
	SE	0.174	0.038	0.182	0.039	0.447
	MSE	0.032	0.002	0.033	0.002	0.201
算法改进 3	Bias	0.015	0.018	-0.001	-0.005	-0.022
	SD	0.076	0.018	0.024	0.074	0.062
	SE	0.173	0.038	0.179	0.039	0.424
	MSE	0.030	0.002	0.032	0.002	0.180
算法改进 4	Bias	-0.007	0.012	0.001	-0.011	-0.004
	SD	0.079	0.018	0.025	0.079	0.066
	SE	0.172	0.037	0.179	0.039	0.425
	MSE	0.030	0.002	0.032	0.002	0.181
算法改进 5	Bias	0.019	0.007	-0.003	0.038	-0.006
	SD	0.076	0.016	0.020	0.0793	0.058
	SE	0.081	0.018	0.018	0.082	0.056
	MSE	0.007	0.000	0.007	0.002	0.003

表 5-15 折点数 $K = 1$, 异方差 $e_i \sim t(3)$, $\tau = 0.25$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	δ_1
MKQR	Bias	0.021	0.007	-0.004	0.060	-0.068
	SD	0.099	0.025	0.024	0.103	0.092
	SE	0.102	0.031	0.027	0.106	0.098
	MSE	0.011	0.001	0.001	0.014	0.014
算法改进 1	Bias	0.022	0.005	-0.003	0.054	-0.061
	SD	0.093	0.020	0.022	0.096	0.084
	SE	0.217	0.047	0.219	0.049	0.523
	MSE	0.047	0.002	0.048	0.005	0.277
算法改进 2	Bias	0.043	0.011	-0.002	0.026	-0.048
	SD	0.089	0.020	0.024	0.095	0.081
	SE	0.220	0.048	0.228	0.049	0.574
	MSE	0.050	0.002	0.052	0.003	0.332
算法改进 3	Bias	-0.010	0.010	-0.004	0.024	-0.031
	SD	0.110	0.060	0.062	0.188	0.082
	SE	0.215	0.047	0.222	0.049	0.525
	MSE	0.046	0.002	0.049	0.003	0.276
算法改进 4	Bias	-0.038	0.001	-0.006	0.020	-0.009
	SD	0.118	0.074	0.076	0.023	0.084
	SE	0.214	0.047	0.222	0.049	0.525
	MSE	0.047	0.002	0.049	0.003	0.276
算法改进 5	Bias	0.012	0.006	-0.009	0.052	-0.020
	SD	0.092	0.020	0.023	0.096	0.083
	SE	0.098	0.022	0.100	0.023	0.077
	MSE	0.010	0.001	0.010	0.003	0.006

的, 而偏的绝对值的最大值也没有超过 0.1。尽管估计的偏较小, 但是异方差情形下的估计效果总体而言还是不及同方差情形, 这和数据噪声有关系, 属于正常现象。总体而言, 还是在 $e_i \sim N(0, 1)$ 情况下的估计的偏最小。而对于不同的分位数情形, 估计效果

表 5-16 折点数 $K = 1$ ，异方差 $e_i \sim \mathbf{LA}(0, 1)$ ， $\tau = 0.25$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	δ_1
MKQR	Bias	0.036	0.008	0.004	-0.020	-0.030
	SD	0.144	0.051	0.055	0.081	0.072
	SE	0.149	0.072	0.077	0.080	0.091
	MSE	0.023	0.005	0.006	0.007	0.009
算法改进 1	Bias	0.023	0.006	-0.001	0.044	-0.055
	SD	0.090	0.019	0.023	0.100	0.085
	SE	0.221	0.048	0.226	0.051	0.534
	MSE	0.049	0.002	0.051	0.005	0.288
算法改进 2	Bias	0.040	0.010	-0.001	0.025	-0.046
	SD	0.090	0.019	0.023	0.102	0.085
	SE	0.223	0.049	0.234	0.051	0.579
	MSE	0.052	0.002	0.055	0.003	0.337
算法改进 3	Bias	-0.008	0.014	-0.000	0.003	-0.023
	SD	0.092	0.022	0.029	0.096	0.079
	SE	0.220	0.048	0.229	0.050	0.537
	MSE	0.048	0.003	0.052	0.003	0.289
算法改进 4	Bias	-0.028	0.010	0.002	-0.002	-0.007
	SD	0.095	0.022	0.029	0.098	0.084
	SE	0.219	0.048	0.230	0.050	0.537
	MSE	0.049	0.002	0.052	0.002	0.288
算法改进 5	Bias	0.012	0.007	-0.011	0.047	-0.009
	SD	0.090	0.021	0.023	0.093	0.077
	SE	0.098	0.022	0.024	0.099	0.076
	MSE	0.0489	0.002	0.052	0.002	0.006

表 5-17 折点数 $K = 1$ ，异方差 $e_i \sim \mathbf{N}(0, 1)$ ， $\tau = 0.5$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	δ_1
MKQR	Bias	0.004	0.000	0.001	0.000	0.009
	SD	0.099	0.026	0.022	0.106	0.090
	SE	0.099	0.023	0.022	0.102	0.082
	MSE	0.010	0.001	0.010	0.001	0.007
算法改进 1	Bias	0.024	0.005	-0.001	0.025	-0.034
	SD	0.070	0.015	0.017	0.070	0.060
	SE	0.137	0.029	0.141	0.030	0.336
	MSE	0.019	0.001	0.020	0.001	0.114
算法改进 2	Bias	0.033	0.007	-0.002	0.019	-0.033
	SD	0.069	0.015	0.017	0.073	0.063
	SE	0.138	0.030	0.144	0.030	0.356
	MSE	0.020	0.001	0.021	0.001	0.127
算法改进 3	Bias	0.002	0.000	-0.000	0.007	-0.007
	SD	0.072	0.017	0.022	0.068	0.057
	SE	0.137	0.029	0.143	0.030	0.338
	MSE	0.019	0.001	0.020	0.001	0.114
算法改进 4	Bias	-0.007	-0.002	-0.000	0.004	-0.004
	SD	0.080	0.019	0.022	0.068	0.064
	SE	0.138	0.030	0.144	0.030	0.342
	MSE	0.019	0.001	0.021	0.001	0.117
算法改进 5	Bias	-0.000	0.000	0.001	0.041	-0.001
	SD	0.066	0.014	0.018	0.067	0.060
	SE	0.076	0.017	0.076	0.017	0.062
	MSE	0.006	0.000	0.006	0.002	0.004

尽管存在偏差，但是参数估计效果确实很好。这说明了以上算法的普适性是较高的。

接下来，本文基于不同的算法进行分析。尽管所有的算法在各种各样的情形下对于参数的估计效果都不错，但是各种算法之间还是存在着或多或少的效果以及效率上

表 5-18 折点数 $K = 1$, 异方差 $e_i \sim \mathbf{t}(3)$, $\tau = 0.5$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	δ_1
MKQR	Bias	0.000	0.000	0.000	-0.001	-0.001
	SD	0.103	0.027	0.023	0.111	0.091
	SE	0.109	0.026	0.024	0.114	0.091
	MSE	0.012	0.001	0.001	0.013	0.008
算法改进 1	Bias	0.010	0.002	-0.002	0.047	-0.048
	SD	0.077	0.017	0.016	0.082	0.067
	SE	0.149	0.032	0.153	0.032	0.363
	MSE	0.022	0.001	0.023	0.003	0.134
算法改进 2	Bias	0.036	0.008	-0.002	0.012	-0.028
	SD	0.075	0.016	0.017	0.081	0.066
	SE	0.151	0.033	0.160	0.033	0.400
	MSE	0.024	0.001	0.026	0.001	0.161
算法改进 3	Bias	-0.023	-0.012	-0.009	0.053	-0.023
	SD	0.118	0.091	0.092	0.277	0.079
	SE	0.147	0.032	0.153	0.032	0.363
	MSE	0.022	0.001	0.024	0.004	0.132
算法改进 4	Bias	-0.036	-0.015	-0.008	0.053	-0.015
	SD	0.124	0.091	0.092	0.275	0.088
	SE	0.149	0.032	0.155	0.033	0.366
	MSE	0.023	0.001	0.024	0.004	0.134
算法改进 5	Bias	-0.013	-0.006	0.001	0.055	-0.005
	SD	0.099	0.022	0.023	0.092	0.073
	SE	0.105	0.023	0.102	0.024	0.070
	MSE	0.011	0.001	0.010	0.004	0.005

表 5-19 折点数 $K = 1$, 异方差 $e_i \sim \mathbf{LA}(0, 1)$, $\tau = 0.5$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	δ_1
MKQR	Bias	0.000	0.001	0.002	0.004	0.001
	SD	0.070	0.066	0.082	0.023	0.031
	SE	0.074	0.065	0.079	0.022	0.033
	MSE	0.013	0.001	0.010	0.003	0.005
算法改进 1	Bias	0.012	0.002	-0.002	0.029	-0.033
	SD	0.058	0.013	0.014	0.062	0.050
	SE	0.121	0.026	0.125	0.027	0.300
	MSE	0.015	0.001	0.016	0.002	0.089
算法改进 2	Bias	0.031	0.007	-0.001	0.010	-0.025
	SD	0.057	0.012	0.014	0.064	0.051
	SE	0.123	0.027	0.129	0.027	0.320
	MSE	0.016	0.001	0.017	0.001	0.103
算法改进 3	Bias	-0.011	-0.003	0.000	0.017	-0.001
	SD	0.065	0.015	0.021	0.062	0.050
	SE	0.121	0.026	0.126	0.027	0.296
	MSE	0.015	0.001	0.016	0.001	0.088
算法改进 4	Bias	-0.021	-0.005	0.001	0.019	-0.007
	SD	0.086	0.021	0.021	0.072	0.084
	SE	0.123	0.027	0.127	0.027	0.301
	MSE	0.015	0.001	0.016	0.001	0.091
算法改进 5	Bias	0.001	-0.000	-0.001	0.026	-0.001
	SD	0.060	0.014	0.014	0.063	0.052
	SE	0.080	0.175	0.165	0.073	0.051
	MSE	0.006	0.000	0.005	0.001	0.003

的差异。为了更好地综合不同算法的优缺点，本文给出了各算法在不同情形下的一次估计所用时间表，见表5-41所示。

在这部分我们先着眼于表5-5到表5-40的模拟结果。可以看出 Zhong 等人^[24]提出的

表 5-20 折点数 $K = 1$ ，异方差 $e_i \sim N(0, 1)$ ， $\tau = 0.75$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	δ_1
MKQR	Bias	-0.010	0.008	0.023	0.003	-0.009
	SD	0.123	0.168	0.437	0.419	0.267
	SE	0.125	0.169	0.435	0.421	0.258
	MSE	0.016	0.029	0.190	0.178	0.067
算法改进 1	Bias	0.027	0.005	-0.002	0.029	-0.040
	SD	0.076	0.017	0.017	0.076	0.064
	SE	0.173	0.037	0.178	0.038	0.423
	MSE	0.031	0.001	0.032	0.002	0.181
算法改进 2	Bias	0.037	0.007	-0.003	0.021	-0.038
	SD	0.074	0.016	0.017	0.077	0.064
	SE	0.175	0.038	0.181	0.038	0.448
	MSE	0.032	0.001	0.032	0.002	0.202
算法改进 3	Bias	-0.002	-0.015	-0.002	0.021	-0.002
	SD	0.077	0.018	0.023	0.074	0.060
	SE	0.173	0.038	0.180	0.038	0.428
	MSE	0.030	0.002	0.033	0.002	0.183
算法改进 4	Bias	-0.009	-0.016	0.000	0.028	-0.006
	SD	0.086	0.021	0.024	0.073	0.066
	SE	0.173	0.038	0.180	0.039	0.427
	MSE	0.030	0.002	0.030	0.002	0.182
算法改进 5	Bias	-0.013	-0.004	0.007	0.047	0.003
	SD	0.078	0.017	0.024	0.068	0.069
	SE	0.086	0.019	0.083	0.020	0.066
	MSE	0.007	0.000	0.007	0.003	0.004

表 5-21 折点数 $K = 1$ ，异方差 $e_i \sim t(3)$ ， $\tau = 0.75$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	δ_1
MKQR	Bias	0.018	0.004	-0.004	0.060	-0.055
	SD	0.109	0.044	0.027	0.098	0.085
	SE	0.114	0.048	0.030	0.102	0.082
	MSE	0.013	0.002	0.001	0.014	0.010
算法改进 1	Bias	0.016	0.002	-0.003	0.053	-0.052
	SD	0.099	0.021	0.023	0.092	0.079
	SE	0.214	0.047	0.022	0.049	0.522
	MSE	0.046	0.002	0.048	0.005	0.275
算法改进 2	Bias	0.043	0.008	-0.002	0.024	-0.043
	SD	0.092	0.020	0.022	0.097	0.076
	SE	0.218	0.047	0.229	0.048	0.574
	MSE	0.049	0.002	0.052	0.003	0.331
算法改进 3	Bias	-0.023	-0.026	-0.007	0.059	-0.011
	SD	0.125	0.079	0.083	0.247	0.086
	SE	0.211	0.046	0.222	0.049	0.522
	MSE	0.045	0.003	0.050	0.006	0.273
算法改进 4	Bias	-0.031	-0.028	-0.006	0.063	-0.010
	SD	0.131	0.085	0.088	0.264	0.091
	SE	0.215	0.047	0.225	0.049	0.530
	MSE	0.047	0.003	0.051	0.006	0.281
算法改进 5	Bias	-0.013	-0.006	0.001	0.055	-0.005
	SD	0.099	0.022	0.023	0.092	0.073
	SE	0.105	0.023	0.102	0.024	0.070
	MSE	0.011	0.001	0.010	0.004	0.005

MKQR 方法的总体估计效果确实较好，不但参数估计的偏差很小，而且对于方差估计的效果也很好。可以看出，在 MKQR 的模拟结果记录表中，各参数实际估计的标准差 (SD) 与理论参数标准差 (SE) 的差异都非常小。这一现象可以说明两点：(a) Zhong 等

表 5-22 折点数 $K = 1$, 异方差 $e_i \sim \mathbf{LA}(0, 1)$, $\tau = 0.75$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	δ_1
MKQR	Bias	0.032	0.036	0.002	-0.005	-0.042
	SD	0.118	0.047	0.250	0.104	0.082
	SE	0.133	0.052	0.236	0.111	0.096
	MSE	0.019	0.004	0.056	0.012	0.011
算法改进 1	Bias	0.022	0.003	-0.002	0.046	-0.051
	SD	0.094	0.020	0.021	0.094	0.078
	SE	0.220	0.048	0.225	0.049	0.534
	MSE	0.049	0.002	0.051	0.004	0.288
算法改进 2	Bias	0.039	0.007	-0.001	0.028	-0.042
	SD	0.093	0.020	0.020	0.095	0.079
	SE	0.223	0.048	0.233	0.049	0.579
	MSE	0.051	0.002	0.054	0.003	0.337
算法改进 3	Bias	-0.006	-0.018	-0.000	0.031	-0.005
	SD	0.094	0.022	0.027	0.093	0.076
	SE	0.219	0.048	0.229	0.049	0.539
	MSE	0.048	0.003	0.052	0.003	0.290
算法改进 4	Bias	-0.012	-0.019	0.001	0.036	0.084
	SD	0.107	0.026	0.027	0.091	0.084
	SE	0.222	0.048	0.230	0.049	0.544
	MSE	0.049	0.003	0.053	0.004	0.303
算法改进 5	Bias	-0.004	-0.004	0.007	0.046	-0.010
	SD	0.087	0.020	0.022	0.095	0.091
	SE	0.102	0.048	0.026	0.104	0.086
	MSE	0.010	0.001	0.011	0.003	0.007

表 5-23 折点数 $K = 2$, 同方差 $e_i \sim \mathbf{N}(0, 1)$, $\tau = 0.25$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
MKQR	Bias	0.002	0.000	0.000	-0.006	0.003	0.002	-0.004
	SD	0.212	0.042	0.063	0.109	0.129	0.077	0.058
	SE	0.195	0.042	0.059	0.109	0.129	0.070	0.056
	MSE	0.038	0.017	0.035	0.012	0.017	0.005	0.003
算法改进 1	Bias	0.029	0.006	0.000	0.131	-0.216	-0.083	0.018
	SD	0.222	0.066	0.044	0.110	0.132	0.078	0.054
	SE	0.457	0.139	0.240	0.286	0.096	0.464	0.172
	MSE	0.210	0.019	0.057	0.099	0.056	0.222	0.030
算法改进 2	Bias	0.041	0.007	-0.004	0.111	-0.171	-0.077	0.022
	SD	0.232	0.067	0.043	0.113	0.133	0.080	0.052
	SE	0.476	0.144	0.251	0.298	0.097	0.514	0.191
	MSE	0.229	0.021	0.063	0.101	0.039	0.270	0.037
算法改进 3	Bias	-0.014	-0.001	0.001	0.047	-0.075	-0.024	0.004
	SD	0.191	0.058	0.042	0.105	0.128	0.070	0.053
	SE	0.446	0.136	0.245	0.291	0.097	0.047	0.170
	MSE	0.199	0.018	0.060	0.087	0.015	0.218	0.029
算法改进 4	Bias	-0.037	-0.007	0.000	0.029	-0.041	-0.006	-0.002
	SD	0.188	0.057	0.042	0.107	0.124	0.076	0.061
	SE	0.441	0.136	0.245	0.292	0.097	0.463	0.168
	MSE	0.196	0.018	0.060	0.086	0.011	0.215	0.028
算法改进 5	Bias	0.007	0.006	0.013	0.036	-0.109	-0.005	-0.001
	SD	0.211	0.060	0.041	0.100	0.116	0.076	0.053
	SE	0.226	0.064	0.047	0.101	0.118	0.079	0.055
	MSE	0.051	0.004	0.002	0.011	0.026	0.006	0.003

人^[24]提出的 MKQR 可以很好地执行模型中参数的方差标准差估计的任务；(b) 此时的参数估计结果更符合参数具有渐近正态性这一预期。因为估计与实际的差异很小，也说明该方法更好地刻画了参数估计的有效性，进而保证了各参数均方误差 (MSE) 计算

表 5-24 折点数 $K = 2$ ，同方差 $e_i \sim t(3)$ ， $\tau = 0.25$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
MKQR	Bias	0.012	-0.001	0.003	-0.014	0.015	0.000	-0.003
	SD	0.244	0.054	0.074	0.143	0.171	0.091	0.074
	SE	0.243	0.052	0.074	0.137	0.163	0.087	0.070
	MSE	0.059	0.003	0.006	0.019	0.027	0.008	0.005
算法改进 1	Bias	0.019	0.005	-0.003	0.186	-0.295	-0.109	0.026
	SD	0.287	0.086	0.054	0.135	0.160	0.103	0.065
	SE	0.582	0.177	0.297	0.355	0.122	0.583	0.221
	MSE	0.339	0.031	0.088	0.161	0.102	0.352	0.050
算法改进 2	Bias	0.061	0.014	0.000	0.114	-0.180	-0.091	0.027
	SD	0.291	0.085	0.053	0.133	0.159	0.098	0.065
	SE	0.614	0.185	0.320	0.379	0.122	0.674	0.256
	MSE	0.381	0.034	0.103	0.156	0.047	0.463	0.066
算法改进 3	Bias	-0.052	-0.012	-0.004	0.107	-0.146	-0.032	0.005
	SD	0.246	0.090	0.075	0.204	0.262	0.107	0.130
	SE	0.561	0.172	0.302	0.360	0.122	0.583	0.215
	MSE	0.317	0.030	0.091	0.141	0.036	0.341	0.046
算法改进 4	Bias	0.017	0.012	0.005	0.041	-0.135	-0.017	0.000
	SD	0.233	0.069	0.044	0.141	0.161	0.096	0.072
	SE	0.270	0.079	0.155	0.183	0.058	0.297	0.109
	MSE	0.073	0.006	0.024	0.035	0.022	0.088	0.012
算法改进 5	Bias	0.017	0.012	0.005	0.041	-0.135	-0.017	0.000
	SD	0.233	0.069	0.044	0.141	0.161	0.096	0.072
	SE	0.236	0.069	0.048	0.148	0.167	0.100	0.076
	MSE	0.056	0.005	0.002	0.024	0.046	0.010	0.006

表 5-25 折点数 $K = 2$ ，同方差 $e_i \sim \mathbf{LA}(0,1)$ ， $\tau = 0.25$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
MKQR	Bias	-0.003	0.003	0.002	-0.008	0.011	0.001	0.000
	SD	0.248	0.056	0.073	0.139	0.167	0.093	0.075
	SE	0.249	0.053	0.076	0.139	0.166	0.089	0.072
	MSE	0.062	0.003	0.006	0.019	0.028	0.008	0.005
算法改进 1	Bias	0.001	0.000	0.001	0.176	-0.274	-0.100	0.024
	SD	0.277	0.081	0.056	0.135	0.167	0.103	0.064
	SE	0.588	0.179	0.302	0.363	0.123	0.588	0.223
	MSE	0.346	0.032	0.091	0.163	0.091	0.356	0.050
算法改进 2	Bias	0.027	0.006	0.002	0.126	-0.192	-0.086	0.026
	SD	0.278	0.081	0.056	0.139	0.174	0.104	0.064
	SE	0.620	0.186	0.322	0.383	0.123	0.671	0.253
	MSE	0.385	0.035	0.104	0.162	0.052	0.458	0.065
算法改进 3	Bias	-0.056	-0.009	0.002	0.078	-0.107	-0.027	0.010
	SD	0.235	0.069	0.052	0.126	0.158	0.090	0.067
	SE	0.569	0.174	0.308	0.369	0.124	0.590	0.216
	MSE	0.327	0.030	0.095	0.142	0.027	0.349	0.047
算法改进 4	Bias	-0.075	-0.014	0.003	0.057	-0.076	-0.010	0.000
	SD	0.238	0.069	0.052	0.129	0.159	0.095	0.073
	SE	0.562	0.173	0.310	0.369	0.123	0.587	0.215
	MSE	0.322	0.030	0.096	0.140	0.021	0.345	0.046
算法改进 5	Bias	-0.009	-0.012	0.010	0.032	-0.132	-0.002	-0.004
	SD	0.241	0.074	0.050	0.122	0.115	0.088	0.064
	SE	0.254	0.076	0.054	0.124	0.117	0.076	0.058
	MSE	0.065	0.006	0.003	0.016	0.031	0.006	0.003

的准确性。

对于其他的改进算法，可以看出改进 1 的示性函数光滑化算法，参数估计的偏很小，但是参数实际估计的标准差 (SD) 与理论的标准差 (SE) 之间存在一定的差异。

表 5-26 折点数 $K = 2$, 同方差 $e_i \sim N(0, 1)$, $\tau = 0.5$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
MKQR	Bias	-0.005	0.001	0.000	-0.006	0.007	0.004	-0.001
	SD	0.178	0.040	0.055	0.102	0.120	0.067	0.053
	SE	0.179	0.039	0.055	0.101	0.120	0.064	0.052
	MSE	0.032	0.002	0.003	0.010	0.014	0.004	0.003
算法改进 1	Bias	0.024	0.004	-0.004	0.113	-0.181	-0.067	0.018
	SD	0.203	0.061	0.041	0.099	0.117	0.071	0.051
	SE	0.365	0.110	0.193	0.232	0.078	0.375	0.138
	MSE	0.133	0.012	0.037	0.066	0.039	0.145	0.019
算法改进 2	Bias	0.044	0.009	-0.002	0.080	-0.128	-0.059	0.019
	SD	0.208	0.063	0.040	0.106	0.120	0.071	0.051
	SE	0.379	0.114	0.201	0.237	0.078	0.412	0.150
	MSE	0.146	0.013	0.040	0.063	0.022	0.173	0.023
算法改进 3	Bias	-0.011	-0.003	-0.002	0.043	-0.062	-0.017	0.006
	SD	0.172	0.051	0.039	0.094	0.112	0.064	0.051
	SE	0.356	0.108	0.196	0.234	0.079	0.375	0.136
	MSE	0.127	0.012	0.038	0.057	0.010	0.141	0.018
算法改进 4	Bias	-0.030	-0.008	0.001	0.030	-0.041	-0.005	-0.002
	SD	0.185	0.056	0.031	0.096	0.114	0.076	0.058
	SE	0.358	0.109	0.199	0.236	0.079	0.378	0.136
	MSE	0.129	0.012	0.040	0.057	0.008	0.143	0.019
算法改进 5	Bias	-0.004	-0.002	0.001	0.025	-0.040	0.003	-0.007
	SD	0.162	0.048	0.038	0.084	0.114	0.065	0.051
	SE	0.160	0.058	0.044	0.044	0.086	0.112	0.059
	MSE	0.026	0.003	0.002	0.003	0.009	0.013	0.004

表 5-27 折点数 $K = 2$, 同方差 $e_i \sim t(3)$, $\tau = 0.5$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
MKQR	Bias	0.013	0.001	0.005	-0.012	0.003	-0.001	-0.003
	SD	0.193	0.043	0.059	0.112	0.129	0.075	0.060
	SE	0.198	0.043	0.060	0.110	0.130	0.070	0.056
	MSE	0.039	0.002	0.004	0.012	0.017	0.005	0.003
算法改进 1	Bias	0.015	0.003	-0.001	0.152	-0.246	-0.087	0.018
	SD	0.230	0.068	0.044	0.111	0.129	0.077	0.054
	SE	0.403	0.122	0.208	0.246	0.085	0.408	0.151
	MSE	0.162	0.015	0.043	0.084	0.068	0.174	0.023
算法改进 2	Bias	0.045	0.010	-0.000	0.077	-0.124	-0.062	0.019
	SD	0.228	0.068	0.043	0.107	0.131	0.078	0.054
	SE	0.425	0.128	0.225	0.264	0.085	0.473	0.175
	MSE	0.183	0.016	0.051	0.075	0.023	0.228	0.031
算法改进 3	Bias	-0.050	-0.018	-0.005	0.110	-0.153	-0.024	-0.0005
	SD	0.209	0.096	0.088	0.247	0.323	0.106	0.164
	SE	0.389	0.118	0.210	0.248	0.085	0.405	0.148
	MSE	0.154	0.014	0.044	0.074	0.031	0.165	0.022
算法改进 4	Bias	-0.087	-0.026	-0.007	0.100	-0.138	-0.001	-0.014
	SD	0.234	0.105	0.093	0.269	0.350	0.125	0.182
	SE	0.387	0.118	0.214	0.253	0.086	0.409	0.150
	MSE	0.157	0.015	0.046	0.074	0.026	0.167	0.023
算法改进 5	Bias	0.014	0.006	-0.003	0.028	-0.041	-0.007	-0.001
	SD	0.191	0.057	0.037	0.109	0.119	0.062	0.056
	SE	0.195	0.051	0.040	0.112	0.132	0.067	0.049
	MSE	0.038	0.003	0.002	0.013	0.019	0.005	0.002

尽管差异不大,基本可以控制在 0.2 以内,但是实际标准差估计相较于理论总是会有低估的情况发生。这也许会对后续的参数的置信区间构造也产生一些影响。但是该方法由于对示性函数进行了光滑化,运用了梯度下降算法,可以看到估计的计算效率大大

表 5-28 折点数 $K = 2$ ，同方差 $e_i \sim \mathbf{LA}(0, 1)$ ， $\tau = 0.5$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
MKQR	Bias	-0.006	0.001	-0.001	-0.003	0.004	0.003	-0.001
	SD	0.153	0.034	0.045	0.090	0.102	0.058	0.047
	SE	0.158	0.034	0.048	0.089	0.106	0.057	0.046
	MSE	0.025	0.001	0.002	0.008	0.011	0.003	0.002
算法改进 1	Bias	0.005	0.001	-0.001	0.125	-0.196	-0.066	0.018
	SD	0.173	0.051	0.035	0.088	0.102	0.062	0.045
	SE	0.324	0.098	0.171	0.203	0.070	0.332	0.123
	MSE	0.105	0.010	0.029	0.057	0.043	0.115	0.015
算法改进 2	Bias	0.035	0.008	0.000	0.065	-0.105	-0.049	0.016
	SD	0.170	0.050	0.034	0.092	0.109	0.063	0.046
	SE	0.339	0.102	0.182	0.215	0.070	0.377	0.139
	MSE	0.116	0.010	0.033	0.050	0.016	0.145	0.020
算法改进 3	Bias	-0.038	-0.010	0.001	0.072	-0.097	-0.021	0.007
	SD	0.153	0.045	0.033	0.085	0.099	0.057	0.045
	SE	0.315	0.096	0.172	0.205	0.069	0.330	0.121
	MSE	0.101	0.009	0.030	0.047	0.014	0.109	0.015
算法改进 4	Bias	-0.058	-0.015	0.002	0.060	-0.077	-0.010	0.000
	SD	0.203	0.057	0.034	0.102	0.118	0.095	0.073
	SE	0.319	0.097	0.177	0.210	0.071	0.337	0.124
	MSE	0.105	0.010	0.031	0.048	0.0141	0.114	0.015
算法改进 5	Bias	-0.002	-0.001	-0.004	0.021	-0.032	0.004	-0.001
	SD	0.146	0.044	0.030	0.098	0.109	0.060	0.045
	SE	0.147	0.047	0.035	0.120	0.111	0.055	0.041
	MSE	0.022	0.002	0.001	0.015	0.013	0.003	0.002

表 5-29 折点数 $K = 2$ ，同方差 $e_i \sim \mathbf{N}(0, 1)$ ， $\tau = 0.75$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
MKQR	Bias	0.008	0.001	0.003	-0.008	0.004	0.000	0.001
	SD	0.192	0.044	0.058	0.110	0.131	0.073	0.059
	SE	0.194	0.042	0.059	0.109	0.129	0.070	0.056
	MSE	0.038	0.002	0.003	0.001	0.002	0.005	0.003
算法改进 1	Bias	0.024	0.004	-0.002	0.124	-0.201	-0.074	0.019
	SD	0.232	0.069	0.043	0.111	0.133	0.076	0.054
	SE	0.461	0.140	0.240	0.286	0.097	0.486	0.170
	MSE	0.213	0.020	0.058	0.097	0.050	0.225	0.029
算法改进 2	Bias	0.063	0.014	-0.003	0.095	-0.165	-0.075	0.019
	SD	0.227	0.068	0.043	0.111	0.134	0.078	0.057
	SE	0.476	0.143	0.250	0.295	0.097	0.512	0.188
	MSE	0.230	0.021	0.062	0.096	0.037	0.267	0.036
算法改进 3	Bias	-0.002	-0.004	-0.000	0.047	-0.067	-0.020	0.006
	SD	0.198	0.059	0.040	0.102	0.127	0.071	0.053
	SE	0.448	0.137	0.244	0.290	0.097	0.467	0.168
	MSE	0.201	0.019	0.060	0.086	0.014	0.218	0.028
算法改进 4	Bias	-0.006	-0.005	-0.000	0.029	-0.041	-0.010	-0.002
	SD	0.195	0.059	0.040	0.106	0.117	0.072	0.058
	SE	0.450	0.137	0.250	0.298	0.099	0.477	0.172
	MSE	0.203	0.019	0.063	0.090	0.011	0.228	0.030
算法改进 5	Bias	0.006	-0.002	-0.002	0.039	-0.014	-0.001	-0.011
	SD	0.205	0.062	0.045	0.104	0.131	0.073	0.061
	SE	0.222	0.067	0.048	0.102	0.140	0.076	0.064
	MSE	0.049	0.004	0.002	0.012	0.020	0.006	0.004

提高了，对于 $K = 1$ 的情形一次估计甚至用不到 1 秒，效率相较于 Zhong 等人^[24] 的方法大大提高。同理还有改进 3，即利用光滑函数替代原有分位数损失函数的方法，虽然和算法改进 1 存在着类似的不足之处，但是运算效率大大提高。

表 5-30 折点数 $K = 2$, 同方差 $e_i \sim t(3)$, $\tau = 0.75$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
MKQR	Bias	0.013	-0.004	0.004	-0.009	0.004	0.001	0.000
	SD	0.247	0.051	0.074	0.140	0.156	0.089	0.074
	SE	0.244	0.052	0.074	0.137	0.162	0.088	0.070
	MSE	0.060	0.003	0.005	0.019	0.026	0.008	0.005
算法改进 1	Bias	0.041	0.010	-0.002	0.164	-0.284	-0.100	0.017
	SD	0.289	0.085	0.052	0.130	0.155	0.093	0.068
	SE	0.578	0.175	0.296	0.353	0.121	0.582	0.217
	MSE	0.335	0.031	0.088	0.152	0.096	0.349	0.047
算法改进 2	Bias	0.057	0.012	-0.002	0.105	-0.174	-0.080	0.022
	SD	0.284	0.083	0.054	0.127	0.162	0.096	0.068
	SE	0.613	0.183	0.322	0.380	0.121	0.681	0.255
	MSE	0.379	0.034	0.104	0.155	0.045	0.470	0.065
算法改进 3	Bias	-0.020	-0.012	-0.003	0.105	-0.154	-0.032	0.002
	SD	0.248	0.090	0.073	0.202	0.258	0.125	0.129
	SE	0.558	0.171	0.300	0.357	0.121	0.580	0.213
	MSE	0.311	0.029	0.090	0.139	0.038	0.337	0.045
算法改进 4	Bias	-0.064	-0.022	-0.002	0.086	-0.114	-0.009	-0.006
	SD	0.242	0.085	0.068	0.183	0.234	0.105	0.116
	SE	0.557	0.170	0.308	0.365	0.122	0.589	0.216
	MSE	0.315	0.029	0.095	0.141	0.028	0.347	0.047
算法改进 5	Bias	0.012	-0.006	-0.017	0.057	-0.022	-0.000	-0.015
	SD	0.246	0.078	0.046	0.147	0.169	0.085	0.079
	SE	0.254	0.079	0.057	0.152	0.177	0.102	0.080
	MSE	0.065	0.005	0.004	0.026	0.032	0.010	0.007

表 5-31 折点数 $K = 2$, 同方差 $e_i \sim \mathbf{LA}(0,1)$, $\tau = 0.75$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
MKQR	Bias	-0.014	0.002	-0.004	-0.005	0.008	0.007	-0.003
	SD	0.247	0.053	0.074	0.149	0.168	0.097	0.077
	SE	0.247	0.053	0.075	0.140	0.166	0.090	0.071
	MSE	0.061	0.003	0.006	0.020	0.028	0.008	0.005
算法改进 1	Bias	0.021	0.004	0.000	0.157	-0.262	-0.091	0.017
	SD	0.290	0.085	0.052	0.140	0.160	0.095	0.071
	SE	0.586	0.178	0.304	0.363	0.123	0.591	0.221
	MSE	0.344	0.032	0.092	0.156	0.084	0.358	0.049
算法改进 2	Bias	0.042	0.008	0.002	0.111	-0.188	-0.076	0.018
	SD	0.283	0.084	0.052	0.141	0.168	0.095	0.071
	SE	0.617	0.185	0.324	0.383	0.123	0.674	0.252
	MSE	0.382	0.034	0.105	0.159	0.050	0.460	0.064
算法改进 3	Bias	-0.022	-0.011	0.001	0.079	-0.112	-0.025	0.006
	SD	0.245	0.073	0.049	0.135	0.154	0.091	0.070
	SE	0.569	0.174	0.309	0.369	0.124	0.593	0.216
	MSE	0.325	0.030	0.096	0.142	0.028	0.352	0.047
算法改进 4	Bias	-0.042	-0.016	0.002	0.064	-0.085	-0.010	-0.000
	SD	0.238	0.071	0.049	0.132	0.152	0.095	0.073
	SE	0.573	0.174	0.317	0.376	0.125	0.605	0.220
	MSE	0.330	0.031	0.101	0.146	0.023	0.366	0.048
算法改进 5	Bias	-0.016	-0.014	-0.010	0.066	-0.030	0.007	-0.006
	SD	0.217	0.071	0.044	0.174	0.211	0.099	0.078
	SE	0.227	0.079	0.052	0.169	0.223	0.098	0.077
	MSE	0.052	0.006	0.003	0.033	0.051	0.010	0.006

算法改进 2 的 BentCable 结光滑化算法, 参数估计偏同样很小, 问题仍旧在于方差估计存在误差。事实上, 这一方法的方差估计存在误差是可以预料的。因为 BentCable 结本身就是在折点处用二次函数进行逼近的算法, 以至于在本质上改变了模型的形式,

表 5-32 折点数 $K = 2$ ，异方差 $e_i \sim N(0, 1)$ ， $\tau = 0.25$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
MKQR	Bias	0.006	0.000	0.002	-0.006	0.009	-0.001	-0.002
	SD	0.095	0.018	0.020	0.106	0.018	0.056	0.084
	SE	0.090	0.016	0.019	0.101	0.185	0.054	0.078
	MSE	0.008	0.000	0.000	0.010	0.034	0.003	0.006
算法改进 1	Bias	0.014	0.003	-0.001	0.150	-0.028	-0.008	0.012
	SD	0.092	0.020	0.017	0.093	0.190	0.058	0.073
	SE	0.211	0.046	0.021	0.101	0.039	0.361	0.245
	MSE	0.045	0.002	0.047	0.033	0.083	0.136	0.060
算法改进 2	Bias	0.035	0.008	-0.001	0.127	-0.254	-0.078	0.012
	SD	0.086	0.018	0.018	0.100	0.202	0.061	0.071
	SE	0.214	0.046	0.233	0.428	0.039	0.392	0.269
	MSE	0.047	0.002	0.054	0.199	0.066	0.160	0.072
算法改进 3	Bias	0.013	0.018	0.000	0.038	-0.092	-0.029	0.006
	SD	0.095	0.021	0.023	0.099	0.182	0.057	0.079
	SE	0.209	0.045	0.225	0.420	0.038	0.365	0.237
	MSE	0.044	0.002	0.051	0.178	0.100	0.134	0.056
算法改进 4	Bias	-0.006	0.013	0.001	0.004	-0.026	-0.004	0.002
	SD	0.094	0.021	0.025	0.102	0.179	0.060	0.080
	SE	0.206	0.044	0.228	0.420	0.038	0.367	0.234
	MSE	0.043	0.002	0.052	0.176	0.002	0.134	0.055
算法改进 5	Bias	0.043	0.016	-0.012	0.025	-0.097	-0.002	-0.002
	SD	0.097	0.022	0.023	0.090	0.164	0.056	0.076
	SE	0.102	0.025	0.024	0.105	0.156	0.062	0.079
	MSE	0.012	0.001	0.001	0.012	0.034	0.004	0.006

表 5-33 折点数 $K = 2$ ，异方差 $e_i \sim t(3)$ ， $\tau = 0.25$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
MKQR	Bias	0.008	0.000	0.003	-0.015	0.033	0.001	0.001
	SD	0.112	0.022	0.024	0.133	0.234	0.070	0.104
	SE	0.113	0.021	0.024	0.128	0.233	0.068	0.099
	MSE	0.013	0.000	0.001	0.017	0.055	0.005	0.010
算法改进 1	Bias	0.012	0.003	-0.002	0.194	-0.366	-0.107	0.008
	SD	0.112	0.024	0.021	0.112	0.227	0.073	0.085
	SE	0.264	0.057	0.267	0.519	0.049	0.444	0.311
	MSE	0.047	0.011	-0.001	0.130	-0.262	-0.087	0.015
算法改进 2	Bias	0.061	0.014	0.000	0.114	-0.180	-0.091	0.027
	SD	0.108	0.023	0.021	0.119	0.241	0.073	0.083
	SE	0.273	0.059	0.303	0.543	0.050	0.518	0.358
	MSE	0.077	0.004	0.092	0.312	0.071	0.276	0.129
算法改进 3	Bias	-0.022	0.008	-0.003	0.093	-0.171	-0.041	-0.001
	SD	0.126	0.061	0.062	0.199	0.307	0.089	0.145
	SE	0.260	0.056	0.275	0.518	0.048	0.450	0.299
	MSE	0.068	0.003	0.076	0.277	0.032	0.205	0.089
算法改进 4	Bias	0.045	0.018	-0.021	0.023	-0.115	-0.005	-0.003
	SD	0.107	0.023	0.025	0.131	0.221	0.076	0.099
	SE	0.129	0.028	0.280	0.517	0.048	0.453	0.294
	MSE	0.069	0.003	0.079	0.271	0.013	0.205	0.087
算法改进 5	Bias	0.045	0.018	-0.021	0.023	-0.115	-0.005	-0.003
	SD	0.107	0.023	0.025	0.131	0.221	0.076	0.099
	SE	0.109	0.028	0.026	0.137	0.250	0.079	0.101
	MSE	0.014	0.001	0.001	0.019	0.076	0.006	0.010

这势必会对模型的方差估计存在一定影响。而且从表5-41看出该算法的计算开销还是很大的，因为算法过程中涉及带宽 h 的迭代。该算法的优势在于在折点附近更灵活地刻画折线的急剧变化，使得折点现象趋于平缓，而且由于此时的模型参数处处可导，也

表 5-34 折点数 $K = 2$, 异方差 $e_i \sim \mathbf{LA}(0, 1)$, $\tau = 0.25$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
MKQR	Bias	0.002	0.001	0.002	-0.010	0.018	0.000	0.001
	SD	0.116	0.023	0.025	0.132	0.241	0.072	0.111
	SE	0.115	0.021	0.025	0.130	0.239	0.069	0.101
	MSE	0.013	0.000	0.001	0.017	0.057	0.005	0.010
算法改进 1	Bias	0.007	0.003	0.000	0.191	-0.361	-0.104	0.013
	SD	0.111	0.024	0.022	0.114	0.237	0.079	0.085
	SE	0.268	0.058	0.271	0.529	0.050	0.450	0.315
	MSE	0.072	0.003	0.074	0.316	0.133	0.214	0.099
算法改进 2	Bias	0.036	0.009	0.000	0.145	-0.291	-0.093	0.015
	SD	0.100	0.023	0.022	0.122	0.246	0.080	0.084
	SE	0.275	0.059	0.300	0.553	0.050	0.511	0.358
	MSE	0.077	0.004	0.090	0.327	0.087	0.270	0.128
算法改进 3	Bias	-0.020	0.012	0.001	0.074	-0.143	-0.039	0.009
	SD	0.113	0.025	0.029	0.119	0.235	0.071	0.097
	SE	0.265	0.057	0.281	0.534	0.049	0.458	0.303
	MSE	0.071	0.003	0.079	0.290	0.023	0.212	0.092
算法改进 4	Bias	-0.043	0.007	0.002	0.027	-0.061	-0.006	-0.001
	SD	0.118	0.027	0.029	0.125	0.231	0.073	0.105
	SE	0.262	0.056	0.286	0.534	0.048	0.460	0.298
	MSE	0.070	0.003	0.082	0.287	0.006	0.212	0.089
算法改进 5	Bias	0.027	0.014	-0.018	0.030	-0.130	0.002	-0.006
	SD	0.103	0.022	0.026	0.113	0.177	0.068	0.087
	SE	0.120	0.027	0.024	0.123	0.182	0.064	0.083
	MSE	0.015	0.001	0.001	0.016	0.050	0.004	0.007

表 5-35 折点数 $K = 2$, 异方差 $e_i \sim \mathbf{N}(0, 1)$, $\tau = 0.5$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
MKQR	Bias	0.002	0.000	0.000	-0.007	0.014	0.002	0.001
	SD	0.083	0.017	0.018	0.101	0.168	0.052	0.080
	SE	0.083	0.014	0.018	0.094	0.171	0.050	0.072
	MSE	0.007	0.000	0.000	0.009	0.029	0.003	0.005
算法改进 1	Bias	0.012	0.003	0.000	0.131	-0.257	0.069	0.006
	SD	0.083	0.018	0.016	0.089	0.179	0.053	0.071
	SE	0.166	0.035	0.175	0.332	0.030	0.287	0.195
	MSE	0.028	0.001	0.031	0.127	0.067	0.087	0.038
算法改进 2	Bias	0.024	0.005	-0.001	0.103	-0.211	-0.062	0.006
	SD	0.080	0.017	0.016	0.092	0.177	0.054	0.067
	SE	0.169	0.036	0.188	0.340	0.030	0.316	0.213
	MSE	0.029	0.001	0.035	0.126	0.045	0.104	0.045
算法改进 3	Bias	-0.014	-0.003	0.001	0.045	-0.074	-0.017	0.002
	SD	0.086	0.019	0.022	0.094	0.167	0.051	0.075
	SE	0.165	0.035	0.181	0.335	0.030	0.292	0.190
	MSE	0.027	0.001	0.033	0.114	0.006	0.085	0.036
算法改进 4	Bias	-0.026	-0.006	0.000	0.022	-0.032	-0.001	0.001
	SD	0.100	0.023	0.022	0.097	0.154	0.057	0.085
	SE	0.166	0.036	0.186	0.340	0.029	0.300	0.191
	MSE	0.028	0.001	0.035	0.116	0.002	0.089	0.037
算法改进 5	Bias	-0.000	0.000	0.002	0.028	-0.047	0.004	-0.011
	SD	0.084	0.019	0.021	0.086	0.159	0.050	0.075
	SE	0.085	0.025	0.019	0.088	0.170	0.053	0.059
	MSE	0.007	0.001	0.000	0.030	0.010	0.007	0.004

增加了这一模型的应用上的广泛性。

值得一提的是, 本文对于算法改进 1 和 2 中的带宽 h 的选择进行了试验研究, 具体结果见图 5-2 和 5-3。两幅图的左列为同方差情形, 右列为异方差情形。图中的红色、蓝

表 5-36 折点数 $K = 2$, 异方差 $e_i \sim t(3)$, $\tau = 0.5$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
MKQR	Bias	0.003	0.000	0.001	-0.008	0.011	0.000	-0.003
	SD	0.088	0.017	0.019	0.109	0.190	0.058	0.086
	SE	0.090	0.016	0.019	0.103	0.186	0.055	0.079
	MSE	0.008	0.000	0.000	0.011	0.035	0.003	0.006
算法改进 1	Bias	-0.003	-0.001	-0.002	0.174	-0.325	-0.086	0.007
	SD	0.092	0.019	0.019	0.093	0.189	0.056	0.073
	SE	0.183	0.039	0.188	0.359	0.033	0.313	0.215
	MSE	0.033	0.002	0.035	0.159	0.107	0.106	0.046
算法改进 2	Bias	0.035	0.008	-0.001	0.092	-0.194	-0.060	0.008
	SD	0.090	0.019	0.018	0.101	0.197	0.060	0.076
	SE	0.186	0.040	0.215	0.380	0.033	0.364	0.248
	MSE	0.036	0.002	0.046	0.153	0.039	0.136	0.061
算法改进 3	Bias	-0.050	-0.015	-0.004	0.113	-0.171	-0.029	-0.003
	SD	0.111	0.066	0.068	0.207	0.302	0.085	0.149
	SE	0.179	0.038	0.193	0.361	0.033	0.315	0.210
	MSE	0.034	0.002	0.037	0.143	0.030	0.100	0.044
算法改进 4	Bias	-0.075	-0.024	-0.007	0.092	-0.129	-0.001	-0.019
	SD	0.141	0.092	0.092	0.284	0.390	0.113	0.205
	SE	0.179	0.038	0.199	0.363	0.033	0.319	0.210
	MSE	0.038	0.002	0.040	0.140	0.018	0.102	0.044
算法改进 5	Bias	0.002	0.000	-0.002	0.032	-0.038	-0.001	-0.005
	SD	0.089	0.020	0.021	0.103	0.161	0.051	0.082
	SE	0.092	0.028	0.022	0.121	0.127	0.057	0.078
	MSE	0.008	0.001	0.000	0.016	0.018	0.003	0.006

表 5-37 折点数 $K = 2$, 异方差 $e_i \sim \mathbf{LA}(0, 1)$, $\tau = 0.5$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
MKQR	Bias	-0.004	0.001	-0.001	-0.004	0.008	0.003	-0.001
	SD	0.070	0.014	0.015	0.086	0.143	0.046	0.066
	SE	0.073	0.013	0.016	0.083	0.152	0.045	0.064
	MSE	0.005	0.000	0.000	0.007	0.023	0.002	0.004
算法改进 1	Bias	-0.006	-0.002	-0.001	0.136	-0.248	-0.064	0.008
	SD	0.071	0.015	0.014	0.080	0.149	0.048	0.059
	SE	0.149	0.032	0.156	0.294	0.027	0.256	0.173
	MSE	0.022	0.001	0.024	0.105	0.062	0.070	0.030
算法改进 2	Bias	0.024	0.005	-0.000	0.080	-0.165	-0.048	0.007
	SD	0.069	0.014	0.014	0.087	0.159	0.050	0.060
	SE	0.151	0.032	0.173	0.307	0.027	0.291	0.196
	MSE	0.023	0.001	0.030	0.101	0.028	0.087	0.038
算法改进 3	Bias	-0.039	-0.009	0.001	0.072	-0.103	-0.021	0.006
	SD	0.081	0.018	0.021	0.082	0.142	0.047	0.063
	SE	0.146	0.031	0.159	0.296	0.027	0.257	0.169
	MSE	0.023	0.001	0.025	0.093	0.011	0.067	0.029
算法改进 4	Bias	-0.054	-0.013	0.001	0.050	-0.068	-0.006	-0.001
	SD	0.110	0.026	0.021	0.108	0.160	0.073	0.105
	SE	0.147	0.032	0.165	0.302	0.027	0.263	0.173
	MSE	0.025	0.001	0.027	0.094	0.005	0.069	0.030
算法改进 5	Bias	-0.001	-0.000	-0.002	0.026	-0.042	0.004	-0.000
	SD	0.070	0.015	0.017	0.096	0.145	0.049	0.058
	SE	0.073	0.022	0.015	0.102	0.135	0.052	0.063
	MSE	0.005	0.000	0.000	0.011	0.020	0.003	0.004

色、绿色线分别表示分位数位置为 $\tau = 0.25, 0.5, 0.75$ 时模拟结果的总均方误差 TMSE 随着带宽初值选择的变化, 带宽初值选择为 $h = n^{-d}$, 从 $d = 0$ 开始以每次增加 0.5 进行变动, 直到 $d = 5$ 。可以看到除了算法改进 1 中对于 $e_i \sim t(3)$ 时当带宽影响参数设

表 5-38 折点数 $K = 2$, 异方差 $e_i \sim N(0, 1)$, $\tau = 0.75$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
MKQR	Bias	-0.008	0.000	-0.002	-0.005	0.020	0.002	0.001
	SD	0.094	0.018	0.020	0.105	0.179	0.055	0.083
	SE	0.090	0.016	0.019	0.102	0.186	0.054	0.079
	MSE	0.008	0.000	0.000	0.010	0.035	0.003	0.006
算法改进 1	Bias	0.018	0.003	-0.002	0.137	-0.283	-0.075	0.002
	SD	0.093	0.020	0.017	0.096	0.190	0.055	0.072
	SE	0.209	0.045	0.219	0.414	0.038	0.360	0.244
	MSE	0.044	0.002	0.048	0.181	0.082	0.135	0.059
算法改进 2	Bias	0.032	0.006	-0.002	0.112	-0.242	-0.069	0.003
	SD	0.091	0.019	0.017	0.094	0.184	0.056	0.070
	SE	0.212	0.046	0.235	0.426	0.038	0.395	0.267
	MSE	0.046	0.002	0.055	0.194	0.060	0.161	0.071
算法改进 3	Bias	-0.041	0.024	-0.000	0.072	-0.094	-0.015	0.002
	SD	0.093	0.021	0.024	0.099	0.181	0.055	0.075
	SE	0.207	0.044	0.227	0.419	0.038	0.366	0.239
	MSE	0.044	0.003	0.051	0.181	0.010	0.134	0.057
算法改进 4	Bias	-0.042	-0.024	0.000	0.049	-0.051	-0.006	-0.005
	SD	0.105	0.025	0.025	0.099	0.168	0.056	0.080
	SE	0.210	0.045	0.233	0.426	0.039	0.373	0.242
	MSE	0.046	0.003	0.054	0.184	0.004	0.139	0.059
算法改进 5	Bias	-0.034	-0.014	0.017	0.056	-0.036	-0.005	-0.018
	SD	0.092	0.020	0.025	0.098	0.173	0.053	0.082
	SE	0.094	0.032	0.022	0.100	0.196	0.056	0.092
	MSE	0.010	0.001	0.001	0.013	0.040	0.003	0.009

表 5-39 折点数 $K = 2$, 异方差 $e_i \sim t(3)$, $\tau = 0.75$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
MKQR	Bias	-0.004	0.000	-0.002	-0.002	0.007	0.000	0.002
	SD	0.115	0.023	0.025	0.236	0.135	0.074	0.110
	SE	0.114	0.021	0.024	0.129	0.233	0.069	0.099
	MSE	0.013	0.000	0.001	0.017	0.054	0.005	0.010
算法改进 1	Bias	0.016	0.002	-0.002	0.175	-0.342	-0.095	0.000
	SD	0.119	0.025	0.021	0.109	0.236	0.067	0.089
	SE	0.262	0.057	0.269	0.514	0.049	0.445	0.308
	MSE	0.069	0.003	0.072	0.295	0.119	0.207	0.095
算法改进 2	Bias	0.043	0.008	-0.001	0.116	-0.254	-0.074	0.003
	SD	0.110	0.023	0.023	0.111	0.233	0.066	0.089
	SE	0.268	0.058	0.306	0.542	0.048	0.518	0.357
	MSE	0.074	0.003	0.094	0.307	0.067	0.274	0.127
算法改进 3	Bias	-0.068	-0.039	-0.009	0.145	-0.188	-0.022	0.013
	SD	0.145	0.095	0.099	0.295	0.426	0.119	0.212
	SE	0.257	0.056	0.274	0.515	0.048	0.447	0.300
	MSE	0.071	0.005	0.075	0.286	0.038	0.200	0.090
算法改进 4	Bias	-0.089	-0.040	-0.002	0.106	-0.121	-0.006	-0.010
	SD	0.142	0.068	0.070	0.221	0.318	0.097	0.161
	SE	0.261	0.056	0.286	0.525	0.048	0.461	0.305
	MSE	0.076	0.005	0.082	0.287	0.017	0.213	0.093
算法改进 5	Bias	-0.030	-0.016	-0.015	0.073	-0.047	0.005	-0.021
	SD	0.105	0.025	0.026	0.136	0.222	0.070	0.113
	SE	0.108	0.035	0.032	0.154	0.254	0.077	0.117
	MSE	0.013	0.001	0.001	0.029	0.067	0.006	0.014

为 0 时的波动较大, 其余时候总均方误差 TMSE 随着带宽的变动影响很小。而事实上当 $d = 0$ 时, 带宽 $h = 1$, 而往往实际问题的解决中带宽不会选择为 $h = 1$ 。因此可以认为这两种算法对于带宽的敏感性并不高, 这也是这两种算法改进存在的优势。

表 5-40 折点数 $K = 2$ ，异方差 $e_i \sim \mathbf{LA}(0, 1)$ ， $\tau = 0.75$ 情形下各方法模拟结果比较

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
MKQR	Bias	-0.015	0.001	-0.004	-0.008	0.015	0.008	-0.005
	SD	0.115	0.021	0.025	0.143	0.023	0.077	0.110
	SE	0.115	0.021	0.025	0.131	0.238	0.070	0.100
	MSE	0.013	0.000	0.001	0.017	0.057	0.005	0.010
算法改进 1	Bias	0.011	0.001	-0.001	0.164	-0.335	-0.087	-0.004
	SD	0.119	0.025	0.021	0.115	0.227	0.068	0.093
	SE	0.266	0.057	0.276	0.528	0.049	0.454	0.312
	MSE	0.071	0.003	0.076	0.306	0.115	0.214	0.098
算法改进 2	Bias	0.034	0.006	-0.000	0.122	-0.276	-0.073	-0.003
	SD	0.115	0.024	0.020	0.118	0.232	0.067	0.094
	SE	0.272	0.059	0.307	0.552	0.049	0.517	0.355
	MSE	0.075	0.003	0.094	0.320	0.078	0.273	0.126
算法改进 3	Bias	-0.050	-0.028	0.001	0.100	-0.137	-0.023	0.000
	SD	0.115	0.026	0.028	0.125	0.219	0.070	0.100
	SE	0.264	0.057	0.285	0.534	0.048	0.463	0.306
	MSE	0.072	0.004	0.081	0.295	0.021	0.215	0.094
算法改进 4	Bias	-0.064	-0.032	0.001	0.075	-0.083	-0.006	-0.001
	SD	0.130	0.030	0.028	0.125	0.211	0.073	0.105
	SE	0.267	0.058	0.295	0.543	0.049	0.473	0.312
	MSE	0.076	0.004	0.087	0.301	0.009	0.224	0.097
算法改进 5	Bias	-0.038	-0.017	0.017	0.077	-0.041	0.012	-0.011
	SD	0.101	0.024	0.024	0.165	0.250	0.083	0.103
	SE	0.103	0.027	0.027	0.155	0.246	0.088	0.106
	MSE	0.012	0.001	0.001	0.030	0.062	0.008	0.011

表 5-41 算法一次估计所用时间

K	e_i	MKQR	改进 1	改进 2	改进 3	改进 4	改进 5
K=1	N(0, 1) 同方差	7.242	0.815	9.805	0.512	3.743	5.966
	N(0, 1) 异方差	7.363	0.934	7.762	0.603	3.825	6.031
	t(3) 同方差	7.532	0.993	7.529	0.487	4.120	5.875
	t(3) 异方差	7.498	1.012	4.281	0.575	4.097	5.993
	LA(0, 1) 同方差	8.542	1.121	2.933	0.532	2.715	5.778
	LA(0, 1) 异方差	8.993	1.110	5.893	0.549	3.084	5.946
K=2	N(0, 1) 同方差	14.426	2.902	27.892	2.272	7.452	9.735
	N(0, 1) 异方差	14.205	2.687	41.723	2.214	8.005	10.568
	t(3) 同方差	14.894	2.563	25.327	1.976	8.631	9.998
	t(3) 异方差	14.923	2.899	20.487	1.998	8.464	10.126
	LA(0, 1) 同方差	13.672	2.967	17.098	1.835	7.899	10.163
	LA(0, 1) 异方差	14.223	3.106	23.679	2.003	7.764	10.001

表 5-42 Wald 置信区间构造方法与 SRS 置信区间构造方法对比

τ	方法	覆盖率		区间长度	
		δ_1	δ_2	δ_1	δ_2
0.25	Wald	93.2%	91.3%	0.451	0.357
0.25	SRS	99.3%	98.2%	0.554	0.457
0.5	Wald	92.1%	91.4%	0.394	0.315
0.5	SRS	98.6%	97.4%	0.483	0.384
0.75	Wald	89.2%	90.9%	0.553	0.447
0.75	SRS	100.0%	99.2%	0.625	0.555

算法改进 4 所用到的损失函数和算法改进 3 一致，但是可以认为是为了应对特殊和高维情况下的算法改进 3 的 Boosting 版本。该算法对参数估计的效果同样比较好，虽然其运算效率不如算法 3，但是该算法本身具有良好的特征选择机制^[45]，这是该算法的

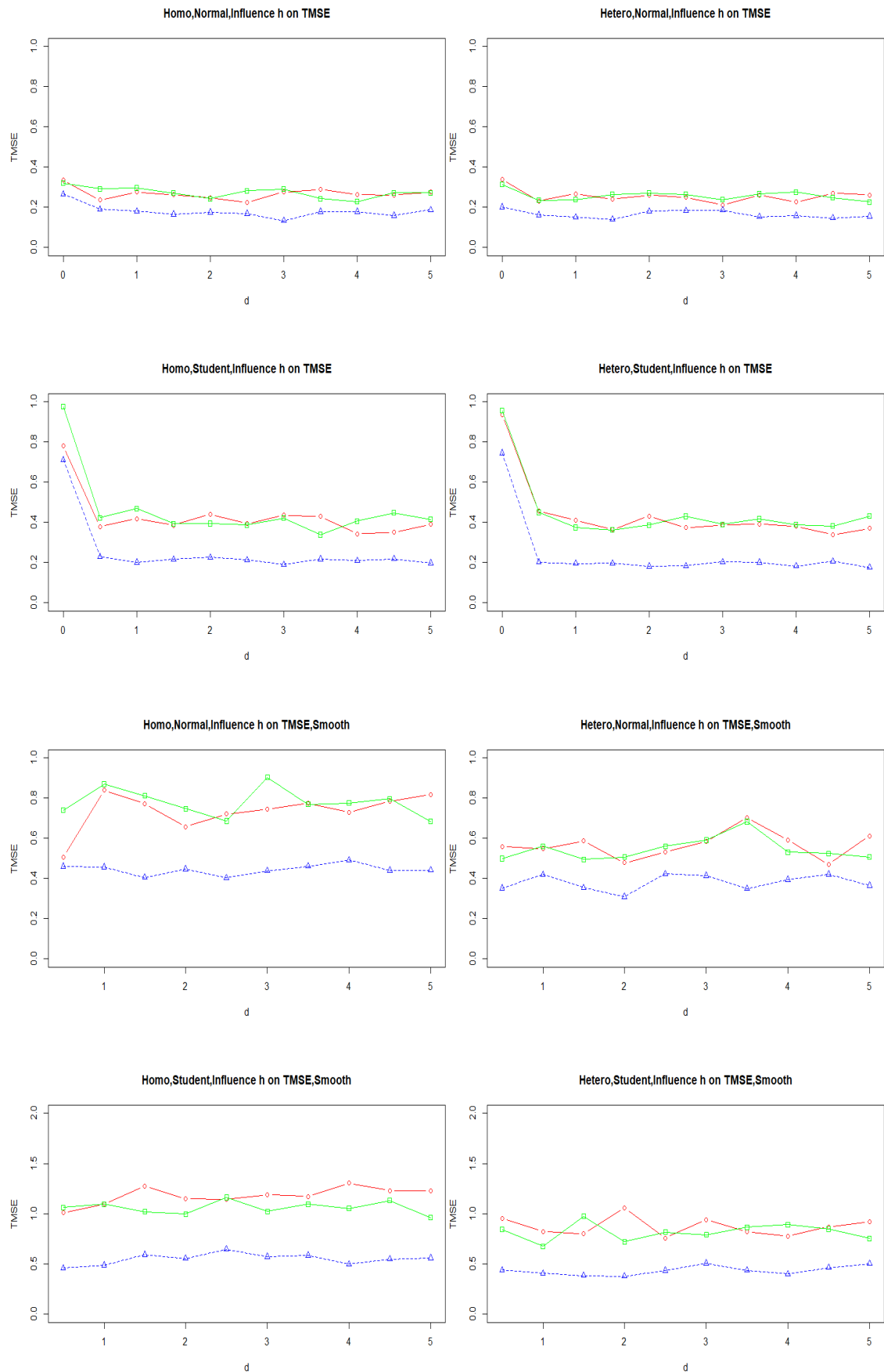
优势。由于篇幅原因，在此不做赘述。

算法改进 5 是本人在改进当中较为推崇的方法。这种结合了 Wang 与 Brown^[48] 诱导光滑化的方法，不但对于参数估计的效果很好 (参数估计的偏差很小)，而且对于方差估计也具有很好的效果。这主要得益于该方法在光滑化的同时并没有改变模型的本质，因此一定意义上对所有的估计效果都很好，间接说明了参数估计的渐近正态性质。尽管该算法看上去较为复杂，但是算法在为数不多的迭代步数中即可完成收敛，因而计算开销并不大，而且该算法在迭代的过程中，可以直接通过 \hat{H} 的迭代完成所有参数的标准差估计，无需再另外对标准差估计，具有“一步到位”的良好效果。与此同时，该迭代算法对于初值的选取也并无苛刻要求， \hat{H} 的初值选择单位阵即可，这说明算法的泛化性质较好。

与此同时，Zhong 等人^[24] 的方法和算法改进 4、5 的方法是先行利用线性化技巧估计出折点位置的方法，因而对于折点数 K 较大的情形也可以从容应对；而算法改进 1、2、3 对于 K 较大的情形，在设定最大迭代数为 10000 的情形下也很难实现算法的收敛，这也是这些算法的不足之处。因此 $K = 3$ 的模拟结果本人不在正文当中放出，对于 Zhong 等人^[24] 的方法和算法改进 4、5 的方法本人进行 $K = 3$ 的模拟结果放在文章的附录位置供读者参阅。

而对于多折点分位数回归模型的折点位置置信区间构造，本人比对了传统的 Wald 置信区间构造方法与 SRS 置信区间构造方法，如表 5-42 所示。对于两种方法，都构造了折点的 95% 置信区间。可以看到，Wald 置信区间构造方法会有些许欠覆盖的情况，而 SRS 方法的置信区间长度会稍长一些，可以说各有优劣，依据不同情形可以采用不同的方法。

综上，Zhong 等人^[24] 的折点估计方法以及本人结合他人研究成果进行的改进都存在着或多或少的优劣之处。但是不难看出 Zhong 等人^[24] 的方法以及诱导光滑化改进算法的优势较为明显，综合考虑是更值得推广的算法。



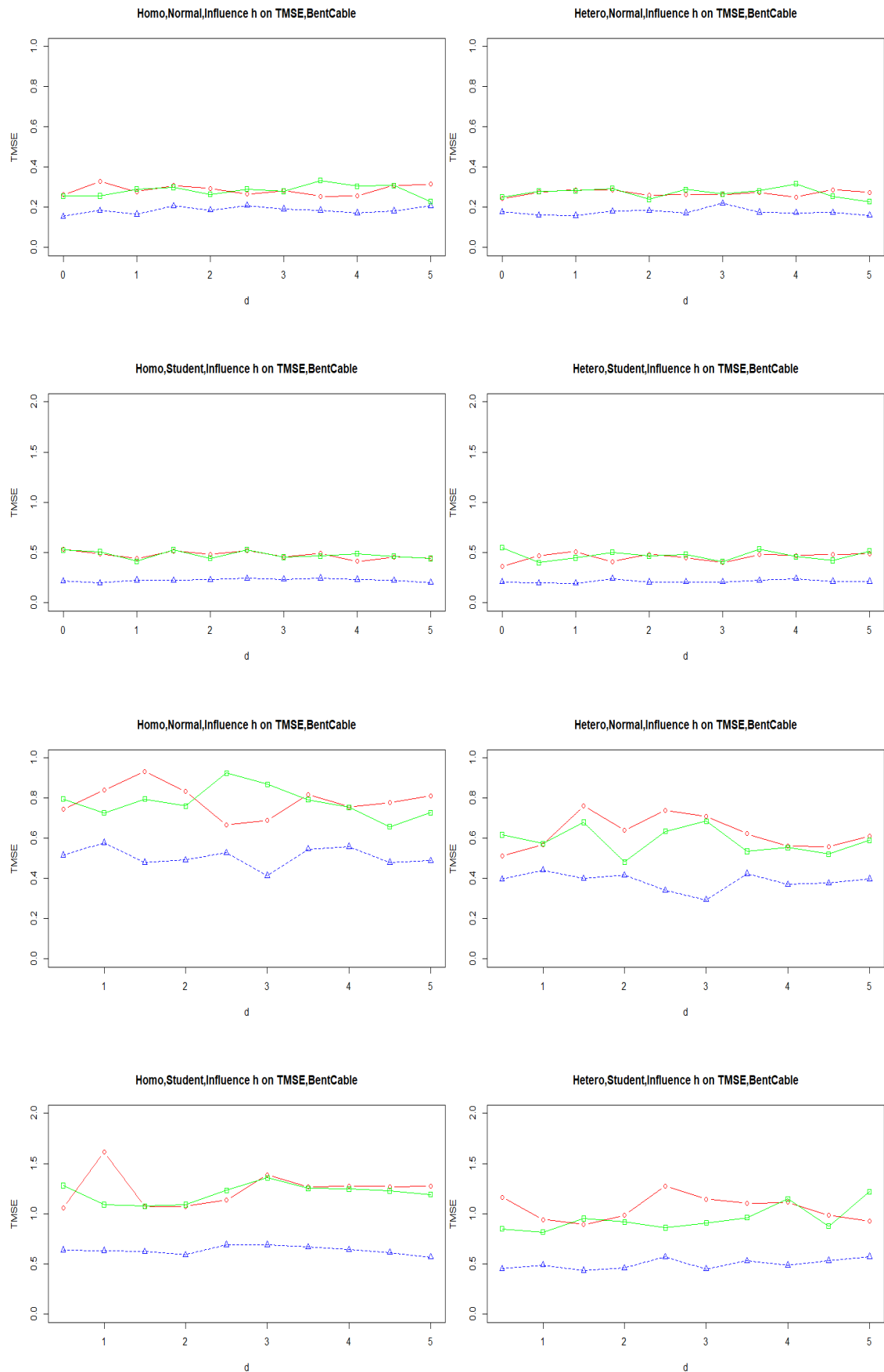


图 5-3 算法改进 2 中带宽 $h = n^{-d}$ 中 d 的选择对估计结果影响

6 实际数据分析

在这一部分，应用此前提到的模型和算法进行实证分析。对于单折点回归模型，本文选用我国国民生产总值与第二产业比重数据，希望通过单折点模型分析我国各城市的国民生产总值与第二产业产值占比之间的变化关系；而对于多折点分位数回归模型，本人选用冈比亚女性肱三头肌皮褶厚度的生物数据集，意在研究冈比亚女性的肱三头肌皮褶厚度随年龄的变化关系。这两个数据集在 Zhong 等人^[24] 的文章中都有出现。

6.1 单折点回归模型

6.1.1 数据简介

自 1978 年中国实行改革开放政策以来，人均 GDP 飞速增长，产业结构也发生了巨大的变化。现已有不少学者对中国产业结构的变化进行了相关研究工作。传统发展经济学理论告诉我们，一国在发展过程中，第一产业所占比重逐渐下降，而第三产业所占比重反而逐渐上升。同时，第二产业的比重也经历了一个先快速增长，然后逐渐停止增长甚至下降的过程。这意味着人均 GDP 与第二产业占比之间极有可能存在着显著的变化关系。研究这一关系有助于我国经济社会以及产业结构的转型，从而提升产业经济效益。因此就这一变化关系，采用折点模型进行探究。

此处用到的数据集来自于世界合作发展组织的数据库，由 Zhong 等人^[24] 进行搜集整理。该数据集记录了我国 280 个地级市的人均 GDP 以及各产业 GDP。在此处解释变量 x_i 为各城市的人均 GDP（单位：万元），响应变量 y_i 为各城市的第二产业占该城市的总 GDP 比重，其余解释变量 z_i 为各城市的财政支出与固定资产投资占城市的总 GDP 比重。

首先利用算法2-1的折点效应检验，发现该数据集确实存在折点效应。再通过算法3-2的后项消除算法，得到该数据集的折点数估计为 $\hat{K} = 1$ 。因此，本人决定用单折点均值回归模型与单折点分位数回归模型分别分析该数据集，其中对于单折点分位数回归模型，分位数 τ 分别为 0.1, 0.3, 0.5, 0.7, 0.9。

6.1.2 结果展示与分析

首先，本人利用 Hansen^[21] 提出的单折点均值回归模型分析数据。利用网格搜索算法，得到均值回归情况下的折点位置估计为 $\hat{\delta} = 3.7$ 。由此通过单折点最小二乘估计模型系数，得到单折点均值回归模型形式为：

$$y = 0.4656 + 0.0866(x - 3.7000)_- + 0.0059(x - 3.7000)_+ - 0.8411z_1 + 0.0761z_2, \quad (6-1)$$

其中 z_1 表示财政支出占比这一解释变量，而 z_2 表示固定资产投资占比这一解释变量。该模型在图6-1上用红线表示。可以得到各个参数估计的标准差估计，其中常数项的标

准差为 0.0262, β_1 的标准差为 0.0149, β_2 的标准差为 0.0021, γ_1 的标准差为 0.2361, γ_2 的标准差为 0.0186, 折点位置 δ 的标准差为 0.1992, 其中 γ_1 , γ_2 分别对应两个解释变量 z_1 和 z_2 的系数。与此同时, 可以利用算法2-3得到估计(6-1)式的回归函数形式的逐点 90% 置信区间, 将其与单折点均值回归模型一起呈现在图6-1上, 用蓝线表示。具体见图6-1。

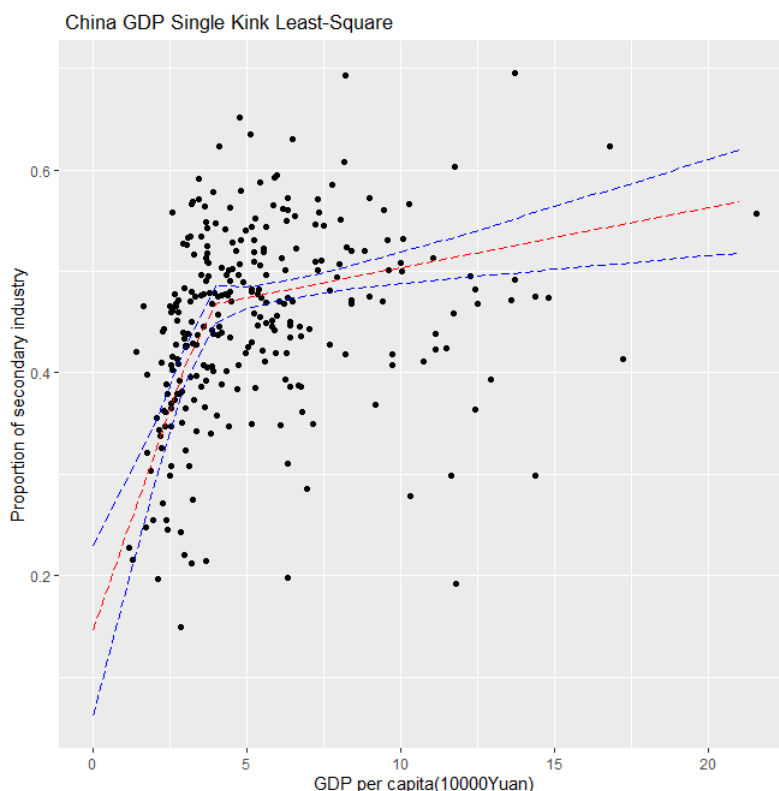


图 6-1 第二产业占比与人均 GDP 数据集单折点均值回归模型及回归函数置信区间

此时的单折点均值回归模型的折点位置为 $\delta = 3.7$ 附近, 这说明在中国, 当一座城市的人均 GDP 达到 3.7 万元左右时, 第二产业比重随 GDP 增长而增长的幅度会大幅减缓。这一点有效论证了传统发展经济中第二产业比重随 GDP 增长先快速增长, 接着逐渐减缓增长幅度, 以至于停止增长甚至下降的观点。而对于单折点均值回归模型的置信区间而言, 也近乎呈现相似的变化。

但是这是基于数据的均值回归而言的。显然均值并不能代表一组数据的真实水平。对于较为发达的一线城市, 例如北京、上海、深圳, 他们的第二产业占比随 GDP 增长的变化情况就与一些相对落后的四五线城市之间存在着明显差别, 而且基于社会经济发展环境的不同, 这些城市的产业占比增长变化幅度之间也存在着明显差异。我们渴望显现并且比较这种差异, 因而引进了分位数回归。在此处, 对于不同水平的分位数采用算法3-2得到的折点数估计均为 $\hat{K} = 1$, 因此考虑利用 Li 等人^[22]给出的单折点分位数回归的方法对该数据集进行分位数回归。具体的结果如下图6-2所示。具体的不同分位数水平下的参数估计及其方差估计结果参见表6-1。

从图6-2中, 我们可以明显发现, 在不同分位数水平下, 模型的折点位置估计 $\hat{\delta}$ 不

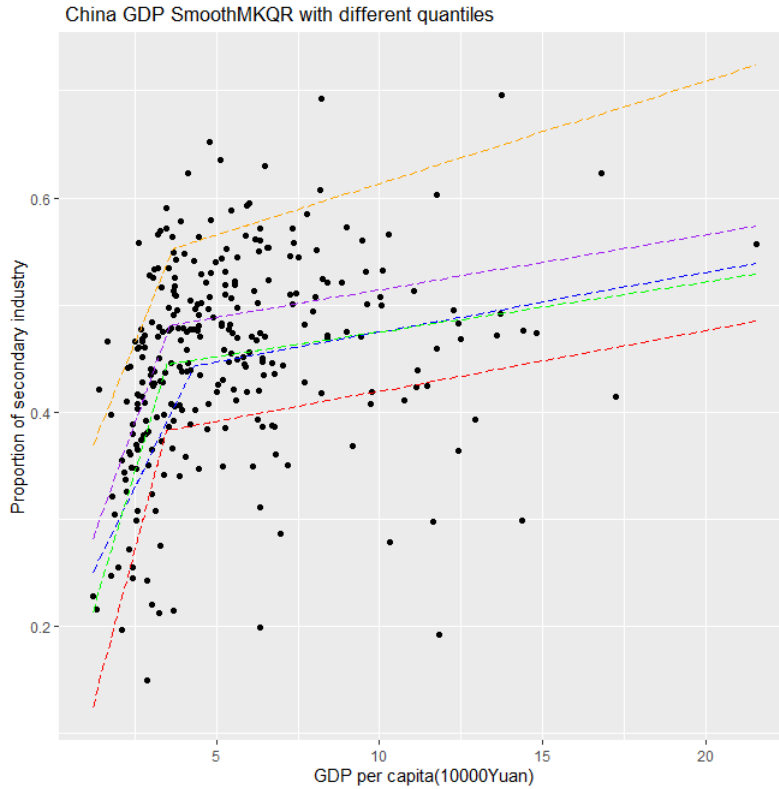


图 6-2 第二产业占比与人均 GDP 数据集单折点分位数回归。红色 $\tau = 0.1$ ，蓝色 $\tau = 0.3$ ，绿色 $\tau = 0.5$ ，紫色 $\tau = 0.7$ ，橙色 $\tau = 0.9$ 。

表 6-1 第二产业占比与人均 GDP 数据集单折点分位数回归估计结果

	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
\hat{K}	1	1	1	1	1
$\hat{\alpha}_0$	-0.012	0.176	0.091	0.182	0.279
SE	0.115	0.056	0.058	0.042	0.071
$\hat{\alpha}_1$	0.114	0.062	0.102	0.084	0.075
SE	0.037	0.014	0.018	0.013	0.021
$\hat{\beta}$	-0.109	-0.056	-0.097	-0.079	-0.065
SE	0.037	0.015	0.019	0.013	0.021
$\hat{\delta}$	3.457	4.315	3.471	3.571	3.664
SE	0.351	0.432	0.197	0.186	0.389
$\hat{\gamma}_1$	-0.985	-0.910	-0.711	-0.828	-0.707
SE	0.397	0.261	0.199	0.152	0.275
$\hat{\gamma}_2$	0.078	0.087	0.090	0.098	0.056
SE	0.032	0.021	0.016	0.012	0.022

尽相同，其中 $\tau = 0.3$ 分位数水平下的折点位置估计 $\hat{\delta} = 4.371$ ，是所有分位数水平中的最大位置，这表明对于数据集的 280 座城市中第二产业占比相对中等偏下位置的城市，达成产业转型的人均 GDP 要求会稍许苛刻一些。总体而言，在五个不同的分位数水平下，折点位置估计 $\hat{\delta}$ 以及折点两侧的线性模型斜率估计 $\hat{\alpha}_1$ 与 $\hat{\alpha}_1 + \hat{\beta}$ 都较为接近，这都可以间接证明此前提到的传统经济学发展论调。事实上，在该模型中的折点位置可以被认为是经济发展与产业结构转换当中存在的重要转折点和门槛。当数据点处于这些分位数水平下的折点位置附近时，我们可以认为这些对应城市的经济主体正处于一个重要的转型期。在此期间，如果该城市的经济发展能够跳过这一门槛，就可以有效实现

城市经济结构调整,从而促进该城市居民进入小康社会乃至高收入群体行列,实现经济发展的快速腾飞。否则,该城市可能会因为产业结构的转型失败从而导致经济发展停滞不前,进而导致该城市居民可能陷入中等收入陷阱的瓶颈当中。因此,通过该单折点分位数回归模型的应用,可以有效帮助位于数据集当中的 280 所城市找准属于自身的经济发展定位,加速其经济发展与产业结构转变进程,为其未来经济发展提供有效的建议。例如,基于城市目前所处的分位数水平,若位于该分位数水平下模型折点位置的左侧,可以考虑加大第二产业的发展力度;而位于折点位置右侧,亦或是人均 GDP 已经达到较高水平的发达城市,可以考虑放缓第二产业的发展,转而大力发展城市的第三产业,进一步优化产业结构。由此通过单折点分位数回归模型,做到“对症下药”,具体问题具体分析。

6.2 多折点分位数回归模型

6.2.1 数据简介

肱三头肌皮褶厚度是反映身体指标的一个重要生物学参数。这一生物学参数可以有效反映人体的肥胖情况。科学研究表明,肱三头肌皮褶厚度位于人类 85% 分位数水平及以上的人更易受到富贵病、高血压等基于肥胖而产生的病症的不良影响,而肱三头肌皮褶厚度位于人类较低分位数水平及以下的人则通常很瘦,且容易被营养不良等病症所困扰。鉴于肱三头肌皮褶厚度的重要性,在此利用多折点分位数回归模型研究这一生物学参数随年龄变化的趋势。

在此之前,Durnin 与 Womersley^[51]通过线性回归方法,Cole 与 Green^[52]通过样条插值方法研究过这一问题,但是最终的研究结果都没有凸显出肱三头肌皮褶厚度随年龄增长而变化的阈值,而且得到的模型结果可解释性较差^[24]。在这里通过本文研究的多折点分位数回归模型对这一数据集进行研究。

本文所用数据集来自于 Royston 与 Sauerbrei^[37]的文献。他们收集了 892 位冈比亚种族的不同年龄段女性的肱三头肌皮褶厚度数据,以此研究肱三头肌皮褶厚度随年龄变化的影响。在此处的肱三头肌皮褶厚度经过对数处理,数据可视化后坐标横轴(解释变量 x_i)为冈比亚女性的年龄,纵轴(响应变量 y_i)为冈比亚女性的肱三头肌皮褶厚度的对数值。该数据集解释变量只有 x_i 。

6.2.2 结果展示与分析

在此处,首先利用算法3-3对原数据集进行检验,发现该数据集确实存在折点效应。接着利用算法3-2得到该数据集的折点数估计 $\hat{K} = 2$ 。由此利用(3-1)式 $K = 2$ 情形对该数据集进行分位数回归。在此处,本人为了比较 Zhong 等人^[24]提供的 BRISQ 算法3-1与其他算法改进之间的优劣,利用算法3-1,算法4-1,算法4-2,算法4-3,算法4-4,算法4-5分别对这一数据集进行多折点分位数回归,得到的各方法的参数估计与方差估计的数值结果对比参见表6-2所示,多折点分位数回归的图像参见下图6-3—图6-8。在此处本文

选择的分位数水平仍旧为 $\tau = 0.1, 0.3, 0.5, 0.7, 0.9$ ，而图中的各分位数水平的回归线代表的颜色与 6.1 节单折点分位数回归中的颜色一致。

事实上，可以简单地认为，分位数水平某种程度上反映了肱三头肌皮褶厚度的水平，进而反映了冈比亚女性的胖瘦程度，其中 $\tau = 0.9$ 对应体重较重的女性，自然而然地， $\tau = 0.1$ 对应体重较轻的女性。可以简单地以图6-3为例，该图反映的是利用 Zhong 等人^[24] 提供的 BRISQ 算法对这一数据集进行多折点分位数回归得到的结果。可以看到，无论在什么分位数水平下，两个折点的位置都大致在 10 岁与 20 岁左右。这两个折点位置将冈比亚女性的生命历程分解为了 3 个部分，可以理解为幼年期、青年时期和青年时期以后。基本上，在幼年期，肱三头肌皮褶厚度呈明显的随年龄增长而减小的趋势，而在青年时期，这一趋势被完全逆转，此时的肱三头肌皮褶厚度呈明显的随年龄增长而增大的趋势。而在第二个折点位置之后，肱三头肌皮褶厚度随年龄增长所体现出的增长与下降趋势都不是特别明显，此时的肱三头肌皮褶厚度趋于稳定。而在不同分位数水平下，两个折点估计的位置也有所不同。可以很明显地发现，处于较低分位数水平位置的冈比亚女性（即相对较瘦的那些），她们的折点位置估计 $\hat{\delta}_1, \hat{\delta}_2$ 都较大，而相对地，对于处于较高分位数水平位置的冈比亚女性（即相对较胖的那些），她们的折点位置估计 $\hat{\delta}_1, \hat{\delta}_2$ 都较小。由此有理由相信，肱三头肌皮褶厚度随年龄的变化趋势与人体的胖瘦程度之间存在着密不可分的关系。对于较瘦的女性，他们肱三头肌皮褶厚度的发育程度较为迟缓，发育周期会较她们的平均水平而言向后延迟；而对于较胖的女性，她们的肱三头肌皮褶厚度的发育程度则较为迅速，发育周期较平均水平而提前了。由此在得出这一结论的前提下，就可以针对不同人群的胖瘦情况，依据他们所面临的健康状况以及身体发育问题对症下药。进一步，从表6-2也可以发现， $\hat{\delta}_2$ 处的方差估计要远大于 $\hat{\delta}_1$ 处的方差估计，这说明该数据集存在着较大的异方差性。而针对异方差性较大的数据进行回归分析，采用分位数方法显然是更为恰当的，这从另一方面体现了该模型的优势。

在关注了这一数据集的总体情况后，本文将这 6 种不同的算法结果进行了对比。不难发现，在这几张结果图表当中，总体效果较好的是算法3-1，算法4-4以及算法4-5。这几张图表下的折点位置估计是采用线性化技巧进行估计的，得到的大致是随着分位数水平的提高，折点位置估计减小的趋势，而且这几种估计方法得到的方差估计也较小，有效地保证了估计的稳健性。

对于算法改进 1 和算法改进 2 而言，尽管算法改进 1 采用核函数对整体模型的光滑化进行改进，算法改进 2 在折点估计位置处采用了二次函数进行了逼近，在折点位置处灵活描述了变量间关系的急剧变化，但是这两种算法的参数的方差估计较大，尤其是对于折点位置估计 $\hat{\delta}_2$ ，在 $\tau = 0.1$ 的分位数水平下，方差估计达到了 9 左右，显然此时得到的不是一个稳健的估计，但是这两种算法给出的多折点分位数回归模型的形式大致与算法3-1，算法4-4以及算法4-5呈现的结果吻合。而对于算法改进 3 而言，尽管该算法的运算效率较高，但是该算法得到的在不同分位数水平下的折点位置估计 $\hat{\delta}_1$ 和 $\hat{\delta}_2$ 几乎没有变动， $\hat{\delta}_1$ 基本上在 7.7 附近， $\hat{\delta}_2$ 基本上在 18.12 附近，这显然是不符合该生

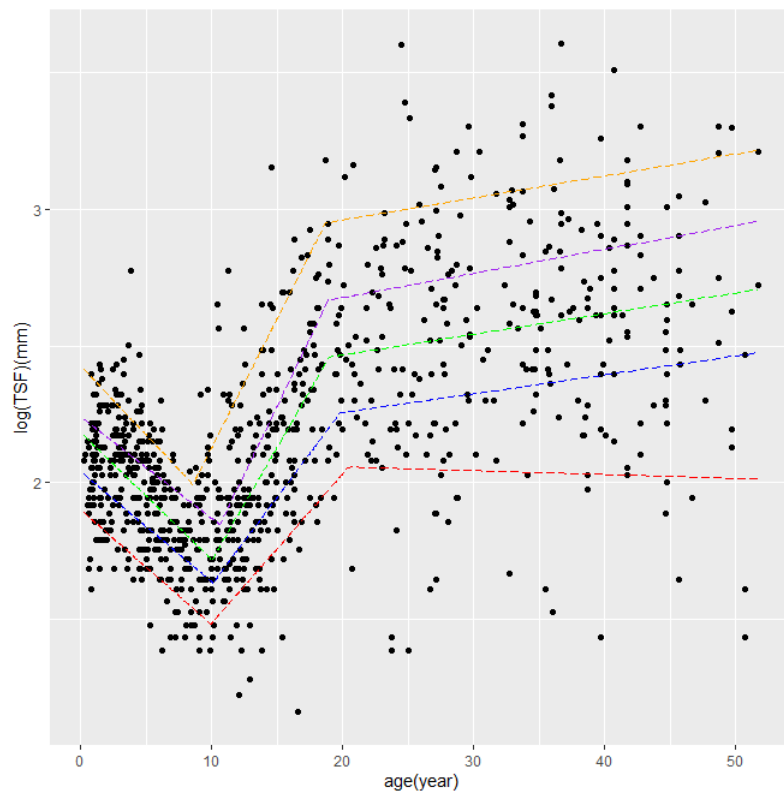


图 6-3 冈比亚女性年龄与肱三头肌皮褶厚度数据集多折点分位数回归 (MKQR 算法)

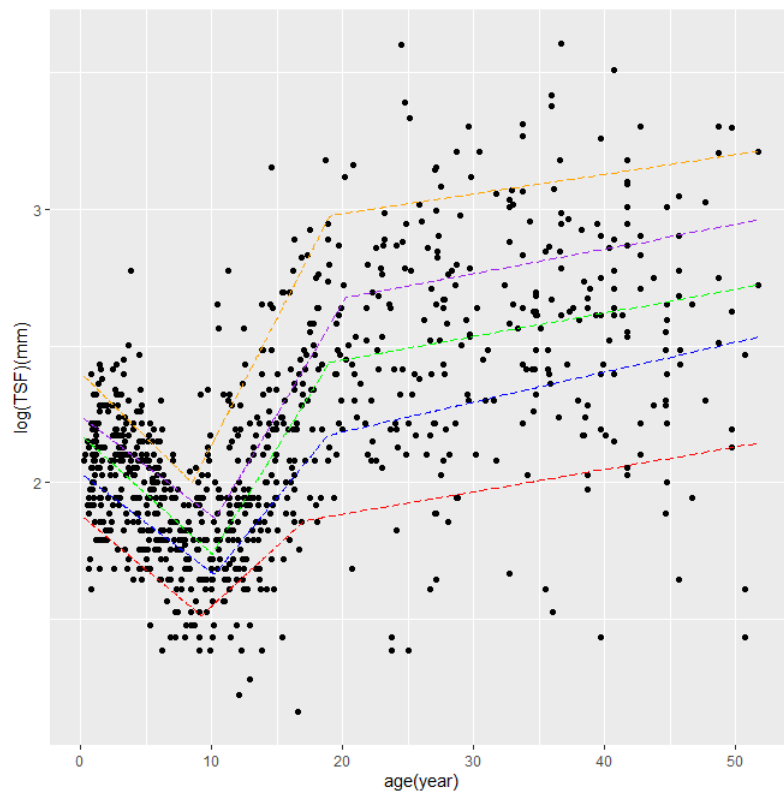


图 6-4 冈比亚女性年龄与肱三头肌皮褶厚度数据集多折点分位数回归 (算法改进 1)

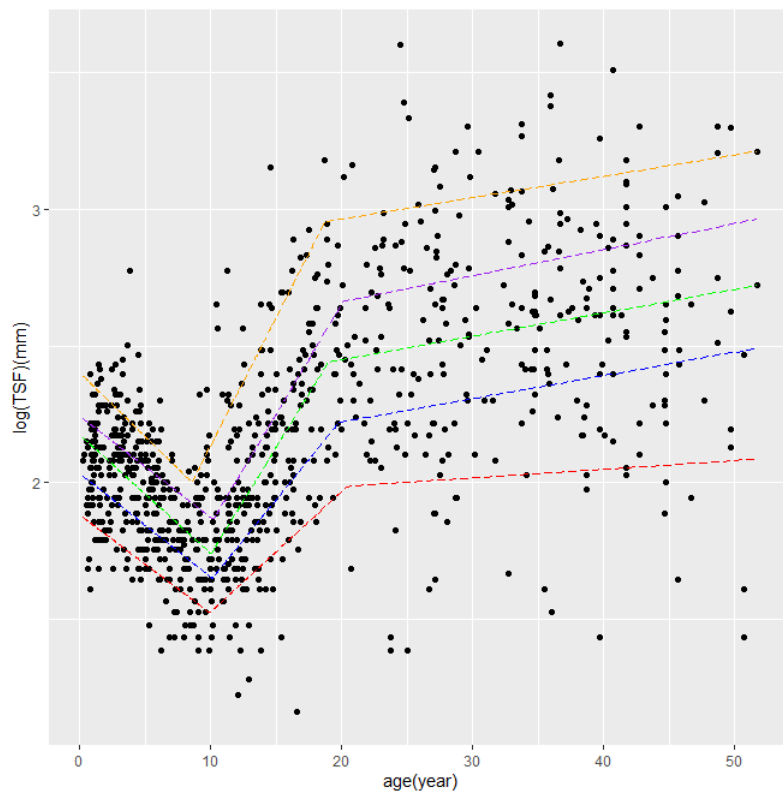


图 6-5 冈比亚女性年龄与肱三头肌皮褶厚度数据集多折点分位数回归 (算法改进 2)

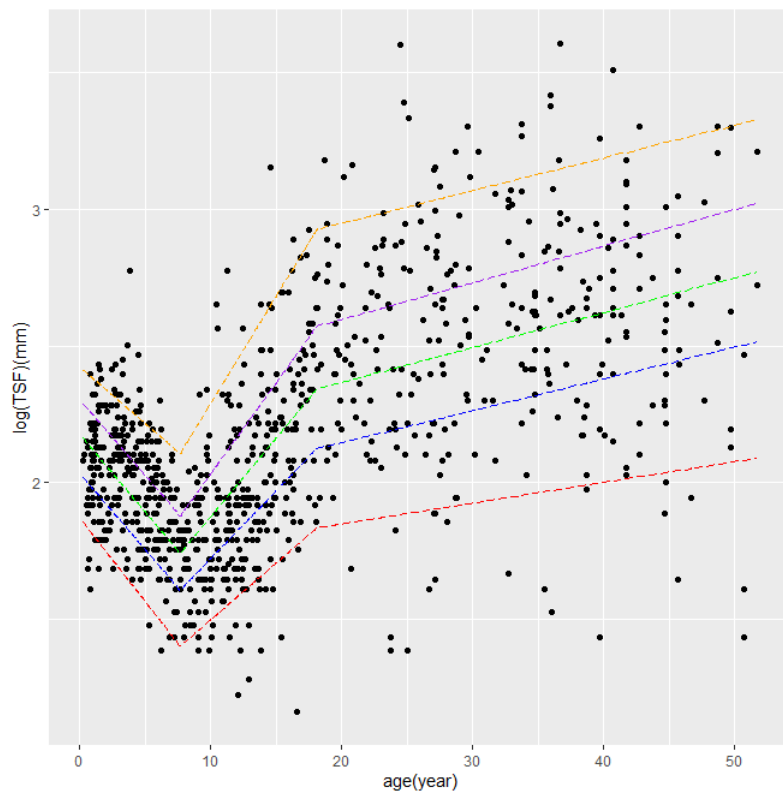


图 6-6 冈比亚女性年龄与肱三头肌皮褶厚度数据集多折点分位数回归 (算法改进 3)

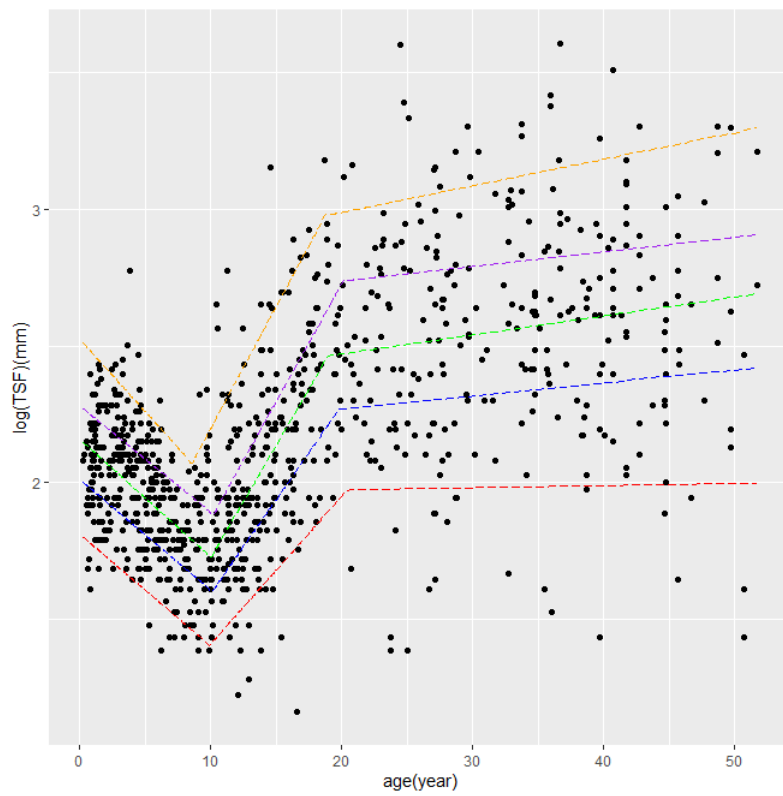


图 6-7 冈比亚女性年龄与肱三头肌皮褶厚度数据集多折点分位数回归 (算法改进 4)

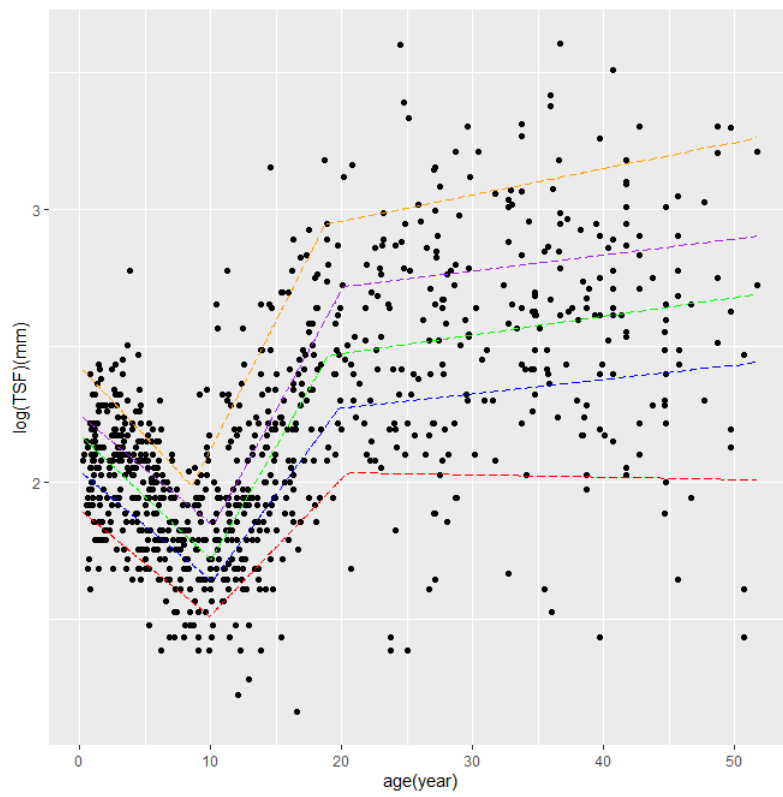


图 6-8 冈比亚女性年龄与肱三头肌皮褶厚度数据集多折点分位数回归 (算法改进 5)

表 6-2 Triceps 数据集各算法结果比较

	Method	$\hat{\alpha}_0$	SE	$\hat{\alpha}_1$	SE	$\hat{\beta}_1$	SE	$\hat{\beta}_2$	SE	$\hat{\delta}_1$	SE	$\hat{\delta}_2$	SE
$\tau = 0.1$	MKQR	1.896	0.021	-0.041	0.002	0.097	0.014	-0.057	0.015	10.042	0.176	20.332	2.941
	算法改进 1	1.882	0.100	-0.040	0.022	0.084	0.050	-0.036	0.048	9.215	2.554	17.084	8.806
	算法改进 2	1.883	0.097	-0.036	0.021	0.080	0.038	-0.041	0.038	9.923	2.691	20.458	8.116
	算法改进 3	1.871	0.106	-0.061	0.024	0.102	0.037	-0.034	0.036	7.671	1.521	18.120	10.700
	算法改进 4	1.812	0.096	-0.041	0.021	0.095	0.038	-0.053	0.039	9.923	2.254	20.458	6.147
	算法改进 5	1.902	0.028	-0.039	0.005	0.089	0.012	-0.051	0.013	9.856	0.611	20.684	2.594
$\tau = 0.3$	MKQR	2.042	0.029	-0.040	0.005	0.106	0.010	-0.058	0.010	10.120	0.373	19.690	1.533
	算法改进 1	2.035	0.065	-0.036	0.010	0.095	0.025	-0.048	0.024	10.198	0.980	18.886	4.098
	算法改进 2	2.035	0.065	-0.038	0.010	0.097	0.022	-0.050	0.021	10.115	0.931	19.809	3.825
	算法改进 3	2.034	0.080	-0.056	0.018	0.105	0.024	-0.038	0.017	7.671	1.052	18.120	3.851
	算法改进 4	2.012	0.065	-0.040	0.010	0.109	0.022	-0.064	0.021	10.115	0.802	19.809	2.989
	算法改进 5	2.043	0.031	-0.040	0.006	0.109	0.010	-0.060	0.010	10.486	0.489	19.917	1.300
$\tau = 0.5$	MKQR	2.183	0.027	-0.046	0.005	0.129	0.010	-0.075	0.009	10.030	0.307	18.993	1.051
	算法改进 1	2.175	0.054	-0.043	0.009	0.121	0.020	-0.069	0.019	10.030	0.638	18.993	2.279
	算法改进 2	2.175	0.054	-0.043	0.009	0.121	0.020	-0.070	0.019	10.014	0.652	18.992	2.297
	算法改进 3	2.177	0.072	-0.057	0.016	0.114	0.022	-0.045	0.016	7.671	0.818	18.120	3.646
	算法改进 4	2.159	0.626	-0.043	0.009	0.125	0.020	-0.075	0.019	10.052	0.626	18.992	2.131
	算法改进 5	2.175	0.031	-0.045	0.005	0.125	0.012	-0.076	0.012	10.037	0.413	19.005	1.173
$\tau = 0.7$	MKQR	2.249	0.027	-0.040	0.003	0.136	0.008	-0.088	0.008	10.391	0.432	18.969	0.417
	算法改进 1	2.244	0.043	-0.036	0.009	0.117	0.022	-0.072	0.022	10.277	0.906	20.272	2.642
	算法改进 2	2.244	0.044	-0.037	0.010	0.117	0.023	-0.071	0.023	10.202	0.995	20.150	2.838
	算法改进 3	2.304	0.072	-0.055	0.015	0.122	0.022	-0.053	0.017	7.680	0.744	18.123	3.244
	算法改进 4	2.281	0.044	-0.039	0.010	0.125	0.023	-0.080	0.023	10.202	0.906	20.175	2.458
	算法改进 5	2.248	0.028	-0.040	0.005	0.127	0.011	-0.082	0.011	10.202	0.373	20.009	1.156
$\tau = 0.9$	MKQR	2.426	0.037	-0.050	0.008	0.144	0.015	-0.086	0.013	8.602	0.471	18.688	1.482
	算法改进 1	2.402	0.126	-0.047	0.027	0.140	0.050	-0.085	0.044	8.490	1.580	19.053	4.934
	算法改进 2	2.402	0.124	-0.047	0.027	0.140	0.050	-0.086	0.044	8.572	1.627	18.742	4.563
	算法改进 3	2.423	0.126	-0.041	0.034	0.120	0.046	-0.066	0.033	7.672	2.044	18.120	3.901
	算法改进 4	2.524	1.558	-0.053	0.027	0.142	0.050	-0.079	0.043	8.572	1.558	18.743	4.885
	算法改进 5	2.426	0.036	-0.051	0.009	0.145	0.014	-0.084	0.011	8.771	0.515	18.755	0.976

物参数的常理的。此时的估计可以认为是无效估计，因为并没有表征出在不同分位数水平下的人群之间存在的差异。而且该算法的估计也同样不够稳健，估计方差较大。与此同时，该算法得到的图像结果6-6显示在达到第二个折点位置之后的肱三头肌皮褶厚度并没有随着年龄的增长而趋于稳定，反倒是呈现着较为明显的上升趋势，这也与常理有一些不符。因此，对于该数据集，用算法4-3进行估计并不是一个好的选择。

事实上，不同的算法有着其不同的适用的数据集。但是由于文章篇幅的原因，就不再对其余的实际数据集进行分析了。在实际数据分析中，对于不同的数据集，需要具体问题具体分析，选择合适的算法进行回归分析。

7 结论与展望

本文针对折点回归模型进行了系统性的研究。从单折点的回归模型出发,先对单折点均值回归模型进行了研究,基于均值回归模型在数据分析过程中存在的致命缺陷,将研究重心转移到折点分位数回归模型上来。而对于折点分位数回归模型,同样从单折点的情形出发,向多折点分位数回归模型过渡,进而系统地从折点数估计、参数估计、方差估计以及统计推断等多个方面研究多折点分位数回归模型。随后,本文基于前人提出的一些方法,对多折点分位数回归模型的估计算法进行了一些改进,将主要的改进重点放在了针对原有模型的参数估计的损失函数不可导的缺点上,提出了一些针对性的光滑化算法改进。这些算法都存在着理论依据。

针对前人已经提出的一些折点回归模型以及其中涉及到的和折点效应检验、置信区间构造以及折点数估计等相关的算法,本文通过一些数值模拟试验验证了这些算法存在的合理性,以及其性能的优劣。与此同时,对原有的模型参数估计算法和本文进行改进的参数估计算法,进行了大量的模拟仿真试验,通过这些模拟试验研究了这些算法的性能,并对这些算法处理数据的能力进行评估。本文分别模拟了折点数为 $K = 1, 2, 3$ 时,在同方差情形和异方差情形下,误差项来自三种不同分布的数据,基于不同的分位数水平,通过估计的偏、实际标准差、估计标准差以及均方误差对这些算法的性能进行评估,比较了这些算法的优劣。

最后,运用我国不同城市的 GDP 与产业比重数据集和冈比亚女性肱三头肌皮褶厚度数据集这两个实际数据集对这些算法进行了实证分析。这两个实际数据集分别对应模型的单折点情形和多折点情形,进而说明了这些算法存在的优劣性以及实际意义。实际数据分析结果表明,本文所提出的模型和算法能够全面有效地挖掘出数据的折点信息。而且不同算法对于不同数据集的效果存在优劣,对于实际数据集而言,需要具体问题具体分析,选择合适的算法解决问题。

在本文的研究基础上,还可以考虑以下几个问题:一是考虑在数据集存在折点情形下,广义线性模型和非线性分位数回归模型的构建。本文主要针对的是线性的情形,但是线性模型本身还是存在着一些局限性,可以考虑通过识别出折点后利用分段的广义线性模型与非线性模型去解决问题。二是多折点分位数回归模型对自相关误差数据的研究。本文中的模拟仿真试验假设的误差并不相关,但是实际数据当中,误差项自相关的情形是不可避免的,这一点值得深入研究。最后可以考虑回归曲线在折点处发生跳跃时如何估计参数的问题。考虑这一问题同样可以挖掘出很多有用的统计理论与估计算法,这是本文研究工作的另一个扩展方向。

致 谢

时光荏苒，白驹过隙。论文搁笔时是在图书馆的四楼，侧望窗外悠远的天空，关于大学四年的思绪与情感如潮水般涌上心头，撞击着年轻而又炽热的心灵。这一路历程，仿佛一场电影，而此刻的我站在影院的出口，回望那荧幕，眼前闪过的，无论愉悦还是疼殇，都徜徉于内心。唯有此刻，送上最后的鞠躬致谢。

首先感谢我的父母，于丹峰先生与季桦女士。感谢他们在我生命的前二十一年对我无微不至的照料以及悉心的教导。在我成长的过程中，他们为我提供了无忧无虑的环境，使我可以专注于眼前的目标而无后顾之忧；他们给我灌输的教育使我少走了许多弯路，总是帮助我做出正确的选择。正是他们付出一切形成我坚实的后盾，我才能逾越生活与学习中的困难，开启未来的精彩人生。

接着感谢大学四年所有对我悉心教导，循循善诱的老师。感谢我的毕设导师付利亚老师，能遇见这位老师是我此生的荣幸。她高超的授课技巧，认真严谨的学术态度以及平易近人的性格都潜移默化地影响着我，让我领略到统计学领域的博大精深，收获内心那份关于学习与生活的感动。感谢我的辅导员沈飞跃老师对我学习生活中的照顾与帮助以及遇到困难时的开导。除此之外，感谢程晓青、张春霞、段启宏、陆亚明、杨家青、孟德宇、惠永昌、王立周、李辉、朱学虎等老师在学习中对我的答疑、帮助与指导，让我受益匪浅。

还要感谢大学四年与我共同奋斗的朋友们。感谢我的两位室友石志豪与罗海宸，与你们同寝的四年时光永远是我快乐的回忆。同时感谢史长浩、吴致远、宋程、连之瑞、林太阳、刘乾、周彦、杨一鸣、郑宗焜、谢文轩、赵振华、王新宇、李若涵、高赞程等人在学习与生活中对我的帮助。感谢你们与我共度这些阳光灿烂的日子，感谢你们对我的包容、支持与鼓励。希望我们多年以后仍不会忘记彼此。

最后感谢我的母校——西安交通大学，感谢栽培我的数学与统计学院。在这里我收获了真才实学，收获了属于我的青春回忆。毕业在即，我会铭记师长们的教诲，继续不懈努力，未来为社会作出贡献，报答支持与帮助过我的人。愿西安交通大学再创辉煌，愿老师们心想事成，愿同学们前程似锦。

参考文献

- [1] Page E S. Continuous inspection schemes[J]. *Biometrika*, 1954, 41(1):100-115.
- [2] Garland T. The relation between maximal running speed and body mass in terrestrial mammals[J]. *Journal of Zoology*, 1983, 199(2):157-170.
- [3] Hansen J, Lebedeff S. Global trends of measured surface air temperature[J]. *Journal of Geophysical Research: Atmospheres*, 1987, 92(D11):13345-13372.
- [4] Koenker R, Schorfheide F. Quantile spline models for global temperature change[J]. *Climatic Change*, 1994, 28(4):395-404.
- [5] Quandt R E. The estimation of the parameters of a linear regression system obeying two separate regimes[J]. *Journal of the American Statistical Association*, 1958, 53(284):873-880.
- [6] Quandt R E. Tests of the hypothesis that a linear regression system obeys two separate regimes[J]. *Journal of the American Statistical Association*, 1960, 55(290):324-330.
- [7] Sprent P. Some hypotheses concerning two phase regression lines[J]. *Biometrics*, 1961, 17(4):634-645.
- [8] Hinkley D V. Inference about the intersection in two-phase regression[J]. *Biometrika*, 1969, 56(3):495-504.
- [9] Hinkley D V. Inference in two-phase regression[J]. *Journal of the American Statistical Association*, 1971, 66(336):736-743.
- [10] Feder P I. On asymptotic distribution theory in segmented regression problems—identified case[J]. *The Annals of Statistics*, 1975, 3(1):49-83.
- [11] Chappell R. Fitting bent lines to data, with applications to allometry[J]. *Journal of Theoretical Biology*, 1989, 138(2):235-256.
- [12] Tong H. Threshold models in non-linear time series analysis: volume 21[M]. Springer Science & Business Media, 2012.
- [13] Tong H. Non-linear time series: a dynamical system approach[M]. Oxford university press, 1990.
- [14] Chan K S, Tsay R S. Limiting properties of the least squares estimator of a continuous threshold autoregressive model[J]. *Biometrika*, 1998, 85(2):413-426.
- [15] Hansen B E. Inference when a nuisance parameter is not identified under the null hypothesis[J]. *Econometrica: Journal of the Econometric Society*, 1996, 64(2):413-430.
- [16] Lee S, Seo M H, Shin Y. Testing for threshold effects in regression models[J]. *Journal of the American Statistical Association*, 2011, 106(493):220-231.
- [17] Chiu G, Lockhart R, Routledge R. Asymptotic theory for bent-cable regression—the basic case[J]. *Journal of Statistical Planning and Inference*, 2005, 127(1-2):143-156.
- [18] Chiu G, Lockhart R, Routledge R. Bent-cable regression theory and applications[J]. *Journal of the American Statistical Association*, 2006, 101(474):542-553.
- [19] Tishler A, Zang I. A new maximum likelihood algorithm for piecewise regression[J]. *Journal of the American Statistical Association*, 1981, 76(376):980-987.
- [20] Das R, Banerjee M, Nan B, et al. Fast estimation of regression parameters in a broken-stick model for longitudinal data[J]. *Journal of the American Statistical Association*, 2016, 111(515):1132-1143.
- [21] Hansen B E. Regression kink with an unknown threshold[J]. *Journal of Business & Economic Statistics*, 2017, 35(2):228-240.
- [22] Li C, Wei Y, He X, et al. Bent line quantile regression with application to an allometric study of land mammals' speed and mass[J]. *Biometrics*, 2011, 67(1):242-249.

- [23] Koenker R, Bassett Jr G. Regression quantiles[J]. *Econometrica: Journal of the Econometric Society*, 1978, 46(1):33-50.
- [24] Zhong W, Wan C, Zhang W. Estimation and inference for multi-kink quantile regression[J]. *Journal of Business & Economic Statistics*, 2021, 0(0):1-17.
- [25] Muggeo V M. Estimating regression models with unknown break-points[J]. *Statistics in Medicine*, 2003, 22(19):3055-3071.
- [26] Efron B. The jackknife, the bootstrap and other resampling plans[M]. SIAM, 1982.
- [27] Zhou X, Zhang F. Bent line quantile regression via a smoothing technique[J]. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 2020, 13(3):216-228.
- [28] Zhang F, Li Q. A continuous threshold expectile model[J]. *Computational Statistics & Data Analysis*, 2017, 116:49-66.
- [29] Lerman P. Fitting segmented regression models by grid search[J]. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 1980, 29(1):77-84.
- [30] Shapiro A. On concepts of directional differentiability[J]. *Journal of Optimization Theory and Applications*, 1990, 66(3):477-487.
- [31] Fang Z, Santos A. Inference on directionally differentiable functions[J]. *The Review of Economic Studies*, 2019, 86(1):377-412.
- [32] Hong H, Li J. The numerical directional delta method[J]. Working Paper, 2015.
- [33] Koenker R, Bassett G. Quantile regression, 2005[J]. Google Scholar Cross Ref.
- [34] He X, Zhu L X. A lack-of-fit test for quantile regression[J]. *Journal of the American Statistical Association*, 2003, 98(464):1013-1022.
- [35] Koenker R, Bassett G. Tests of linear hypotheses and l^1 estimation[J]. *Econometrica: Journal of the Econometric Society*, 1982, 50(6):1577-1583.
- [36] White H. Maximum likelihood estimation of misspecified models[J]. *Econometrica: Journal of the Econometric Society*, 1982, 50(1):1-25.
- [37] Royston P, Sauerbrei W. Multivariable model-building: a pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables: volume 777[M]. John Wiley & Sons, 2008.
- [38] Akaike H. A bayesian extension of the minimum aic procedure of autoregressive model fitting[J]. *Biometrika*, 1979, 66(2):237-242.
- [39] Hendricks W, Koenker R. Hierarchical spline models for conditional quantiles and the demand for electricity[J]. *Journal of the American Statistical Association*, 1992, 87(417):58-68.
- [40] Hall P, Sheather S J. On the distribution of a studentized quantile[J]. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1988, 50(3):381-391.
- [41] Bofingeb E. Estimation of a density function using order statistics 1[J]. *Australian Journal of Statistics*, 1975, 17(1):1-7.
- [42] Muggeo V M. Interval estimation for the breakpoint in segmented regression: A smoothed score-based approach[J]. *Australian & New Zealand Journal of Statistics*, 2017, 59(3):311-322.
- [43] Aizerman M A, Braverman E M, Rozonoer L I. Theoretical foundation of potential functions method in pattern recognition[J]. *Avtomat. I Telemekh*, 1964, 25(6):917-936.
- [44] Horowitz J L. A smoothed maximum score estimator for the binary response model[J]. *Econometrica: Journal of the Econometric Society*, 1992, 60(3):505-531.
- [45] Zheng S. Gradient descent algorithms for quantile regression with smooth approximation[J]. *International Journal of Machine Learning and Cybernetics*, 2011, 2(3):191-207.

- [46] Friedman J H. Greedy function approximation: a gradient boosting machine[J]. *Annals of Statistics*, 2001, 29(5):1189-1232.
- [47] Mason L, Baxter J, Bartlett P, et al. Boosting algorithms as gradient descent[J]. *Advances in Neural Information Processing Systems*, 1999, 12:512-518.
- [48] Brown B M, Wang Y G. Standard errors and covariance matrices for smoothed rank estimators[J]. *Biometrika*, 2005, 92(1):149-158.
- [49] Fu L, Wang Y G. Quantile regression for longitudinal data with a working correlation model[J]. *Computational Statistics & Data Analysis*, 2012, 56(8):2526-2538.
- [50] Pang L, Lu W, Wang H J. Variance estimation in censored quantile regression via induced smoothing [J]. *Computational Statistics & Data Analysis*, 2012, 56(4):785-796.
- [51] Durnin J V, Womersley J. Body fat assessed from total body density and its estimation from skinfold thickness: measurements on 481 men and women aged from 16 to 72 years[J]. *British Journal of Nutrition*, 1974, 32(1):77-97.
- [52] Cole T J, Green P J. Smoothing reference centile curves: the lms method and penalized likelihood[J]. *Statistics in Medicine*, 1992, 11(10):1305-1319.

附录 A 正文未展示的模拟试验结果展示

表 A-1 折点数 $K = 3$ ，同方差 $e_i \sim \mathbf{N}(0, 1)$ 情形下各方法不同分位数模拟结果比较

τ	Method		α_0	α_1	γ	β_1	β_2	β_3	δ_1	δ_2	δ_3
$\tau = 0.25$	MKQR	Bias	0.069	0.000	0.016	-0.023	0.010	-0.007	-0.002	-0.001	-0.001
		SD	0.718	0.044	0.176	0.195	0.131	0.199	0.092	0.061	0.069
		SE	0.675	0.041	0.165	0.191	0.130	0.190	0.083	0.056	0.062
		MSE	0.461	0.002	0.028	0.037	0.017	0.036	0.007	0.003	0.004
	算法改进 4	Bias	-0.197	-0.043	0.001	0.060	-0.036	0.062	0.016	-0.001	-0.009
		SD	0.712	0.174	0.041	0.185	0.126	0.183	0.096	0.059	0.069
		SE	0.723	0.375	0.430	0.291	0.427	0.096	0.553	0.170	0.139
		MSE	0.497	0.143	0.185	0.088	0.184	0.013	0.306	0.029	0.019
	算法改进 5	Bias	-0.439	-0.094	0.032	0.078	-0.048	0.162	0.017	0.005	-0.007
		SD	0.719	0.175	0.045	0.210	0.128	0.200	0.094	0.048	0.068
		SE	0.720	0.177	0.203	0.138	0.214	0.045	0.262	0.078	0.071
		MSE	0.711	0.040	0.042	0.025	0.048	0.028	0.069	0.006	0.005
$\tau = 0.5$	MKQR	Bias	0.053	0.000	0.013	-0.020	0.014	-0.016	0.001	0.002	-0.001
		SD	0.655	0.041	0.160	0.177	0.119	0.175	0.085	0.056	0.062
		SE	0.622	0.038	0.154	0.176	0.119	0.176	0.076	0.051	0.057
		MSE	0.389	0.001	0.024	0.031	0.014	0.031	0.006	0.003	0.003
	算法改进 4	Bias	-0.215	-0.050	0.002	0.064	-0.038	0.073	0.019	-0.003	-0.010
		SD	0.674	0.165	0.036	0.173	0.119	0.170	0.095	0.060	0.069
		SE	1.225	0.304	0.349	0.238	0.346	0.079	0.452	0.139	0.113
		MSE	1.546	0.095	0.122	0.061	0.121	0.012	0.204	0.019	0.013
	算法改进 5	Bias	-0.218	-0.054	0.002	0.054	0.007	0.064	0.010	0.001	-0.009
		SD	0.568	0.139	0.038	0.157	0.111	0.155	0.082	0.051	0.068
		SE	0.655	0.161	0.184	0.129	0.196	0.042	0.245	0.073	0.066
		MSE	0.476	0.029	0.034	0.020	0.039	0.006	0.060	0.005	0.004
$\tau = 0.75$	MKQR	Bias	0.031	-0.001	0.006	-0.010	0.011	-0.014	0.002	0.001	-0.000
		SD	0.746	0.043	0.182	0.204	0.130	0.195	0.093	0.058	0.068
		SE	0.675	0.041	0.167	0.191	0.129	0.190	0.083	0.056	0.062
		MSE	0.456	0.002	0.028	0.037	0.017	0.036	0.007	0.003	0.004
	算法改进 4	Bias	-0.196	-0.048	0.000	0.058	-0.029	0.074	0.022	-0.002	-0.014
		SD	0.694	0.170	0.040	0.178	0.127	0.188	0.098	0.059	0.072
		SE	1.524	0.377	0.435	0.301	0.437	0.098	0.569	0.174	0.143
		MSE	2.361	0.145	0.189	0.094	0.192	0.015	0.324	0.030	0.021
	算法改进 5	Bias	-0.603	-0.151	-0.015	0.181	0.018	0.185	0.122	-0.001	-0.004
		SD	0.665	0.165	0.049	0.212	0.163	0.191	0.096	0.054	0.066
		SE	0.705	0.174	0.199	0.137	0.210	0.044	0.266	0.082	0.070
		MSE	0.862	0.053	0.040	0.051	0.045	0.036	0.070	0.007	0.005

表 A-2 折点数 $K = 3$, 同方差 $e_i \sim \mathbf{t}(3)$ 情形下各方法不同分位数模拟结果比较

τ	Method		α_0	α_1	γ	β_1	β_2	β_3	δ_1	δ_2	δ_3
$\tau = 0.25$	MKQR	Bias	0.026	-0.002	0.006	-0.016	0.018	-0.018	0.007	0.001	-0.003
		SD	0.924	0.052	0.226	0.242	0.165	0.244	0.124	0.075	0.085
		SE	0.844	0.051	0.209	0.240	0.162	0.241	0.104	0.070	0.078
		MSE	0.713	0.003	0.044	0.058	0.027	0.059	0.011	0.005	0.006
	算法改进 4	Bias	-0.428	-0.096	-0.007	0.151	-0.128	0.170	0.041	-0.005	-0.021
		SD	0.894	0.225	0.073	0.278	0.263	0.308	0.206	0.072	0.186
		SE	1.847	0.462	0.524	0.356	0.521	0.118	0.685	0.217	0.175
		MSE	3.594	0.222	0.275	0.149	0.288	0.043	0.471	0.047	0.031
	算法改进 5	Bias	-0.685	-0.149	0.023	0.119	-0.052	0.230	0.041	-0.006	0.003
		SD	0.792	0.198	0.047	0.243	0.171	0.227	0.115	0.078	0.093
		SE	0.888	0.220	0.251	0.168	0.258	0.055	0.326	0.096	0.086
		MSE	1.256	0.071	0.064	0.042	0.069	0.056	0.108	0.009	0.007
	MKQR	Bias	0.059	-0.001	0.013	-0.022	0.013	-0.006	0.000	-0.002	-0.001
		SD	0.703	0.043	0.173	0.188	0.131	0.193	0.093	0.058	0.068
		SE	0.684	0.042	0.169	0.193	0.131	0.192	0.084	0.056	0.063
		MSE	0.472	0.002	0.029	0.038	0.017	0.037	0.007	0.003	0.004
	算法改进 4	Bias	-0.464	-0.109	-0.007	0.175	-0.148	0.191	0.056	-0.004	-0.035
		SD	0.799	0.208	0.093	0.309	0.350	0.380	0.275	0.074	0.264
		SE	1.305	0.325	0.370	0.255	0.372	0.086	0.490	0.155	0.127
		MSE	1.918	0.117	0.137	0.096	0.160	0.044	0.243	0.024	0.017
	算法改进 5	Bias	-0.317	-0.074	-0.004	0.063	0.009	0.065	0.030	-0.007	-0.001
		SD	0.694	0.172	0.037	0.195	0.127	0.193	0.099	0.062	0.071
		SE	0.715	0.177	0.203	0.141	0.218	0.046	0.269	0.080	0.073
		MSE	0.612	0.037	0.041	0.024	0.048	0.006	0.073	0.006	0.005
$\tau = 0.5$	MKQR	Bias	0.018	0.002	0.004	-0.014	0.020	-0.021	0.006	0.001	-0.001
		SD	0.895	0.053	0.218	0.244	0.169	0.240	0.117	0.079	0.088
		SE	0.847	0.051	0.209	0.241	0.164	0.240	0.105	0.070	0.078
		MSE	0.718	0.003	0.044	0.058	0.027	0.058	0.011	0.005	0.006
	算法改进 4	Bias	-0.468	-0.114	-0.002	0.171	-0.128	0.188	0.040	-0.002	-0.023
		SD	0.882	0.222	0.068	0.261	0.236	0.280	0.183	0.073	0.160
		SE	1.877	0.468	0.533	0.368	0.536	0.121	0.702	0.225	0.182
		MSE	3.744	0.232	0.284	0.164	0.304	0.050	0.495	0.051	0.034
	算法改进 5	Bias	-0.668	-0.169	-0.025	0.208	-0.001	0.213	0.016	-0.003	-0.002
		SD	0.752	0.192	0.055	0.260	0.185	0.251	0.106	0.087	0.087
		SE	0.881	0.217	0.249	0.175	0.258	0.055	0.340	0.104	0.088
		MSE	1.223	0.076	0.063	0.074	0.066	0.048	0.116	0.011	0.008
$\tau = 0.75$	MKQR	Bias	0.018	0.002	0.004	-0.014	0.020	-0.021	0.006	0.001	-0.001
		SD	0.895	0.053	0.218	0.244	0.169	0.240	0.117	0.079	0.088
		SE	0.847	0.051	0.209	0.241	0.164	0.240	0.105	0.070	0.078
		MSE	0.718	0.003	0.044	0.058	0.027	0.058	0.011	0.005	0.006
	算法改进 4	Bias	-0.468	-0.114	-0.002	0.171	-0.128	0.188	0.040	-0.002	-0.023
		SD	0.882	0.222	0.068	0.261	0.236	0.280	0.183	0.073	0.160
		SE	1.877	0.468	0.533	0.368	0.536	0.121	0.702	0.225	0.182
		MSE	3.744	0.232	0.284	0.164	0.304	0.050	0.495	0.051	0.034
	算法改进 5	Bias	-0.668	-0.169	-0.025	0.208	-0.001	0.213	0.016	-0.003	-0.002
		SD	0.752	0.192	0.055	0.260	0.185	0.251	0.106	0.087	0.087
		SE	0.881	0.217	0.249	0.175	0.258	0.055	0.340	0.104	0.088
		MSE	1.223	0.076	0.063	0.074	0.066	0.048	0.116	0.011	0.008

表 A-3 折点数 $K = 3$, 同方差 $e_i \sim \mathbf{LA}(0, 1)$ 情形下各方法不同分位数模拟结果比较

τ	Method		α_0	α_1	γ	β_1	β_2	β_3	δ_1	δ_2	δ_3
$\tau = 0.25$	MKQR	Bias	0.052	0.003	0.014	-0.021	0.015	0.018	0.001	-0.001	-0.003
		SD	0.933	0.056	0.228	0.244	0.173	0.254	0.122	0.079	0.087
		SE	0.864	0.052	0.214	0.246	0.166	0.246	0.107	0.071	0.080
		MSE	0.749	0.003	0.046	0.061	0.028	0.061	0.011	0.005	0.006
	算法改进 4	Bias	-0.368	-0.078	0.003	0.110	-0.074	0.127	0.030	-0.002	-0.019
		SD	0.909	0.220	0.052	0.229	0.163	0.238	0.131	0.077	0.092
		SE	1.895	0.473	0.540	0.368	0.541	0.121	0.698	0.219	0.179
		MSE	3.725	0.229	0.292	0.148	0.299	0.031	0.488	0.048	0.032
	算法改进 5	Bias	-0.717	-0.151	0.026	0.143	-0.056	0.245	0.020	0.007	-0.020
		SD	0.809	0.194	0.053	0.242	0.181	0.263	0.130	0.068	0.099
		SE	0.879	0.218	0.249	0.167	0.242	0.054	0.315	0.100	0.080
		MSE	1.286	0.070	0.063	0.048	0.062	0.063	0.100	0.010	0.007
$\tau = 0.5$	MKQR	Bias	0.021	0.001	0.005	-0.006	0.002	-0.010	-0.001	-0.001	0.002
		SD	0.596	0.034	0.196	0.159	0.104	0.159	0.073	0.047	0.053
		SE	0.561	0.034	0.139	0.158	0.106	0.157	0.069	0.046	0.051
		MSE	0.316	0.001	0.019	0.025	0.011	0.025	0.005	0.002	0.003
	算法改进 4	Bias	-0.377	-0.087	0.002	0.121	-0.077	0.133	0.030	-0.002	-0.019
		SD	0.823	0.197	0.034	0.185	0.133	0.187	0.131	0.077	0.092
		SE	1.091	0.271	0.311	0.214	0.311	0.071	0.405	0.127	0.104
		MSE	1.331	0.081	0.096	0.060	0.123	0.023	0.165	0.016	0.011
	算法改进 5	Bias	-0.225	-0.054	-0.003	0.061	-0.014	0.071	0.007	-0.001	-0.003
		SD	0.537	0.132	0.033	0.155	0.110	0.155	0.072	0.044	0.060
		SE	0.658	0.161	0.183	0.125	0.197	0.041	0.248	0.069	0.067
		MSE	0.484	0.029	0.034	0.019	0.039	0.007	0.061	0.005	0.005
$\tau = 0.75$	MKQR	Bias	0.040	0.002	0.007	-0.013	0.017	-0.035	-0.001	0.001	0.002
		SD	0.911	0.053	0.223	0.243	0.169	0.249	0.116	0.081	0.085
		SE	0.863	0.052	0.213	0.245	0.166	0.246	0.107	0.071	0.080
		MSE	0.746	0.003	0.045	0.060	0.028	0.062	0.011	0.005	0.006
	算法改进 4	Bias	-0.365	-0.090	0.002	0.123	-0.074	0.126	0.030	-0.002	-0.019
		SD	0.896	0.221	0.049	0.226	0.165	0.236	0.131	0.077	0.092
		SE	1.940	0.482	0.553	0.381	0.550	0.124	0.722	0.227	0.184
		MSE	3.896	0.241	0.306	0.160	0.308	0.031	0.522	0.052	0.034
	算法改进 5	Bias	-0.684	-0.174	-0.020	0.225	-0.009	0.235	0.006	0.002	-0.005
		SD	0.663	0.170	0.050	0.233	0.212	0.227	0.108	0.078	0.092
		SE	0.876	0.215	0.248	0.171	0.258	0.055	0.337	0.105	0.085
		MSE	1.235	0.076	0.062	0.080	0.067	0.058	0.114	0.011	0.007

表 A-4 折点数 $K = 3$, 异方差 $e_i \sim \mathbf{N}(0, 1)$ 情形下各方法不同分位数模拟结果比较

τ	Method		α_0	α_1	γ	β_1	β_2	β_3	δ_1	δ_2	δ_3
$\tau = 0.25$	MKQR	Bias	0.131	0.000	0.004	-0.006	0.016	-0.022	0.000	0.003	-0.007
		SD	0.140	0.017	0.029	0.069	0.142	0.342	0.033	0.061	0.112
		SE	0.139	0.016	0.029	0.069	0.135	0.323	0.033	0.056	0.101
		MSE	0.019	0.000	0.001	0.005	0.019	0.105	0.001	0.003	0.010
	算法改进 4	Bias	-0.024	0.010	0.000	0.015	-0.035	0.074	0.002	0.001	-0.025
		SD	0.155	0.034	0.025	0.068	0.136	0.307	0.034	0.056	0.118
		SE	0.312	0.066	0.156	0.304	0.724	0.038	0.217	0.167	0.223
		MSE	0.098	0.004	0.024	0.093	0.525	0.007	0.047	0.028	0.050
	算法改进 5	Bias	-0.074	-0.009	-0.003	0.005	-0.023	0.077	0.000	0.009	-0.012
		SD	0.194	0.043	0.022	0.077	0.129	0.277	0.036	0.047	0.119
		SE	0.260	0.059	0.093	0.142	0.336	0.030	0.139	0.076	0.105
		MSE	0.073	0.004	0.009	0.020	0.114	0.007	0.019	0.006	0.011
	MKQR	Bias	0.002	-0.001	0.000	-0.004	0.011	-0.018	0.001	-0.001	-0.002
		SD	0.127	0.016	0.027	0.063	0.126	0.304	0.031	0.056	0.106
		SE	0.125	0.015	0.026	0.064	0.124	0.301	0.030	0.051	0.093
		MSE	0.016	0.000	0.001	0.004	0.015	0.091	0.001	0.003	0.009
	算法改进 4	Bias	-0.112	-0.025	0.000	0.052	-0.042	0.087	0.000	0.004	-0.027
		SD	0.161	0.035	0.022	0.068	0.129	0.293	0.033	0.061	0.114
		SE	0.249	0.052	0.126	0.249	0.587	0.030	0.177	0.139	0.182
		MSE	0.075	0.003	0.016	0.064	0.347	0.008	0.031	0.019	0.034
	算法改进 5	Bias	-0.122	-0.027	0.002	0.033	0.011	0.084	0.001	0.005	-0.019
		SD	0.139	0.030	0.020	0.064	0.122	0.273	0.032	0.055	0.114
		SE	0.249	0.056	0.090	0.135	0.314	0.027	0.135	0.072	0.097
		MSE	0.077	0.004	0.008	0.019	0.099	0.008	0.018	0.005	0.010
$\tau = 0.5$	MKQR	Bias	-0.017	0.000	-0.004	0.001	0.019	-0.051	0.002	0.003	-0.002
		SD	0.136	0.017	0.029	0.073	0.140	0.330	0.035	0.059	0.114
		SE	0.138	0.016	0.029	0.069	0.134	0.326	0.033	0.056	0.100
		MSE	0.019	0.000	0.001	0.005	0.018	0.109	0.001	0.003	0.010
	算法改进 4	Bias	-0.207	-0.062	-0.000	0.087	-0.034	0.080	0.001	0.006	-0.026
		SD	0.180	0.040	0.025	0.076	0.134	0.299	0.033	0.059	0.121
		SE	0.316	0.067	0.158	0.313	0.744	0.038	0.220	0.175	0.230
		MSE	0.143	0.008	0.025	0.106	0.554	0.008	0.048	0.030	0.054
	算法改进 5	Bias	-0.248	-0.060	0.006	0.079	0.024	0.177	0.002	0.003	-0.019
		SD	0.185	0.042	0.023	0.082	0.154	0.286	0.037	0.055	0.111
		SE	0.241	0.054	0.093	0.145	0.340	0.025	0.141	0.080	0.103
		MSE	0.119	0.006	0.009	0.027	0.116	0.032	0.020	0.006	0.011
$\tau = 0.75$	MKQR	Bias	-0.017	0.000	-0.004	0.001	0.019	-0.051	0.002	0.003	-0.002
		SD	0.136	0.017	0.029	0.073	0.140	0.330	0.035	0.059	0.114
		SE	0.138	0.016	0.029	0.069	0.134	0.326	0.033	0.056	0.100
		MSE	0.019	0.000	0.001	0.005	0.018	0.109	0.001	0.003	0.010
	算法改进 4	Bias	-0.207	-0.062	-0.000	0.087	-0.034	0.080	0.001	0.006	-0.026
		SD	0.180	0.040	0.025	0.076	0.134	0.299	0.033	0.059	0.121
		SE	0.316	0.067	0.158	0.313	0.744	0.038	0.220	0.175	0.230
		MSE	0.143	0.008	0.025	0.106	0.554	0.008	0.048	0.030	0.054
	算法改进 5	Bias	-0.248	-0.060	0.006	0.079	0.024	0.177	0.002	0.003	-0.019
		SD	0.185	0.042	0.023	0.082	0.154	0.286	0.037	0.055	0.111
		SE	0.241	0.054	0.093	0.145	0.340	0.025	0.141	0.080	0.103
		MSE	0.119	0.006	0.009	0.027	0.116	0.032	0.020	0.006	0.011

表 A-5 折点数 $K = 3$, 异方差 $e_i \sim t(3)$ 情形下各方法不同分位数模拟结果比较

τ	Method		α_0	α_1	γ	β_1	β_2	β_3	δ_1	δ_2	δ_3
$\tau = 0.25$	MKQR	Bias	0.017	0.000	0.005	-0.008	0.035	-0.058	-0.001	0.006	-0.012
		SD	0.175	0.020	0.036	0.086	0.172	0.419	0.039	0.072	0.147
		SE	0.174	0.020	0.037	0.087	0.170	0.409	0.041	0.070	0.125
		MSE	0.030	0.000	0.001	0.008	0.030	0.170	0.002	0.005	0.016
	算法改进 4	Bias	-0.205	-0.032	-0.006	0.108	-0.126	0.179	0.018	0.006	-0.056
		SD	0.203	0.081	0.076	0.222	0.321	0.491	0.216	0.073	0.263
		SE	0.370	0.079	0.189	0.374	0.889	0.047	0.263	0.214	0.280
		MSE	0.179	0.007	0.036	0.152	0.805	0.034	0.069	0.046	0.082
	算法改进 5	Bias	-0.143	-0.024	-0.010	0.006	-0.029	0.150	0.005	0.003	-0.032
		SD	0.234	0.051	0.022	0.102	0.193	0.325	0.042	0.077	0.170
		SE	0.281	0.063	0.106	0.175	0.418	0.033	0.156	0.093	0.131
		MSE	0.099	0.005	0.011	0.031	0.176	0.024	0.024	0.009	0.018
$\tau = 0.5$	MKQR	Bias	0.004	0.000	0.001	-0.003	0.009	-0.028	0.000	0.001	-0.029
		SD	0.137	0.018	0.029	0.069	0.137	0.334	0.034	0.057	0.114
		SE	0.139	0.016	0.029	0.070	0.136	0.327	0.033	0.056	0.099
		MSE	0.019	0.000	0.001	0.005	0.019	0.108	0.001	0.003	0.010
	算法改进 4	Bias	-0.290	-0.069	-0.007	0.175	-0.148	0.191	0.056	-0.004	-0.035
		SD	0.210	0.090	0.087	0.252	0.357	0.461	0.254	0.070	0.289
		SE	0.268	0.056	0.135	0.269	0.627	0.033	0.191	0.154	0.201
		MSE	0.156	0.008	0.018	0.095	0.412	0.056	0.037	0.024	0.046
	算法改进 5	Bias	-0.139	-0.031	-0.003	0.029	0.010	0.083	0.006	0.000	-0.014
		SD	0.159	0.035	0.020	0.077	0.135	0.297	0.033	0.064	0.126
		SE	0.269	0.060	0.098	0.149	0.347	0.029	0.153	0.079	0.108
		MSE	0.092	0.005	0.010	0.023	0.120	0.008	0.023	0.006	0.012
$\tau = 0.75$	MKQR	Bias	-0.019	0.001	-0.005	0.002	0.020	-0.060	0.003	0.004	-0.000
		SD	0.177	0.022	0.037	0.090	0.176	0.443	0.045	0.077	0.152
		SE	0.174	0.020	0.037	0.087	0.169	0.412	0.041	0.070	0.125
		MSE	0.031	0.000	0.001	0.008	0.029	0.173	0.002	0.005	0.016
	算法改进 4	Bias	-0.374	-0.102	-0.002	0.175	-0.120	0.215	0.016	0.010	-0.062
		SD	0.227	0.075	0.070	0.203	0.298	0.454	0.194	0.078	0.243
		SE	0.386	0.082	0.195	0.388	0.906	0.048	0.273	0.225	0.290
		MSE	0.289	0.017	0.038	0.181	0.836	0.049	0.075	0.051	0.088
	算法改进 5	Bias	-0.290	-0.070	0.000	0.091	0.019	0.197	0.006	0.001	-0.014
		SD	0.204	0.046	0.024	0.116	0.183	0.375	0.040	0.086	0.154
		SE	0.269	0.059	0.109	0.183	0.414	0.027	0.164	0.101	0.130
		MSE	0.157	0.008	0.012	0.042	0.172	0.039	0.027	0.010	0.017

表 A-6 折点数 $K = 3$, 异方差 $e_i \sim \mathbf{LA}(0, 1)$ 情形下各方法不同分位数模拟结果比较

τ	Method		α_0	α_1	γ	β_1	β_2	β_3	δ_1	δ_2	δ_3
$\tau = 0.25$	MKQR	Bias	0.017	0.001	0.005	-0.006	0.017	-0.049	-0.002	0.003	-0.005
		SD	0.174	0.022	0.037	0.089	0.182	0.437	0.042	0.078	0.144
		SE	0.178	0.021	0.038	0.089	0.172	0.422	0.041	0.071	0.128
		MSE	0.032	0.000	0.001	0.008	0.030	0.180	0.002	0.005	0.017
	算法改进 4	Bias	-0.153	-0.016	0.002	0.069	-0.070	0.137	0.001	0.007	-0.050
		SD	0.185	0.040	0.029	0.081	0.174	0.421	0.043	0.076	0.157
		SE	0.392	0.083	0.195	0.387	0.919	0.048	0.273	0.216	0.286
		MSE	0.177	0.007	0.038	0.154	0.849	0.021	0.074	0.047	0.084
	算法改进 5	Bias	-0.176	-0.030	-0.007	0.018	-0.019	0.152	0.001	0.012	-0.053
		SD	0.240	0.053	0.024	0.100	0.184	0.348	0.050	0.066	0.170
		SE	0.266	0.060	0.103	0.175	0.402	0.033	0.143	0.096	0.126
		MSE	0.102	0.005	0.011	0.031	0.162	0.024	0.021	0.009	0.019
	MKQR	Bias	0.001	0.000	0.000	0.000	0.001	-0.023	-0.001	0.000	0.004
		SD	0.111	0.014	0.023	0.056	0.107	0.269	0.026	0.047	0.086
		SE	0.113	0.013	0.024	0.057	0.110	0.268	0.027	0.046	0.082
		MSE	0.013	0.000	0.001	0.003	0.012	0.073	0.001	0.002	0.007
	算法改进 4	Bias	-0.213	-0.047	0.001	0.096	-0.070	0.158	0.001	0.007	-0.050
		SD	0.196	0.044	0.021	0.076	0.141	0.325	0.043	0.076	0.157
		SE	0.225	0.047	0.112	0.225	0.522	0.027	0.158	0.126	0.165
		MSE	0.096	0.004	0.013	0.060	0.278	0.026	0.025	0.016	0.030
	算法改进 5	Bias	-0.110	-0.024	-0.002	0.031	-0.014	0.092	-0.001	-0.001	-0.008
		SD	0.113	0.024	0.017	0.064	0.115	0.241	0.026	0.043	0.094
		SE	0.277	0.061	0.095	0.130	0.291	0.027	0.156	0.068	0.093
		MSE	0.089	0.004	0.009	0.018	0.085	0.009	0.024	0.005	0.009
$\tau = 0.5$	MKQR	Bias	-0.014	0.001	-0.004	0.003	0.016	-0.086	0.000	0.003	0.008
		SD	0.173	0.021	0.036	0.089	0.180	0.441	0.042	0.081	0.150
		SE	0.176	0.021	0.037	0.089	0.173	0.424	0.042	0.071	0.128
		MSE	0.031	0.000	0.001	0.008	0.030	0.187	0.002	0.005	0.016
	算法改进 4	Bias	-0.267	-0.077	0.001	0.120	-0.062	0.148	0.001	0.007	-0.050
		SD	0.221	0.050	0.028	0.092	0.173	0.408	0.043	0.076	0.157
		SE	0.400	0.085	0.200	0.398	0.930	0.049	0.280	0.266	0.292
		MSE	0.232	0.013	0.040	0.173	0.868	0.024	0.078	0.051	0.088
	算法改进 5	Bias	-0.290	-0.070	0.002	0.103	0.014	0.251	-0.002	0.009	-0.020
		SD	0.205	0.045	0.022	0.121	0.205	0.315	0.038	0.078	0.164
		SE	0.285	0.064	0.110	0.180	0.431	0.027	0.162	0.100	0.132
		MSE	0.165	0.009	0.012	0.043	0.186	0.064	0.026	0.010	0.018
$\tau = 0.75$	MKQR	Bias	-0.014	0.001	-0.004	0.003	0.016	-0.086	0.000	0.003	0.008
		SD	0.173	0.021	0.036	0.089	0.180	0.441	0.042	0.081	0.150
		SE	0.176	0.021	0.037	0.089	0.173	0.424	0.042	0.071	0.128
		MSE	0.031	0.000	0.001	0.008	0.030	0.187	0.002	0.005	0.016
	算法改进 4	Bias	-0.267	-0.077	0.001	0.120	-0.062	0.148	0.001	0.007	-0.050
		SD	0.221	0.050	0.028	0.092	0.173	0.408	0.043	0.076	0.157
		SE	0.400	0.085	0.200	0.398	0.930	0.049	0.280	0.266	0.292
		MSE	0.232	0.013	0.040	0.173	0.868	0.024	0.078	0.051	0.088
	算法改进 5	Bias	-0.290	-0.070	0.002	0.103	0.014	0.251	-0.002	0.009	-0.020
		SD	0.205	0.045	0.022	0.121	0.205	0.315	0.038	0.078	0.164
		SE	0.285	0.064	0.110	0.180	0.431	0.027	0.162	0.100	0.132
		MSE	0.165	0.009	0.012	0.043	0.186	0.064	0.026	0.010	0.018

附录 B 外文文献原文

Regression Kink With an Unknown Threshold

Bruce E. HANSEN

Department of Economics, University of Wisconsin, Madison, WI 53706 (behansen@wisc.edu)

This article explores estimation and inference in a regression kink model with an unknown threshold. A regression kink model (or continuous threshold model) is a threshold regression constrained to be everywhere continuous with a kink at an unknown threshold. We present methods for estimation, to test for the presence of the threshold, for inference on the regression parameters, and for inference on the regression function. A novel finding is that inference on the regression function is nonstandard since the regression function is a nondifferentiable function of the parameters. We apply recently developed methods for inference on nondifferentiable functions. The theory is illustrated by an application to the growth and debt problem introduced by Reinhart and Rogoff, using their long-span time-series for the United States.

KEY WORDS: Model selection/variable selection; Nonlinear; Nonparametric methods.

1. INTRODUCTION

The regression kink model was popularized by Card et al. (2012) as a modification of the regression discontinuity model. In the regression kink model, the regression function is continuous but the slope has a discontinuity at a threshold point, hence a “kink.” This model has gained much empirical attention, including applications by Landaís (2015) and Ganong and Jager (2014). The traditional regression discontinuity model assumes that the threshold is known, but in some cases (such as Card, Mas, and Rothstein 2008) it is unknown and must be estimated. Our article concerns this latter case of an unknown threshold.

Regression discontinuity models are similar to threshold regression models. The latter were introduced by Tong (1983, 1990) in the context of nonlinear autoregression, but can be applied to many nonlinear regression contexts. Most of the literature and methods focus on the discontinuous (unconstrained) threshold model, where the regression model is split into two (or more) groups based on a threshold indicator. A notable exception is the continuous threshold model introduced by Chan and Tsay (1998), which is identical to a regression kink model with piecewise linear regression segments. One economic application of the continuous threshold model is Cox, Hansen, and Jimenez (2004). The regression kink model may be appealing for empirical applications where the threshold effect focuses on one variable and there is no reason to expect a discontinuous regression response at the threshold.

This article extends the theory of Chan and Tsay (1998), considering the problems of testing for a threshold effect, inference on the regression parameters, and inference on the regression function. As in Chan and Tsay (1998), we confine attention to the context where the regression segments are linear rather than nonparametric. This is appropriate in contexts of moderate sample sizes where nonparametric methods may be inappropriate.

There is a large literature on discontinuous threshold regression. For the problem of testing for a threshold effect, relevant contributions include Chan (1990, 1991), Chan and Tong (1990), Hansen (1996), and Lee, Seo, and Shin (2011). For inference on the regression parameters, relevant articles include Chan (1993), Hansen (2000), and Seo and Linton (2007). Panel data methods have been developed by Hansen (1999) and Ramirez-Rondan

(2013). Instrumental variable methods have been developed by Caner and Hansen (2004). An estimation and inference theory for regression discontinuity with unknown thresholds has been developed by Porter and Yu (2015).

In the statistics literature, related classes of models include two-phase regression, segmented regression, the broken stick model, and the bent cable model. Important contributions to this literature include Hinkley (1969, 1971) and Feder (1975). See also the references in Chiu, Lockhart, and Routledge (2006).

For illustration, we apply the regression kink model to the growth and debt problem made famous by Reinhart and Rogoff (2010). These authors argued that there is a nonlinear effect of aggregate debt on economic growth, specifically that as the ratio of debt to GDP increases above some threshold, aggregate economic growth will tend to slow. This can be formalized as a regression kink model, where GDP growth is the dependent variable and the debt/GDP ratio is the key regressor and threshold variable. Econometric analysis of their proposition using threshold regression tools has been pursued by several authors, including Caner, Grennes, and Koehler-Geib (2010), Cecchetti, Mohanty, and Zampolli (2011), Lin (2014), and Lee et al. (2014). These articles investigate the Reinhart–Rogoff proposition using a variety of cross-section, panel, and time-series regressions, but all have focused on the discontinuous threshold regression model. We add to this literature by a small investigation using long-span U.S. time-series data.

During our investigation we encounter one novel technical issue. While the parameter estimates in the regression kink model are asymptotically normal (as shown by Chan and Tsay 1998), the estimates of the regression function itself are not asymptotically normal, since the regression function is a nondifferentiable function of the parameter estimates. Consequently, conventional inference methods cannot be applied to the regression function. To address this issue we use recently developed inference methods by Fang and Santos (2014) and Hong and Li (2015), and present a new nonnormal distribution theory for the regression

© 2017 American Statistical Association
Journal of Business & Economic Statistics
April 2017, Vol. 35, No. 2

DOI: 10.1080/07350015.2015.1073595

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/jbes.

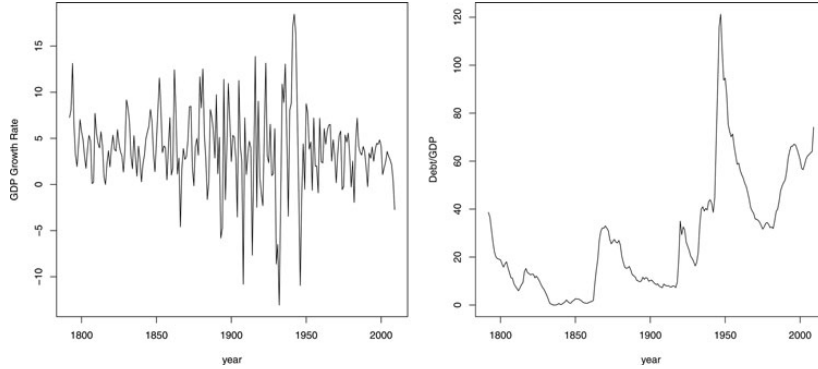


Figure 1. Annual U.S. real GDP growth rate and GDP/debt ratio 1791–2009.

function estimates, and shown how to construct numerical delta method bootstrap confidence intervals for the regression function.

Our organization is as follows. Section 2 introduces the model. Section 3 describes least-square estimation of the model parameters. Section 4 discusses testing for a threshold effect in the context of the model. Section 5 presents an asymptotic distribution theory for the parameter estimates and discusses confidence interval construction. Section 6 discusses inference on the regression function. A formal proof of Theorem 2 is presented in the Appendix.

The data and R code for the empirical and simulation work reported in the article is available on the author's webpage <http://www.ssc.wisc.edu/~bhansen/>.

2. MODEL

Our regression kink model takes the form

$$y_t = \beta_1(x_t - \gamma)_- + \beta_2(x_t - \gamma)_+ + \beta'_3 z_t + e_t, \quad (1)$$

where y_t , x_t , and e_t are scalars, and z_t is an ℓ -vector that includes an intercept. The variables (y_t, x_t, z_t) are observed for $t = 1, \dots, n$. The parameters are β_1 , β_2 , β_3 , and γ . We use $(a)_- = \min[a, 0]$ and $(a)_+ = \max[a, 0]$ to denote the “negative part” and “positive part” of a real number a . In model (1) the slope with respect to the variable x_t equals β_1 for values of x_t less than γ , and equals β_2 for values of x_t greater than γ , yet the regression function is continuous in all variables.

The model (1) is a regression kink model because the regression function is continuous in the variables x and z , but the slope with respect to x is discontinuous (has a kink) at $x = \gamma$. The model (1) specifies the regression segments to be linear, but this could be modified to any parametric form (such as a polynomial). The conventional regression kink design assumes that the threshold point γ is known. (This is suitable in many policy-oriented applications where the threshold is known and determined by policy.) Instead, we treat the parameter γ as an unknown to be estimated. Thus, our methods are appropriate when the threshold is either not determined by the policy, or when the researcher wishes to investigate the robustness of this assumption.

The model (1) has $k = 3 + \ell$ parameters. $\beta = (\beta_1, \beta_2, \beta_3)$ are the regression slopes and are generally unconstrained so that $\beta \in \mathbb{R}^{k-1}$. The parameter γ is called the “threshold,” “knot,” or “kink point.” The model (1) only makes sense if the threshold is in the interior of the support of the threshold variable x_t . We thus assume that $\gamma \in \Gamma$ where Γ is compact and strictly in the interior of the support of x_t .

The regression kink model is nested within the discontinuous threshold model. Therefore, one could imagine testing the assumption of continuity within the threshold model class. This is a difficult problem, one to which we are unaware of a solution, and therefore is not pursued in this article. We simply assume that regression function is continuous and do not explore the issue of testing the assumption of continuity.

As an empirical example, consider the growth/debt regression problem of Reinhart and Rogoff (2010). They argued that economic growth tends to slow when the level of government debt relative to GDP exceeds a threshold. To write this as a regression, we set y_t to be the real GDP growth rate in year t and x_t to be the debt to GDP percentage from the previous year (so that it is plausibly predetermined). We set $z_t = (y_{t-1} \ 1)'$ so that the regression contains a lagged dependent variable to ensure that the error e_t is approximately serially uncorrelated.

We focus on the United States and use the long-span time-series for the years 1792–2009 gathered by Reinhart and Rogoff and posted on their website, so that there are $n = 218$ observations. We display time-series plots of the two series in Figure 1. We follow Lin (2014) and focus on time-series estimates for a single country rather than cross-country or panel estimation, so to not impose parameter homogeneity assumptions.

In Figure 2, we display a scatterplot of (y_t, x_t) along with the fitted regression line and pointwise 90% confidence intervals. (We will discuss estimation in the next section and confidence intervals in Section 6.) We can see that the fitted regression shows a small positive slope for low debt ratios, with a kink (threshold) around 44% (displayed as the red square), switching to a negative slope for debt ratios above that value.

We would like to consider inference in the context of this model, specifically focusing on the following questions: (1) Is the threshold regression statistically different from a linear regression? (2) What is the asymptotic distribution of the parameter estimates, and can we construct confidence intervals for the

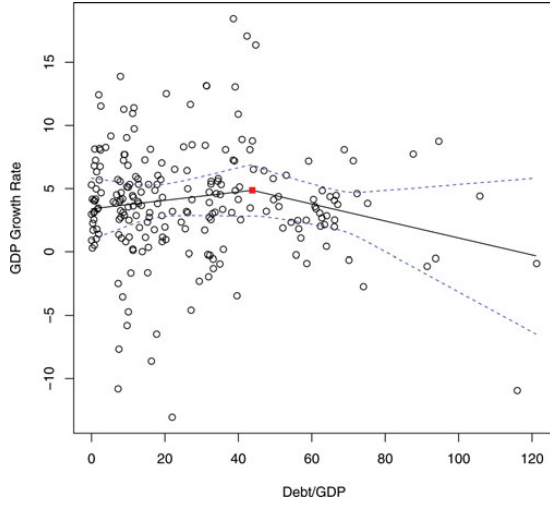


Figure 2. Scatterplot of real GDP growth and debt/GDP, with estimated regression kink model, and 90% confidence intervals.

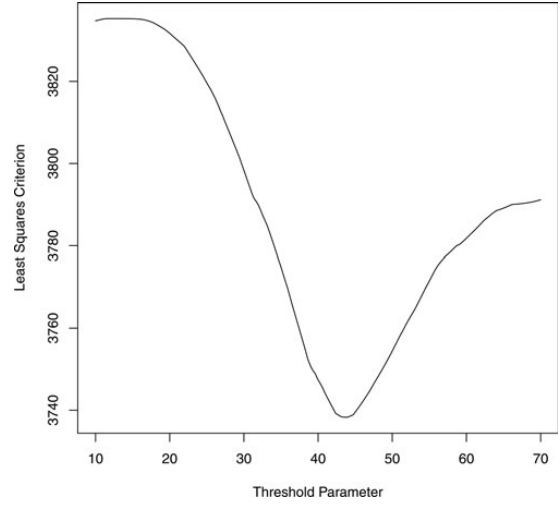


Figure 3. Concentrated least-square criterion for threshold parameter.

parameters? (3) Can we construct confidence intervals for the regression line?

3. ESTIMATION

If the model (1) is interpreted as a conditional mean then the natural estimator of the parameters is least squares. It will be convenient to write

$$x_t(\gamma) = \begin{pmatrix} (x_t - \gamma)_- \\ (x_t - \gamma)_+ \\ z_t \end{pmatrix} \quad (2)$$

so that (1) can be written as $y_t = \beta' x_t(\gamma) + e_t$. The least-square criterion is

$$S_n(\beta, \gamma) = \frac{1}{n} \sum_{t=1}^n (y_t - \beta' x_t(\gamma))^2. \quad (3)$$

The least-square estimator $(\hat{\beta}, \hat{\gamma})$ is the joint minimizer of $S_n(\beta, \gamma)$:

$$(\hat{\beta}, \hat{\gamma}) = \underset{\beta \in \mathbb{R}^{k-1}, \gamma \in \Gamma}{\operatorname{argmin}} S_n(\beta, \gamma) \quad (4)$$

The criterion function $S_n(\beta, \gamma)$ is quadratic in β but non-convex in γ . Thus, it is computationally convenient to use a combination of concentration and grid search, as is typical in the threshold literature. Specifically, notice that by concentration we can write

$$\begin{aligned} \hat{\gamma} &= \underset{\gamma \in \Gamma}{\operatorname{argmin}} \min_{\beta \in \mathbb{R}^{k-1}} S_n(\beta, \gamma) \\ &= \underset{\gamma \in \Gamma}{\operatorname{argmin}} S_n^*(\gamma), \end{aligned} \quad (5)$$

where $\hat{\beta}(\gamma)$ are the least-square coefficients from a regression of y_t on the variables $x_t(\gamma)$ for fixed γ , and

$$S_n^*(\gamma) = S_n(\hat{\beta}(\gamma), \gamma) = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{\beta}(\gamma)' x_t(\gamma))^2$$

is the concentrated sum-of-squared errors function. The solution to (5) can be found numerically by a grid search over γ . After $\hat{\gamma}$ is found the coefficient estimates $\hat{\beta}$ are obtained by standard least squares of y_t on $x_t(\hat{\gamma})$. We write the fitted regression function as

$$y_t = \hat{\beta}' x_t(\hat{\gamma}) + \hat{e}_t. \quad (6)$$

In (6), \hat{e}_t are the (nonlinear) least-square residuals. An estimate of error variance $\sigma^2 = E e_t^2$ is

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n \hat{e}_t^2 = S_n^*(\hat{\gamma}).$$

To illustrate, consider the U.S. growth regression for 1792–2009. First, we set the parameter space Γ for the threshold parameter as $\Gamma = [10, 70]$, so that at least 5% of the sample and 10% of the support of the threshold variable are below the lower bound and above the upper bound. We then approximate the minimization (5) by computing $S_n^*(\gamma)$ on a discrete grid with increments 0.1 (which has 601 gridpoints). At each gridpoint for γ , we estimated the least-square coefficients and computed the least-square criterion $S_n^*(\gamma)$. This criterion is plotted as a function of γ in Figure 3. We observe that the function is reasonably smooth and has a well-defined global minimum, but the criterion is not well described as quadratic. The relative smoothness of the plot suggests that our choice for the grid evaluation is sufficiently fine to obtain the global minimum. The criterion is minimized at $\hat{\gamma} = 43.8$. Interestingly, this threshold estimate is very close to that found by Lin (2014) for the United States using a different data window and quite different

empirical framework, and also very close to the estimate found by Lee et al. (2014) using a median cross-country regression.

The parameter estimates from this fitted regression kink model are as follows.

$$\begin{aligned} y_t = & \begin{pmatrix} 0.033 \\ (0.026) \end{pmatrix} \begin{pmatrix} x_{t-1} - 43.8 \\ (12.1) \end{pmatrix}_- \\ & - 0.067 \begin{pmatrix} x_{t-1} - 43.8 \\ (0.048) \end{pmatrix}_+ + \begin{pmatrix} 0.28 \\ (0.09) \end{pmatrix} y_{t-1} \\ & + \begin{pmatrix} 3.78 \\ (0.69) \end{pmatrix} + \hat{e}_t \\ \hat{\sigma}^2 = & 17.2. \end{aligned}$$

The standard errors will be discussed in Section 5. The estimates show a positive first-order autocorrelation coefficient of 0.28, which is standard for time-series estimates for growth rates. The estimates suggest that high debt ratios (those above 44) induce a moderate slowdown in expected GDP growth rates, consistent with the Reinhart–Rogoff hypothesis.

4. TESTING FOR A THRESHOLD EFFECT

A reasonable question is whether or not the threshold model (1) is “significant” relative to the linear model

$$y_t = \beta_1 x_t + \beta_3' z_t + e_t. \quad (7)$$

The linear model (7) is nested in the threshold model (1), as (7) holds under the restriction $\beta_1 = \beta_2$ (with a redefinition of the intercept). Under this hypothesis the threshold γ disappears and we see the familiar problem of an unidentified parameter under the null hypothesis. This requires nonstandard testing methods; fortunately such issues are well understood and easily applied.

An appropriate estimator of the linear model is least squares. We can write these estimates as

$$y_t = \tilde{\beta}_1 x_t + \tilde{\beta}_3' z_t + \tilde{e}_t. \quad (8)$$

The error variance estimate is

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n \tilde{e}_t^2.$$

For example, the estimates of the linear model for our empirical example are as follows:

$$\begin{aligned} y_t = & \begin{pmatrix} -0.008 \\ (0.013) \end{pmatrix} x_{t-1} + \begin{pmatrix} 0.30 \\ (0.09) \end{pmatrix} y_{t-1} + \begin{pmatrix} 2.90 \\ (0.62) \end{pmatrix} + \tilde{e}_t \\ \tilde{\sigma}^2 = & 17.6. \end{aligned} \quad (9)$$

The estimates show a positive first-order autocorrelation coefficient, and a near zero coefficient on the debt ratio.

A standard test for the null hypothesis of the linear model (7) against the threshold model (1) is to reject $H_0 : \beta_1 = \beta_2$ for large values of the F -type statistic

$$T_n = \frac{n(\tilde{\sigma}^2 - \hat{\sigma}^2)}{\hat{\sigma}^2}.$$

For example, for our empirical example we can compute that $T_n = 5.66$.

To describe the distribution theory for the test statistic, we need to be precise about the stochastic assumptions on the model. We require the following regularity conditions.

Assumption 1. For some $r > 1$,

1. (y_t, z_t, x_t) is strictly stationary, ergodic, and absolutely regular with mixing coefficients $\eta(m) = O(m^{-A})$ for some $A > r/(r-1)$
2. $E|y_t|^{4r} < \infty$, $E|x_t|^{4r} < \infty$, and $E\|z_t\|^{4r} < \infty$
3. $\inf_{\gamma \in \Gamma} \det Q(\gamma) > 0$, where $Q(\gamma) = E(x_t(\gamma)x_t(\gamma)')$
4. x_t has a density function $f(x)$ satisfying $f(x) \leq \bar{f} < \infty$
5. $\gamma \in \Gamma$ where Γ is compact

Assumptions 1.1 and 1.2 are standard weak dependence conditions that are sufficient for a central limit theorem. The choice of r involves a trade-off between the allowable degree of serial dependence and the number of finite moments. For independent observations, we can set r arbitrarily close to one. Assumption 1.3 is an identification condition, requiring that the projection coefficients be well defined for all values of γ in the parameter space Γ . Assumption 1.4 requires that the threshold variable has a bounded density function.

In particular, Assumption 1 allows the regressors x_t and z_t to include lagged dependent variables.

The following is implied by Theorem 3 of Hansen (1996). The only difference is that the latter theorem focused on a heteroscedasticity-robust Wald statistic, while for simplicity we consider the simpler homoscedastic form of the test statistic.

Theorem 1. Suppose that Assumption 1 holds and e_t is a martingale difference sequence. Under $H_0 : \beta_1 = \beta_2$,

$$T_n \rightarrow_d \sup_{\gamma \in \Gamma} G(\gamma)' Q(\gamma)^{-1} G(\gamma) / \sigma^2, \quad (10)$$

where $G(\gamma)$ is a zero-mean Gaussian process with covariance kernel

$$E(G(\gamma_1)G(\gamma_2)) = E(x_t(\gamma_1)x_t(\gamma_2)'e_t^2). \quad (11)$$

Theorem 1 shows that the asymptotic null distribution of the threshold F statistic can be written as the supremum of a stochastic process. In addition to Assumption 1, Theorem 1 adds the regularity condition that the error e_t is a martingale difference sequence. This is a sufficient condition, and convenient, but is not essential. What is important is that the best-fitting approximation in the regression kink model is the linear model so that the central limit theorem applies for all γ , and that the regression scores are uncorrelated so that the covariance kernel takes the simple form (11) rather than an Heteroskedasticity and Autocorrelation Consistent (HAC) form.

The limiting distribution (10) is nonstandard and cannot be tabulated. However, as shown by Hansen (1996), it is simple to simulate approximations to (10) using a multiplier bootstrap, and thus asymptotically valid p -values can be calculated. The following is his recommended algorithm. (Theorem 3 of Hansen (1996) shows that the algorithm produces asymptotically first-order correct p -values under the conditions of Theorem 1.)

Algorithm 1: Testing for a Regression Kink With an Unknown Threshold.

1. Generate n iid draws u_t from the $N(0, 1)$ distribution.
2. Set $y_t^* = \tilde{e}_t u_t$ where \tilde{e}_t are the ordinary least-square (OLS) residuals from the fitted linear regression (8).
3. Using the observations (y_t^*, z_t, x_t) , estimate the linear model (8) and the regression kink model (6) and compute the error variance estimates $\tilde{\sigma}^{*2}$ and $\hat{\sigma}^{*2}$ and the F -statistic

$$T_n^* = \frac{n(\tilde{\sigma}^{*2} - \hat{\sigma}^{*2})}{\hat{\sigma}^{*2}}.$$

4. Repeat this B times, so as to obtain a sample $T_n^*(1), \dots, T_n^*(B)$ of simulated F statistics.
5. Compute the p -value as the percentage of simulated F statistics, which exceed the actual value:

$$p_n = \frac{1}{B} \sum_{b=1}^B 1(T_n^*(b) \geq T_n).$$

6. If desired, compute the level α critical value c_α as the empirical $1 - \alpha$ quantile of the simulated F statistics $T_n^*(1), \dots, T_n^*(B)$.
7. Reject H_0 in favor of H_1 at significance level α if $p_n < \alpha$, or equivalently if $T_n > c_\alpha$.

The number of bootstrap replications B should be set fairly large to ensure accuracy of the p -value p_n . For example, if $B = 10,000$ then p_n is approximately within 0.006 of its limiting (large B) value. Fortunately, the computational cost is minimal. For example, my office computer computed all the empirical calculations reported in this article in just 1 min, including two separate bootstrap simulations using 10,000 replications each. (The calculations were performed in R. The code implements the multiplier bootstrap efficiently by executing all 10,000 regressions simultaneously, exploiting the fact that the regressors are common across the bootstrap replications.)

Returning to the U.S. GDP example, as we said earlier the empirical value of the F statistic is $T_n = 5.66$. The multiplier bootstrap p -value is 0.15, and the bootstrap estimate of the 10% critical value is 7.1. Thus, the test does not reject the null hypothesis of linearity in favor of the regression kink model at the 10% level.

The multiplier bootstrap method does not account for the time-series nature of the observations and thus can be expected to exhibit some finite sample distortions. In particular, Assumption 1 does not allow for nonstationary or near-nonstationary regressors. As argued by Stambaugh (1999) and others, highly persistent regressors will distort normal distribution theory. This is particularly relevant for our example as can be seen in Figure 1, the debt/GDP level displays considerable serial dependence.

To investigate this possibility, we present a simple simulation experiment calibrated on the empirical example. Our data-generating process is

$$y_t = \beta_1(x_t - \gamma)_- + \beta_2(x_t - \gamma)_+ + \beta_3 y_{t-1} + \beta_4 + e_t$$

$$e_t \sim N(0, \sigma^2).$$

To evaluate size, we set $\beta_1 = \beta_2 = 0$, $\beta_3 = 0.3$, $\beta_4 = 3$, and $\sigma^2 = 16$ to match the estimates (9). We fix x_t to equal the empirical values of debt/GDP for the United States, and set $n =$

Table 1. Power of threshold F test with multiplier bootstrap, nominal size 10%

	β_2							
	-0.02	-0.04	-0.06	-0.08	-0.10	-0.12	-0.14	-0.16
Power	0.14	0.24	0.38	0.54	0.72	0.85	0.93	0.98

218 as in the empirical example. Setting x_t to equal the sample values is done precisely to force the simulation to evaluate the performance in a setting with the serial correlation properties of the observed debt/GDP series. We generated 10,000 samples from this process. To speed computation, for all simulations in the article we evaluate $\Gamma = [10, 70]$ using a grid with increments of 1, which reduces the number of gridpoints to 61. The number of bootstrap replications for each simulated sample was set at $B = 1000$.

We first evaluated the size of the threshold test. At the nominal 10% significance level, we found the simulation size to be 10.6%. Thus, the test exhibits no meaningful size distortion.

Second, we evaluated the power of the test. We used the same parameterization as above, but now set $\beta_1 = 0$ and $\gamma = 40$, and vary β_2 from 0 to -0.16 in steps of 0.02. The power is presented in Table 1. What we can see is that the test has increasing power in β_2 , and had reasonable ability to detect changes as small as 0.08. Examining the point estimates of the threshold model, we see that the estimate of the difference in regression slopes is $\hat{\beta}_1 - \hat{\beta}_2 = 0.10$, where the simulation suggests the power should be about 70%, which is reasonable but modest. It follows that our empirical p -value of 15% could be due to the modest power of the test. We conclude that the threshold test is inconclusive regarding the question of whether or not there is a regression kink effect in GDP growth due to high debt.

5. INFERENCE ON THE REGRESSION COEFFICIENTS

In this section, we consider the distribution theory for the least-square estimates of the regression kink model with unknown threshold under the assumption that the threshold is identified.

In an ideal context we might consider (1) as the true regression function, so that the error has conditional mean zero. To allow extra generality, we will not impose this condition. Instead, we will view (1) as the best approximation in the sense that it minimizes squared error loss. We define the best approximation as (pseudo)-true values (β_0, γ_0) , which minimize the squared loss

$$L(\beta, \gamma) = E(y_t - \beta'x_t(\gamma))^2. \quad (12)$$

As in our analysis of estimation, we can define the minimizers by concentration. Let $\beta(\gamma)$ be the minimizer of $L(\beta, \gamma)$ over β for fixed γ , that is, $\beta(\gamma) = E(x_t(\gamma)x_t(\gamma)')^{-1}E(x_t(\gamma)y_t)$. Under Assumption 1.3, $\beta(\gamma)$ is uniquely defined for all $\gamma \in \Gamma$. The concentrated squared loss is then

$$L^*(\gamma) = L(\beta(\gamma), \gamma) = E(y_t - \beta(\gamma)'x_t(\gamma))^2.$$

By concentration, γ_0 is the minimizer of $L^*(\gamma)$ and $\beta_0 = \beta(\gamma_0)$.

We will require that the minimizer γ_0 is unique. This excludes the case of a best-fitting linear model (in which case $L^*(\gamma)$ is a constant function) and the case of multiple best-fitting threshold parameters γ .

To simplify our proof of consistency, we also impose that the parameter space for β is compact, though this assumption could be relaxed by a more detailed argument.

Assumption 2.

1. $\gamma_0 = \operatorname{argmin}_{\gamma \in \Gamma} L^*(\gamma)$ is unique.
2. $\beta \in B \subset \mathbb{R}^{k-1}$ where B is compact.

Chan and Tsay (1998) showed that the least-square estimates in the continuous threshold autoregressive model, including both the slope and threshold coefficients, are jointly asymptotically normal. We extend their distribution theory to the regression kink model with unknown threshold. Set $\theta = (\beta, \gamma)$, $\theta_0 = (\beta_0, \gamma_0)$,

$$\begin{aligned} H_i(\theta) &= -\frac{\partial}{\partial \theta} (y_i - \beta' x_i(\gamma)) \\ &= \begin{pmatrix} x_i(\gamma) \\ -\beta_1 1(x_i < \gamma) - \beta_2 1(x_i > \gamma) \end{pmatrix} \end{aligned}$$

and $H_i = H_i(\theta_0)$.

Theorem 2. Under Assumption 1

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d N(0, V)$$

where $V = Q^{-1} S Q^{-1}$, $S = \sum_{j=-\infty}^{\infty} E(H_i H'_{i+j} e_i e_{i+j})$ and

$$\begin{aligned} Q &= E(H_i H'_i) \\ &+ E \begin{pmatrix} 0 & 0 & 0 & e_i 1(x_i < \gamma_0) \\ 0 & 0 & 0 & e_i 1(x_i > \gamma_0) \\ 0 & 0 & 0 & 0 \\ e_i 1(x_i < \gamma_0) & e_i 1(x_i > \gamma_0) & 0 & 0 \end{pmatrix}. \end{aligned}$$

A formal proof of Theorem 2 is presented in the Appendix.

Notice that the slope and threshold estimates are jointly asymptotically normal with \sqrt{n} convergence rate, and the slope and threshold estimates have a nonzero asymptotic covariance. In contrast, in the conventional noncontinuous threshold model the threshold estimate $\hat{\gamma}$ is rate n consistent with a nonstandard asymptotic distribution, and the slope coefficient estimates are asymptotically independent of the threshold estimate. The difference in the regression kink model is because the regression function is continuous. Consequently, the least-square criterion is continuous in the parameters (though not differentiable) and asymptotically locally quadratic.

Since the threshold estimate $\hat{\gamma}$ has only a \sqrt{n} convergence rate, we should expect its precision to be less accurate than threshold estimates in the noncontinuous case. Thus, it is important to take its sampling distribution into account when constructing confidence intervals.

The asymptotic distribution does not require the model to be correctly specified, so the error e_i need not be a martingale difference sequence. Thus (in general), the covariance matrix S takes an HAC form. When the regression is dynamically well specified (by appropriate inclusion of lagged variables) then the matrix will simplify to $S = E(H_i H'_i e_i^2)$. As our application

includes a lagged dependent variable, we use this simplification in practice.

The second term in the definition of Q is zero when the threshold model is correctly specified so that $E(e_i | x_i) = 0$. However, in the case of model misspecification (so that the regression is a best approximation) then the second term can be nonzero.

We suggest the following estimate of the asymptotic covariance matrix (assuming uncorrelatedness). Set

$$\begin{aligned} \hat{H}_i &= \begin{pmatrix} x_i(\hat{\gamma}) \\ -\hat{\beta}_1 1(x_i < \hat{\gamma}) - \hat{\beta}_2 1(x_i > \hat{\gamma}) \end{pmatrix}, \\ \hat{V} &= \hat{Q}^{-1} \hat{S} \hat{Q}^{-1}, \hat{S} = \frac{1}{n-k} \sum_{i=1}^n \hat{H}_i \hat{H}'_i \hat{e}_i^2, \text{ and} \\ \hat{Q} &= \frac{1}{n} \sum_{i=1}^n (\hat{H}_i \hat{H}'_i \\ &+ \begin{pmatrix} 0 & 0 & 0 & \hat{e}_i 1(x_i < \hat{\gamma}) \\ 0 & 0 & 0 & \hat{e}_i 1(x_i > \hat{\gamma}) \\ 0 & 0 & 0 & 0 \\ \hat{e}_i 1(x_i < \hat{\gamma}) & \hat{e}_i 1(x_i > \hat{\gamma}) & 0 & 0 \end{pmatrix}). \end{aligned}$$

We divide by $n - k$ rather than n for the definition of \hat{S} as an ad hoc degree-of-freedom adjustment. Given \hat{V} , standard errors for the coefficient estimates are found by taking the square roots of the diagonal elements of $n^{-1} \hat{V}$. If the regression is dynamically misspecified (e.g., if no lagged dependent variable is included) then \hat{S} could be formed using a standard HAC estimator.

Asymptotic confidence intervals for the coefficients could then be formed using the conventional rule, for example, for a 95% interval for β_1 , $\hat{\beta}_1 \pm 1.96s(\hat{\beta}_1)$. Theorem 2 shows that such confidence intervals have asymptotic correct coverage.

In small samples, however, the asymptotic confidence intervals may have poor coverage. As shown in Figure 2, the least-square criterion is nonquadratic with respect to the threshold parameter γ , meaning that quadratic (e.g., normal) approximations may be poor unless sample sizes are quite large. In this context better coverage can be obtained by test-inversion confidence sets. This is particularly convenient for the threshold parameter γ , as a test inversion confidence region is a natural by-product of the computation of the least-square minimization. Specifically, to test the hypothesis $H_0 : \gamma = \gamma_0$ against $H_1 : \gamma \neq \gamma_0$, the criterion-based test is to reject for large values of the F -type statistic $F_n(\gamma_0)$, where

$$F_n(\gamma) = \frac{n(\hat{\sigma}^2(\gamma) - \hat{\sigma}^2)}{\hat{\sigma}^2} \quad (13)$$

and $\hat{\sigma}^2(\gamma) = S_n^*(\gamma)$ is the estimator of the error variance when γ is fixed. Given the asymptotic normality of Theorem 2, this test has an asymptotic χ_1^2 distribution under H_0 . Thus, for a nominal level α test we can take the critical value $c_{1-\alpha}$ from the χ_1^2 distribution. A nominal $1 - \alpha$ asymptotic confidence interval for γ can then be formed by test inversion: the set of γ for which $F_n(\gamma)$ is smaller than the χ_1^2 critical value:

$$C_\gamma = \{\gamma : F_n(\gamma) \leq c_{1-\alpha}\}.$$

Given that $\hat{\sigma}^2(\gamma)$ has already been calculated on a grid (for estimation), we have $F_n(\gamma)$ on the same grid as a by-product. The interval C_γ is then obtained as the set of γ gridpoints for which $F_n(\gamma)$ is smaller than $c_{1-\alpha}$.

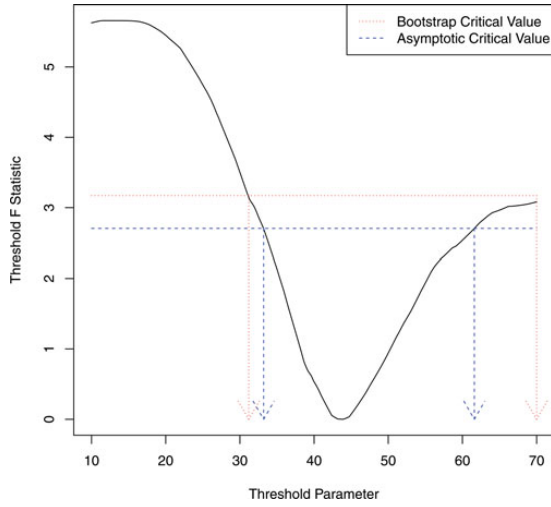


Figure 4. Confidence interval construction for threshold.

To illustrate, examine Figure 4. Here we have plotted the statistic $F_n(\gamma)$ from (13) as a function of γ . By construction, the statistic is nonnegative and hits 0 at $\gamma = \hat{\gamma}$. We have drawn in the asymptotic 90% critical value $c_{.9} = 2.7$ using the blue dashed line. The points of intersection indicate the ranges of the asymptotic confidence interval, and shown on the graph by the blue dashed arrows to the horizontal axes.

Further improvements in coverage accuracy can be obtained via a bootstrap. Since the regressor x_t displays considerable serial dependence, we recommend a wild bootstrap that conditions on the values of x_t . This will preserve the time-series properties better than alternative bootstrap methods such as a block bootstrap that is more suited for cases of mild serial dependence. A model-based bootstrap could also be used, but then an explicit model would be required for the debt/GDP series. To avoid these challenges we adopt a wild bootstrap. Here are the steps.

Algorithm 2: Wild Bootstrap Confidence Intervals for Parameters.

1. Generate n iid draws u_t from the $N(0, 1)$ distribution.
2. Set $e_t^* = \hat{e}_t u_t$ where \hat{e}_t are the LS residuals from the fitted regression kink model (6).
3. Set $y_t^* = \hat{\beta}' x_t(\hat{\gamma}) + e_t^*$, where (z_t, x_t) are the sample observations, and $(\hat{\beta}, \hat{\gamma})$ are the least-square estimates.
4. Using the observations (y_t^*, z_t, x_t) , estimate the regression kink model (6), parameter estimates $(\hat{\beta}^*, \hat{\gamma}^*)$, and $\hat{\sigma}^{*2} = n^{-1} \sum_{t=1}^n \hat{e}_t^{*2}$, where $\hat{e}_t^* = y_t^* - \hat{\beta}^{*'} x_t(\hat{\gamma}^*)$.
5. Calculate the F -statistic for γ

$$F_n^* = \frac{n(\hat{\sigma}^{*2}(\hat{\gamma}) - \hat{\sigma}^{*2})}{\hat{\sigma}^{*2}},$$

where $\hat{\sigma}^{*2}(\hat{\gamma}) = n^{-1} \sum_{t=1}^n \hat{e}_t^{*2}(\hat{\gamma})$ and $\hat{e}_t^*(\hat{\gamma}) = y_t^* - \hat{\beta}^{*'} x_t(\hat{\gamma})$.

6. Repeat this B times, so as to obtain a sample of simulated coefficient estimates $(\hat{\beta}^*, \hat{\gamma}^*)$ and F statistics F_n^* .
7. Create $1 - \alpha$ bootstrap confidence intervals for the coefficients β_1 , β_2 , and β_3 by the symmetric percentile method: the coefficient estimates plus and minus the $(1 - \alpha)$ quantile

of the absolute centered bootstrap estimates. For example, for β_1 the interval is $\hat{\beta}_1 \pm q_1^*$, where q_1^* is the $(1 - \alpha)$ quantile of $|\hat{\beta}_1^* - \hat{\beta}_1|$.

8. Calculate the $1 - \alpha$ quantile $c_{1-\alpha}^*$ of the simulated F statistics F_n^* .
9. Create a $1 - \alpha$ bootstrap confidence interval for γ as the set of γ for which the empirical F statistics $F_n(\gamma)$ (13) are smaller than the bootstrap critical value $c_{1-\alpha}^*$.

$$C_\gamma^* = \{\gamma : F_n(\gamma) \leq c_{1-\alpha}^*\}.$$

This wild bootstrap algorithm can be computed concurrently with the multiplier bootstrap used for the threshold test, resulting in efficient computation. We do not have a formal distribution theory that justifies the use of this (or any other) bootstrap method for confidence interval construction. However, we see no reason for the bootstrap to fail given the asymptotic normality of Theorem 2.

Again for illustration the confidence interval construction can be seen via Figure 4, where the statistic $F_n(\gamma)$ is plotted against γ . The bootstrap 90% critical value 3.3 (calculated with $B=10,000$ bootstrap replications) is plotted as the red dotted line. The points of intersection indicate the ranges of the confidence interval, and are indicated on the figure by the red dotted arrows to the horizontal axis. Since the bootstrap critical value is larger than the asymptotic critical value, the asymptotic interval is a subset of the bootstrap confidence interval.

Before presenting the empirical results we report results from our simulation experiment to assess the performance of the methods. The data-generating process is identical to that used in the previous section. As before, β_2 is the key free parameter, controlling the strength of the threshold effect, and the remaining parameters and variables are set to match the empirical data. In particular, we fixed the regressor x_t to equal the empirical values of debt/GDP for the United States, so to precisely preserve its strong serial dependence properties. As before, we generated 10,000 simulated samples and used $B = 1000$ bootstrap replications for each sample. In Tables 2 and 3, we report the coverage rates of nominal 90% intervals for the parameters β_2 and γ .

The confidence intervals for β_2 are reported in Table 2. The first row reports the coverage rates for the asymptotic “plus or minus” standard error confidence intervals. We see that the confidence intervals undercover, with rates approximately 81%. The second row reports the coverage rates of the percentile confidence interval. These rates are better, especially for moderate values of β_2 . In the third row, we report the coverage rates for the “inverse percentile bootstrap, whose endpoints are twice the coefficient estimates, less the $1 - \alpha/2$ and $\alpha/2$ quantiles of the simulated coefficient estimates $\hat{\beta}_2^*$. Its performance is not superior to the percentile interval. In the fourth row, we report the coverage rates for the “symmetric percentile bootstrap” whose endpoints equal the coefficient estimates, plus and minus the $1 - \alpha$ quantile of the simulated statistics $|\hat{\beta}_2^* - \hat{\beta}_2|$. These intervals only slightly uncover, with coverage rates about 86%–88%. These are our recommended confidence intervals for the slope coefficients.

In Table 3, we report coverage rates of nominal 90% intervals for the parameter γ . The first row reports the coverage rates for the asymptotic “plus or minus” standard error confidence

Table 2. Coverage rates of nominal 90% confidence intervals for β_2

	β_2							
	−0.02	−0.04	−0.06	−0.08	−0.10	−0.12	−0.14	−0.16
$\hat{\beta}_2 \pm 1.645s(\hat{\beta}_2)$	0.81	0.80	0.81	0.81	0.81	0.80	0.81	0.81
Percentile bootstrap	0.81	0.83	0.85	0.87	0.86	0.86	0.86	0.86
Inverse percentile bootstrap	0.84	0.84	0.85	0.83	0.82	0.82	0.82	0.83
Symmetric percentile bootstrap	0.86	0.86	0.87	0.88	0.87	0.86	0.86	0.86

intervals. These severely undercover, with coverage rates as low as 67%. The second and third rows report coverage rates of the conventional percentile interval and the inverse percentile interval. Neither performs well, with the percentile interval substantially overcovering and the inverse percentile interval severely undercovering for small β_2 . The fourth row reports the coverage rates of the symmetric percentile interval. Its performance is better than the other percentile intervals, but coverage rates are sensitive to the value of β_2 . The fifth and sixth rows report the coverage rates for the test-inversion confidence intervals C_γ and C_γ^* . These two have quite good coverage, especially C_γ^* that uses the bootstrap critical values. The interval only mildly undercovers, and has roughly uniform coverage across β_2 . The intervals C_γ^* are our recommended confidence intervals for γ .

Our simulation results show that the recommended wild bootstrap confidence intervals work reasonably well, but with some moderate undercoverage. This is likely because our simulated data are a time-series autoregression (and thus have time-series dependence) while the wild bootstrap does not account for the time-series dependence. The reason we recommend a wild bootstrap (rather than a model-based bootstrap) is that the wild bootstrap is able to condition on the observed regressor processes and (most importantly) nonparametrically handle conditional heteroscedasticity. These are important advantages and conveniences.

We now present in Table 4 the confidence intervals for the coefficient estimates from the empirical regression, computed with our recommended bootstrap methods with $B=10,000$ bootstrap replications. In our assessment, the two most important coefficients are β_2 (the slope effect of debt on growth for high debt ratios) and γ (the threshold level). The confidence interval for β_2 is $[-0.18, -0.01]$. This range is sufficiently wide that we cannot infer with precision the magnitude of the impact of debt on expected growth. The 90% confidence interval for the threshold is 31% to 70%, also indicating substantial uncertainty.

The wide confidence interval for the threshold could be a feature of the \sqrt{n} convergence rate for the threshold estimate, or it could be due to the small sample size.

6. INFERENCE ON THE REGRESSION KINK FUNCTION

In this section, we consider inference on the regression kink function

$$g(\theta) = \beta'x(\gamma), \quad (14)$$

where

$$x(\gamma) = \begin{pmatrix} (x - \gamma)_- \\ (x - \gamma)_+ \\ z \end{pmatrix}.$$

Our theory will treat (x, z) as fixed, even though we will display the estimates as a function of x . (Thus, we focus on pointwise confidence intervals for the regression function.)

The plug-in estimate of $g(\theta)$ is $g(\hat{\theta}) = \hat{\beta}'x(\hat{\gamma})$. In our empirical example, we display in Figure 2 $g(\hat{\theta})$ as a function of x , with z fixed at the sample mean of z_t .

Since Theorem 2 shows that $\hat{\theta}$ is asymptotically normal, it might be conjectured that $g(\hat{\theta})$ will be as well. This conjecture turns out to be incorrect. While $g(\theta)$ is a continuous function of θ , it is not differentiable at $\gamma = x$. As discussed in a recent series of articles (Hirano and Porter 2012; Woutersen and Ham 2013; Fang and Santos 2014; Fang 2014; Hong and Li 2015), the nondifferentiability means that $g(\hat{\theta})$ will not be asymptotically normal at $\gamma_0 = x$, and asymptotic normality is likely to be a poor approximation for γ_0 close to x .

While the regression kink function $g(\theta)$ is not differentiable at $\gamma = x$, it is directionally differentiable at all points, meaning that both left and right derivatives are defined. The directional derivative of a function $\phi(\theta) : \mathbb{R}^k \rightarrow \mathbb{R}$ in the direction $h \in \mathbb{R}^k$ is

$$\phi_\theta(h) = \lim_{\varepsilon \downarrow 0} \frac{\phi(\theta + h\varepsilon) - \phi(\theta)}{\varepsilon} \quad (15)$$

(see Shapiro 1990). The primary difference with the conventional notion of a derivative is that $\phi_\theta(h)$ is allowed to depend on the direction h in which the derivative is taken. For continuously differentiable functions the directional derivative is linear in h .

Table 3. Coverage rates of nominal 90% confidence intervals for γ

	β_2							
	−0.02	−0.04	−0.06	−0.08	−0.10	−0.12	−0.14	−0.16
$\hat{\gamma} \pm 1.645s(\hat{\gamma})$	0.67	0.68	0.69	0.70	0.71	0.73	0.74	0.75
Percentile bootstrap	0.99	0.99	0.97	0.96	0.94	0.93	0.92	0.91
Inverse percentile bootstrap	0.33	0.40	0.50	0.58	0.67	0.74	0.77	0.81
Symmetric percentile bootstrap	0.96	0.94	0.91	0.89	0.87	0.86	0.85	0.86
C_γ	0.90	0.88	0.87	0.86	0.84	0.85	0.84	0.85
C_γ^*	0.91	0.89	0.88	0.88	0.87	0.88	0.87	0.88

Table 4. Coefficient estimates and bootstrap 90% confidence intervals

	Estimate	s.e.	Interval
β_1	0.033	0.035	[−0.002, 0.136]
β_2	−0.067	0.037	[−0.178, −0.006]
y_{i-1}	0.28	0.088	[0.14, 0.43]
Intercept	3.78	0.79	[2.58, 4.94]
γ	43.8	12.3	[30.8, 70.0]

For example, if $\phi(\theta) = x'\theta$ is linear, then $\phi_\theta(h) = x'h$. If $\phi(\theta)$ is continuously differentiable in θ then $\phi_\theta(h) = (\frac{\partial}{\partial \theta} \phi(\theta))' h$. However, if $\phi(\theta)$ is continuous but not continuously differentiable the directional derivative will be nonlinear in h .

In the case of (14), we can calculate that the directional derivative of $g(\theta)$ in the direction $h = (h_\beta, h_\gamma)$ is

$$g_\theta(h) = x(\gamma)'h_\beta + g_\gamma(h_\gamma), \quad (16)$$

where

$$g_\gamma(h_\gamma) = \begin{cases} -\beta_1 h_\gamma, & \text{if } x < \gamma \\ -\beta_1 h_\gamma 1(h_\gamma < 0) - \beta_2 h_\gamma 1(h_\gamma \geq 0), & \text{if } x = \gamma \\ -\beta_2 h_\gamma, & \text{if } x > \gamma. \end{cases} \quad (17)$$

The directional derivative $g_\gamma(h_\gamma)$ is linear for $x \neq \gamma$ but nonlinear for $x = \gamma$, with a slope to the left of $-\beta_1$ and a slope to the right of $-\beta_2$.

Conventional asymptotic theory for functions of the form $\phi(\theta)$ requires that the function $\phi(\theta)$ be continuously differentiable. Shapiro (1991, Theorem 2.1) and Fang and Santos (2014, Corollary 2.1) had generalized this to the case of directional differentiability. They have established that if $\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d Z \sim N(0, V)$ and $\phi_\theta(h)$ is continuous in h then $\sqrt{n}(\phi(\hat{\theta}) - \phi(\theta_0)) \rightarrow_d \phi_\theta(Z)$. In the classic case where $\phi(\theta)$ is continuously differentiable, the limiting distribution specializes to $\phi_\theta(Z) = (\frac{\partial}{\partial \theta} \phi(\theta))' Z \sim N(0, (\frac{\partial}{\partial \theta} \phi(\theta))' V (\frac{\partial}{\partial \theta} \phi(\theta)))$, so their result is a strict generalization of the Delta Method.

Since (16)–(17) is continuous in h , we can immediately deduce the asymptotic distribution of the regression estimate.

Theorem 3. $\sqrt{n}(g(\hat{\theta}) - g(\theta_0)) \rightarrow_d g_\theta(Z)$ where $Z \sim N(0, V)$ and $g_\theta(h)$ is defined in (16)–(17).

For $x \neq \gamma_0$ the asymptotic distribution in Theorem 3 is normal, but for $x = \gamma_0$ it is a nonlinear transformation of a normal random vector. At $x = \gamma_0$ the asymptotic distribution will be biased, with the direction of bias depending on the relative magnitudes of β_1 and β_2 . For example, if $\beta_1 = 1$ and $\beta_2 = -1$ then $g_\gamma(h_\gamma) = |h_\gamma|$ so the asymptotic distribution in Theorem 3 will have a positive mean while if the signs of the coefficients are reversed then $g_\gamma(h_\gamma) = -|h_\gamma|$ and the asymptotic distribution in Theorem 3 will have a negative mean.

This nonnormality implies that classical confidence intervals will have incorrect asymptotic coverage. We illustrate this in Figure 5 by a continuation of our simulation experiment. Using the same simulation design as in the previous section, and varying β_2 from -0.04 to -0.16 in steps of 0.04 , we plot the coverage probability of classical (naive) nominal 90% pointwise confidence intervals, plotting the coverage as a function of

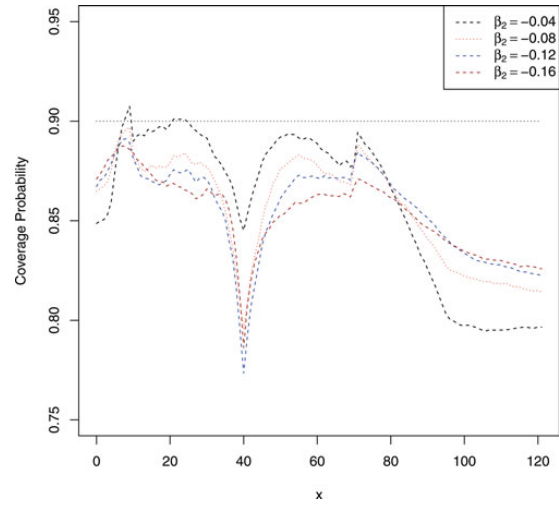


Figure 5. Coverage probabilities of nominal 90% naive asymptotic confidence intervals for regression kink function.

x . The asymptotic theory predicts that the confidence intervals will have asymptotically correct 90% coverage for $x \neq \gamma_0$ but not for $x = \gamma_0$, so we should expect coverage to be better for x distinct from $\gamma = 40$, but deteriorating for x near $\gamma = 40$. Indeed, we see that the coverage rates vary from near 90% for small x to approximately 77% at $x = 40$. The distortion from the nominal coverage is sensitive to β_2 , with the distortion steepening as β_2 increases. We can also see that the coverage rates are less than the nominal 90% for large values of x , which is not predicted from the asymptotic theory, and these distortions are more severe for small values of β_2 . This appears to be a small-sample issue (a finite sample bias in the regression estimate; there are only seven observations where $x_i \geq 80$) and thus unrelated to the asymptotic nonnormality of Theorem 3.

We might hope that bootstrap methods would improve the coverage probabilities, but this is a false hope. As shown by Fang and Santos (2014, Corollary 3.1), the nonnormality of Theorem 3 implies that the conventional bootstrap will be inconsistent. Indeed, we investigated the coverage probabilities of confidence intervals constructed using the percentile and inverse percentile methods, and their coverage rates are similar to that shown in Figure 5 (though less distortion for large x), so are not displayed here.

Constructively, Fang and Santos (2014) suggested an alternative bootstrap method that is consistent for the asymptotic distribution. Their suggestion is to approximate the distribution of $\sqrt{n}(g(\hat{\theta}) - g(\theta_0))$ by that of $\hat{g}_\theta(\sqrt{n}(\hat{\theta}^* - \hat{\theta}))$, where $\hat{\theta}^*$ is the bootstrap distribution of $\hat{\theta}$ and $\hat{g}_\theta(h)$ is an estimate of $g_\theta(h)$. Confidence intervals can be constructed from this alternative bootstrap distribution.

The Fang–Santos alternative bootstrap requires that the directional derivative $g_\theta(h)$ be either known or consistently estimated. Fortunately, the problem of estimating the directional derivative has been solved in a recent article by Hong and Li (2015). Their suggestion is to estimate (15) with a discrete

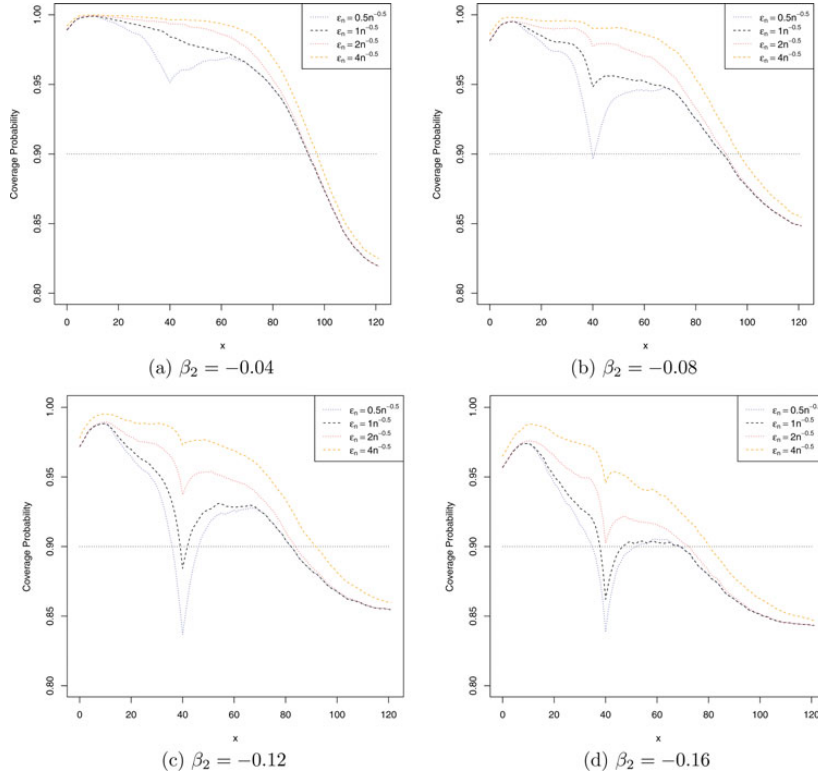


Figure 6. Coverage probabilities of nominal 90% numerical delta method bootstrap intervals.

analog. Specifically, for some sequence $\varepsilon_n > 0$ satisfying $\varepsilon_n \rightarrow 0$ and $\sqrt{n}\varepsilon_n \rightarrow \infty$, estimate the directional derivative $\phi_\theta(h)$ of $\phi(\theta)$ by

$$\hat{\phi}_\theta(h) = \frac{\phi(\hat{\theta} + h\varepsilon_n) - \phi(\hat{\theta})}{\varepsilon_n}.$$

Hong and Li (2015) showed that this produces a consistent estimate of $\phi_\theta(h)$ when $\phi(\theta)$ is Lipschitz continuous. They called $\hat{\phi}_\theta(h)$ “the numerical delta method.”

In our case, the regression function $g(\theta)$ is Lipschitz continuous in θ and the directional derivative $g_\theta(h)$ takes the semilinear form (16), so we only need to estimate the nonlinear component (17). The Hong–Li estimate is

$$\hat{g}_\gamma(h_\gamma) = \frac{g(\hat{\beta}, \hat{\gamma} + h_\gamma \varepsilon_n) - g(\hat{\beta}, \hat{\gamma})}{\varepsilon_n},$$

where we have written $g(\hat{\beta}, \hat{\gamma}) = g(\hat{\theta})$. Thus, our estimate of the full directional derivative is

$$\hat{g}_\theta(h) = x(\hat{\gamma})' h_\beta + \frac{g(\hat{\beta}, \hat{\gamma} + h_\gamma \varepsilon_n) - g(\hat{\beta}, \hat{\gamma})}{\varepsilon_n}.$$

Evaluated at $h = (h_\beta, h_\gamma) = (\sqrt{n}(\hat{\beta}^* - \hat{\beta}), \sqrt{n}(\hat{\gamma}^* - \hat{\gamma}))$, the bootstrap estimate of the distribution of $g(\hat{\theta}) - g(\theta_0)$ is

$$r^* = x(\hat{\gamma})' (\hat{\beta}^* - \hat{\beta}) + \frac{g(\hat{\beta}, \hat{\gamma} + \sqrt{n}\varepsilon_n(\hat{\gamma}^* - \hat{\gamma})) - g(\hat{\beta}, \hat{\gamma})}{\sqrt{n}\varepsilon_n}.$$

Hong and Li (2015) showed that if the function $\phi(\theta)$ is a convex function of θ then upper one-sided confidence intervals for θ constructed using r^* uniformly control size, but lower one-sided confidence intervals will not, and the converse holds when $\phi(\theta)$ is a concave function. In our case, the function $g(\theta)$ is a convex function of γ if $\beta_1 \leq \beta_2$ but is concave if $\beta_1 \geq \beta_2$. Thus, neither one-sided confidence intervals will uniformly control size. Their suggestion is to instead use two-sided symmetric confidence intervals, as these will be asymptotically conservative.

Let $q_{1-\alpha}^*$ denote the $(1-\alpha)$ th quantile of the variable $|r^*|$. The symmetric bootstrap confidence interval for $g(\theta_0)$ is $g(\hat{\theta}) \pm q_{1-\alpha}^*$. We call this the numerical delta method bootstrap interval.

An important issue is setting ε_n , the increment for the numerical derivative. While the general theory of Hong and Li (2015) requires $\sqrt{n}\varepsilon_n \rightarrow \infty$, in the context of this model they recommended $\varepsilon_n = cn^{-1/2}$, though they provide no guidance for selection of c . We follow their advice and set $\varepsilon_n = cn^{-1/2}$, and assess the choice of the constant c via simulation. We implemented the numerical delta method bootstrap, and calculated in the same simulation as Figure 5 the pointwise coverage probabilities of nominal 90% confidence intervals with $c = 0.5$, $c = 1$, $c = 2$, and $c = 4$. The coverage probabilities are plotted in Figure 6, where each panel corresponds to a value of β_2 .

Figure 6 shows that the coverage probabilities are monotonically increasing in c , so a more conservative (and larger)

confidence interval can be found by increasing c . The coverage probabilities are also nonuniform in the regression argument x , with overcoverage for small x and undercoverage for large x . There is also a downward spike in coverage probabilities at $x = 40$, which is the location of the true threshold and is consistent with the nonnormal distribution theory. This downward spike is lessened for larger c but is also increasing with β_2 . The simulations indicate the appropriate choice of c depends on β_2 thus making a generic choice difficult. In our application, our point estimate for $\beta_2 - \beta_1$ (which is the appropriate analog) is -0.10 , which is roughly intermediate between panels (b) and (c). If we focus on obtaining correct coverage at $x = 40$, the results suggest $c = 1$ is reasonable, with over coverage for $\beta_2 = -0.08$ and mild undercoverage for $\beta_2 = -0.12$. We thus set $c = 1$ and hence $\varepsilon_n = n^{-1/2}$. For other applications, we suggest evaluating the coverage via simulation.

The theoretical argument for the numerical delta method bootstrap is that it has asymptotically conservative coverage for all values of x , even at $x = \gamma_0$. For values of x distant from γ_0 , however, it is possible that the classical bootstrap will have better (e.g., less conservative) coverage. There is no clear way, however, to mix bootstrap methods, and the value of γ_0 is unknown in practice, so we recommend using the numerical delta method bootstrap for confidence intervals for all values of x .

We now summarize the steps for the numerical delta method bootstrap.

Algorithm 3: Numerical Delta Method Bootstrap Confidence Intervals for Regression Kink Function at a Fixed Value of (x, z) .

1. Follow Steps 1–3 of the wild bootstrap of Algorithm 2
2. Set $x(\hat{\gamma}) = ((x - \hat{\gamma})_-, (x - \hat{\gamma})_+, z')'$
3. Set $\varepsilon_n = cn^{-1/2}$
4. Set $r^* = x(\hat{\gamma})'(\hat{\beta}^* - \hat{\beta}) + (g(\hat{\beta}, \hat{\gamma}) + \sqrt{n}\varepsilon_n(\hat{\gamma}^* - \hat{\gamma})) - g(\hat{\beta}, \hat{\gamma})/\sqrt{n}\varepsilon_n$
5. Repeat B times, so as to obtain a sample of simulated estimates r^*
6. Calculate $q_{1-\alpha}^*$, the $(1 - \alpha)$ th quantile of the simulated $|r^*|$
7. Set the $1 - \alpha$ confidence interval for $g(\theta_0)$ as $[g(\hat{\theta}) - q_{1-\alpha}^*, g(\hat{\theta}) + q_{1-\alpha}^*]$

This is numerically quite simple to implement, although step 4 is unusual for a bootstrap procedure. For confidence interval bands, steps 2 through 7 need to be repeated for each value of (x, z) considered.

Formally, the merging of the Hong–Li numerical delta method with this specific wild bootstrap for serially dependent data has not been studied. More work would be welcome to understand its properties.

We implemented the numerical delta method bootstrap using the rule $\varepsilon_n = cn^{-1/2}$ on the U.S. growth regression for $c = 0.5$, $c = 1$, $c = 2$, and $c = 4$, and plotted the pointwise 90% confidence intervals in Figure 7. Numerically, this means that Algorithm 3 was applied with z set at the sample mean of z_t and x evaluated on a grid from 1 to 120. The confidence intervals widen as ε_n increases, except at $x = \hat{\gamma}$.

It is important to interpret these confidence intervals as pointwise in x (not uniform). There have been recent advances in developing uniform inference methods for nonparametric re-

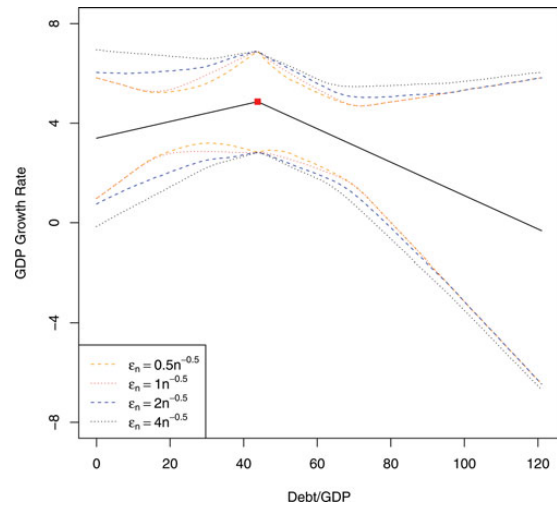


Figure 7. Estimated regression kink function and 90% numerical delta method confidence intervals.

gression by Belloni et al. (in press), but these methods do not apply to the numerical delta method bootstrap.

Following the recommendation of our simulation study, we take the second estimator (with $\varepsilon_n = n^{-1/2}$) as our preferred choice, and this is plotted in Figure 2 with the dashed blue lines. The confidence intervals are sufficiently wide that it is unclear if the true regression function is flat or downward sloping. The confidence intervals reveal that by using the U.S. long-span data alone, the estimates of the regression kink model are not sufficiently precise to make a strong conclusion about whether or not there is a negative effect of debt levels on GDP growth rates.

7. CONCLUSION

This article developed a theory of estimation and inference for the regression kink model with an unknown threshold and applied it to study the growth and debt threshold problem of Reinhart and Rogoff (2010). An interesting theoretical contribution is the finding that the estimate of the regression function is nonnormal due to nondifferentiability, and confidence intervals can be formed using the recent inference methods of Fang and Santos (2014) and Hong and Li (2015).

We apply the method to the long-span time-series U.S. data developed by Reinhart and Rogoff (2010). Our point estimates are consistent with the Reinhart–Rogoff hypothesis of a growth slowdown when debt levels exceed a threshold. However, the formal evidence for the presence of the threshold effect is inconclusive, and our confidence intervals for the regression function are sufficiently wide that the effect of debt on growth is difficult to detect.

An important caveat is that our empirical results are based only on a single time-series (the United States), thus ignoring the information in other nations' experiences. This has the advantage of not imposing homogeneity, but also may reduce

precision and power. It would be useful to extend the results here to panel data analysis.

APPENDIX: MATHEMATICAL PROOFS

Proof of Theorem 2: By the definition (4), $\hat{\theta}$ minimizes $S_n(\theta)$ defined in (3), which we can write as $S_n(\theta) = \frac{1}{n} \sum_{i=1}^n e_i(\theta)^2$ with $e_i(\theta) = y_i - \beta' x_i(\gamma)$. Thus, $\hat{\theta}$ approximately solves the first-order condition $\frac{1}{n} \sum_{i=1}^n m_i(\hat{\theta}) = 0$ where $m_i(\theta) = H_i(\theta) e_i(\theta)$. The pseudo-true value θ_0 minimizes $L(\theta) = L(\beta, \gamma)$ and thus solves $m(\theta_0)$ where $m(\theta) = \frac{\partial}{\partial \theta} L(\theta) = E m_i(\theta)$. Define $Q(\theta) = -\frac{\partial}{\partial \theta'} m(\theta)$.

Andrews (1994), Section 3.2, showed that Theorem 2 will hold under the following conditions.

Condition 1. $\hat{\theta} \rightarrow_p \theta_0$

Condition 2. $\frac{1}{\sqrt{n}} \sum_{i=1}^n H_i e_i \rightarrow_d N(0, S)$

Condition 3. $Q(\theta)$ is continuous in θ and $Q(\theta_0) = Q$

Condition 4. $v_n(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (m_i(\theta) - m(\theta))$ is stochastically equicontinuous

We first establish Condition 1. Since $x_i(\gamma)$ is continuous in γ , $e_i(\theta)$, and $e_i(\theta)^2$ are continuous in θ . Recalling definition (2) we have the simple bound $\|x_i(\gamma)\|^2 = z_i' z_i + (x_i - \gamma)^2 \leq \|z_i\|^2 + x_i^2 + \gamma^2$. Then by the C_r and Cauchy–Schwarz inequalities,

$$e_i(\theta)^2 \leq 2y_i^2 + 2|\beta' x_i(\gamma)|^2 \leq 2y_i^2 + 2\bar{\beta}^2 (\|z_i\|^2 + x_i^2 + \bar{\gamma}^2), \quad (\text{A.1})$$

where $\bar{\beta} = \sup \{\|\beta\| : \beta \in B\}$ and $\bar{\gamma} = \sup \{\|\gamma\| : \gamma \in \Gamma\}$ that are finite under Assumption 2.2. The bound in (A.1) has finite expectation under Assumption 1.2. Then by Lemma 2.4 of Newey and McFadden (1994) (which is based on Lemma 1 of Tauchen (1985)), $L(\theta) = E e_i(\theta)^2$ is continuous and $\sup_{\theta \in B \times \Gamma} |S_n(\theta) - L(\theta)| \rightarrow_p 0$ as $n \rightarrow \infty$. Lemma 2.4 of Newey and McFadden is stated for iid observations, but the result only requires application of a weak law of large numbers, which holds under Assumption 1 by the ergodic theorem.

Given the compactness of $B \times \Gamma$ and the uniqueness of the minimum θ_0 (by Assumptions 1.2 and 1.1, respectively), Theorem 2.1 of Newey and McFadden (1994) establishes that $\hat{\theta} \rightarrow_p \theta_0$ as $n \rightarrow \infty$, which is Condition 1.

Condition 2 follows by the Herrndorf's (1984) central limit theorem for strong mixing processes, which holds under Assumption 1.1–1.2.

The following bound will be useful to establish Conditions 3 and 4. Let w_i be any vector whose elements are pairwise products of the variables (y_i, x_i, z_i) , and set $q = r/(r-1)$. Note that Assumption 1.2 implies $(E \|w_i\|^{2r})^{1/r} \leq C$ for some $C < \infty$. Let $F(\gamma)$ denote the distribution function of x_i , which satisfies $F(\gamma_2) - F(\gamma_1) \leq \bar{f} |\gamma_2 - \gamma_1|$ under Assumption 1.4. By Hölder's inequality, we find

$$\begin{aligned} E \|w_i 1(\gamma_1 \leq x_i \leq \gamma_2)\|^2 &\leq (E \|w_i\|^{2r})^{1/r} (E |1(\gamma_1 \leq x_i \leq \gamma_2)|^q)^{1/q} \\ &\leq C (F(\gamma_2) - F(\gamma_1))^{1/q} \\ &\leq C \bar{f}^{1/q} |\gamma_2 - \gamma_1|^{1/q}. \end{aligned} \quad (\text{A.2})$$

One implication of this bound is that $E(w_i 1(x_i \leq \gamma))$ is uniformly continuous in γ .

We now establish Condition 3. Note that

$$\begin{aligned} Q(\theta) &= -\frac{\partial}{\partial \theta'} E(H_i(\theta) e_i(\theta)) \\ &= E(H_i(\theta) H_i(\theta')) \\ &\quad + E \begin{pmatrix} 0 & 0 & 0 & e_i(\theta) 1(x_i \leq \gamma) \\ 0 & 0 & 0 & e_i(\theta) 1(x_i > \gamma) \\ 0 & 0 & 0 & 0 \\ e_i(\theta) 1(x_i \leq \gamma) & e_i(\theta) 1(x_i > \gamma) & 0 & 0 \end{pmatrix}. \end{aligned}$$

The elements of this matrix are quadratic functions of β , and functions of γ through moments of the form $E(w_i 1(x_i \leq \gamma))$ where w_i is a product of the variables (y_i, x_i, z_i) . Since (A.2) holds we see that $Q(\theta)$

is continuous in θ . Evaluated at θ_0 we find $Q(\theta_0) = Q$. Thus Condition 3 holds.

We now establish Condition 4 by appealing to Theorem 1, Application 4, case (2.15) of Doukhan, Massart, and Rio (1995). Notice that $m_i(\theta)$ is a quadratic function of β , so the only issue is establishing stochastic equicontinuity with respect to γ . We thus so we simplify notation by writing $m_i(\theta)$ as $m_i(\gamma)$. Notice that we can write $m_i(\gamma) = w_i 1(x_i \leq \gamma)$ where w_i is vector whose elements are products of the variables (y_i, x_i, z_i) . Under Assumption 1.2, $m_i(\gamma)$ has a bounded 2 r th moment and thus satisfies the needed envelope condition.

For any $\delta > 0$ set $N(\delta) = \delta^{-2/q}$ and set $\gamma_k, k = 1, \dots, N$, to be an equally spaced grid on Γ . Note that the distance between the grid-points is $O(N(\delta)^{-1})$. Define $m_{ik}^* = \min[m_i(\gamma_{k-1}), m_i(\gamma_k)]$ and $m_{ik}^{**} = \max[m_i(\gamma_{k-1}), m_i(\gamma_k)]$. Then for each $\gamma \in \Gamma$ there is a $k \in \{1, \dots, N\}$ such that $m_{ik}^* \leq m_i(\gamma) \leq m_{ik}^{**}$. Thus $[m_{ik}^*, m_{ik}^{**}]$ brackets $m_i(\gamma)$. By their construction, the fact that $m_i(\gamma) = w_i 1(x_i \leq \gamma)$, and applying (A.2),

$$\begin{aligned} E \|m_{ik}^{**} - m_{ik}^*\|^2 &= E \|m_i(\gamma_k) - m_i(\gamma_{k-1})\|^2 \\ &\leq E \|w_i 1(\gamma_{k-1} \leq x_i \leq \gamma_k)\|^2 \\ &\leq C \bar{f}^{1/q} |\gamma_k - \gamma_{k-1}|^{1/q} \\ &\leq O(N(\delta)^{-q}) = O(\delta^2). \end{aligned}$$

This means that $N(\delta) = \delta^{-2/q}$ are the L^2 bracketing numbers and $H_2(\delta) = \ln N(\delta) = |\log \delta|$ is the metric entropy with bracketing for the class $\{m_i(\gamma) : \gamma \in \Gamma\}$. This and Assumption 1.1 imply eq. (2.15) of Doukhan, Massart, and Rio (1995) and thus their Theorem 1, establishing stochastic equicontinuity of $v_n(\theta)$ and hence Condition 4.

We have established that Conditions 1–4 hold under Assumptions 1 and 2. As discussed above, this is sufficient to establish Theorem 2. \square

ACKNOWLEDGMENTS

The author acknowledges the support by the National Science Foundation. The author thanks the associate editor, two referees, Zheng Fang, Han Hong, Jessie Li, Tzu-Chi Lin, and Andres Santos for very helpful comments, suggestions, and correspondence that were critical for the results developed here.

[Received February 2015. Revised June 2015.]

REFERENCES

- Andrews, D. W. K. (1994), "Empirical Process Methods in Econometrics," in *Handbook of Econometrics* (Vol IV), eds. R. F. Engle and D. L. McFadden, New York: Elsevier, pp. 2247–2294. [239]
- Belloni, A., Chernozhukov, V., Chetverikov, D., and Kata, K. (in press), "On the Asymptotic Theory for Least Squares Series: Pointwise and Uniform Results," *Journal of Econometrics*. [238]
- Caner, M., Grennes, T., and Koehler-Geib, F. (2010), "Finding the Tipping Point-When Sovereign Debt Turns Bad," Policy Research Working Paper Series 5391, The World Bank. [228]
- Caner, M., and Hansen, B. E. (2004), "Instrumental Variable Estimation of a Threshold Model," *Econometric Theory*, 20, 813–843. [228]
- Card, D., Lee, D., Pei, Z., and Weber, W. (2012), "Nonlinear Policy Rules and the Identification and Estimation of Causal Effects in a Generalized Regression Kind Design," NBER Working Paper 18564. [228]
- Card, D., Mas, A., and Rothstein, R. (2008), "Tipping and the Dynamics of Segregation," *Quarterly Journal of Economics*, 123, 177–218. [228]
- Cecchetti, S. G., Mohanty, M. S., and Zampolli, F. (2011), "The Real Effects of Debt," in *Economic Symposium Conference Proceedings: Federal Reserve Bank of Kansas City*, pp. 145–196. [228]
- Chan, K.-S. (1990), "Testing for Threshold Autoregression," *The Annals of Statistics*, 18, 1886–1894. [228]
- (1991), "Percentage Points of Likelihood Ratio Tests for Threshold Autoregression," *Journal of the Royal Statistical Society, Series B*, 53, 691–696. [228]

- (1993), "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model," *The Annals of Statistics*, 21, 520–533. [228]
- Chan, K.-S., and Tong, H. (1990), "On Likelihood Ratio Tests for Threshold Autoregression," *Journal of the Royal Statistical Society, Series B*, 52, 469–476. [228]
- Chan, K.-S., and Tsay, R. S. (1998), "Limiting Properties of the Least Squares Estimator of a Continuous Threshold Autoregressive Model," *Biometrika*, 45, 413–426. [228,233]
- Chiu, G., Lockhart, R., and Routledge, R. (2006), "Bent-Cable Regression Theory and Applications," *Journal of the American Statistical Association*, 101, 542–553. [228]
- Cox, D., Hansen, B. E., and Jimenez, E. (2004), "How Responsive are Private Transfers to Income?" *Journal of the Public Economics*, 88, 2193–2219. [228]
- Doukhan, P., Massart, P., and Rio, E. (1995), "Invariance Principles for Absolutely Regular Empirical Processes," *Annales de l'Institut H. Poincaré, Probabilités et Statistiques*, 31, 393–427. [239]
- Fang, Z. (2014), "Optimal Plug-in Estimators of Directionally Differentiable Functionals," Working Paper, UCSD. [235]
- Fang, Z., and Santos, A. (2014), "Inference on Directionally Differentiable Functions," Working Paper, UCSD. [228,235,236,238]
- Feder, P. I. (1975), "On Asymptotic Distribution Theory in Segmented Regression Problems—Identified Case," *Annals of Statistics*, 3, 49–83. [228]
- Ganong, P., and Jager, S. (2014), "A Permutation Test and Estimation Alternatives for the Regression Kink Design," Working Paper, Harvard University. [228]
- Hansen, B. E. (1996), "Inference When a Nuisance Parameter is not Identified Under the Null Hypothesis," *Econometrica*, 64, 413–430. [228,231]
- (1999), "Threshold Effects in Non-Dynamic Panels: Estimation, Testing and Inference," *Journal of Econometrics*, 93, 345–368. [228]
- (2000), "Sample Splitting and Threshold Estimation," *Econometrica*, 68, 575–603. [228]
- Herndorf, N. (1984), "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables," *The Annals of Probability*, 12, 141–153. [239]
- Hinkley, D. V. (1969), "Inference About the Intersection in Two-Phase Regression," *Biometrika*, 56, 495–504. [228]
- (1971), "Inference in Two-Phase Regression," *Journal of the American Statistical Association*, 66, 736–743. [228]
- Hirano, K., and Porter, J. (2012), "Impossibility Results for Nondifferentiable Functionals," *Econometrica*, 80, 1769–1790. [235]
- Hong, H., and Li, J. (2015), "The Numerical Directional Delta Method," Working Paper, Stanford University. [228,235,236,237,238]
- Landais, C. (2015), "Assessing the Welfare Effects of Unemployment Benefits Using the Regression Kink Design," *American Economic Review: Economic Policy*, 7, 243–278. [228]
- Lee, S., Park, H., Seo, M. H., and Shin, Y. (2014), "A Contribution to the Reinhart and Rogoff Debate: Not 90 Percent but Maybe 30 Percent," CEMMAP Working Paper. [228,231]
- Lee, S., Seo, M. H., and Shin, Y. (2011), "Testing for Threshold Effects in Regression Models," *Journal of the American Statistical Association*, 106, 220–231. [228]
- Lin, T.-C. (2014), "High-Dimensional Threshold Quantile Regression With an Application to Debt Overhang and Economic Growth," Working Paper, University of Wisconsin. [228,229,230]
- Newey, W. K., and McFadden, D. (1994), "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics* (Vol. IV), eds. R. F. Engle and D. L. McFadden, New York: Elsevier, pp. 2113–2245. [239]
- Porter, J., and Yu, P. (2015), "Regression Discontinuity Designs With Unknown Discontinuity Points: Testing and Estimation," *Journal of Econometrics*, 189, 132–147. [228]
- Ramirez-Rondan, N. (2013), "Maximum Likelihood Estimation of a Dynamic Panel Threshold Model," Working Paper, Central Bank of Peru, <https://sites.google.com/site/eltorramirezrondan/>. [228]
- Reinhart, C. M., and Rogoff, K. S. (2010), "Growth in a Time of Debt," *American Economic Review: Papers and Proceedings*, 100, 573–578. [228,229,238]
- Seo, M. H., and Linton, O. (2007), "A Smoothed Least Squares Estimator for Threshold Regression Models," *Journal of Econometrics*, 141, 704–735. [228]
- Shapiro, A. (1990), "On Concepts of Directional Differentiability," *Journal of Optimization Theory and Applications*, 66, 477–487. [235]
- (1991), "Asymptotic Analysis of Stochastic Programs," *Annals of Operations Research*, 30, 169–186. [236]
- Stambaugh, R. F. (1999), "Predictive Regressions," *Journal of Financial Economics*, 54, 375–421. [232]
- Tauchen, G. (1985), "Diagnostic Testing and Evaluation of Maximum Likelihood Models," *Journal of Econometrics*, 30, 415–443. [239]
- Tong, H. (1983), *Threshold Models in Non-linear Time Series Analysis. Lecture Notes in Statistics* (Vol. 21), Berlin: Springer. [228]
- (1990), *Non-Linear Time Series: A Dynamical System Approach*, Oxford: Oxford University Press. [228]
- Woutersen, T. and Ham, J. C. (2013), "Calculating Confidence Intervals for Continuous and Discontinuous Functions of Parameters," Working Paper, University of Arizona. [235]

附录 C 外文文献译文

题目：具有未知门限的折点回归

作者：Bruce • E • Hansen

摘要：这篇文章探索了未知门限的折点回归模型的估计与推断问题。折点回归模型（或者说连续门限模型）是受限制于处处连续且带有在未知门限处的折点的门限回归模型。我们呈现了估计方法、门限存在性的检验、回归参数的推断与回归函数的推断这四个方面的内容。因为回归函数是一个参数的不可微函数，所以对回归函数的推断是不标准的。我们应用近期提出的方法对不可微函数进行推断。对于由 Reinhart 与 Rogoff 引入的关于经济增长与债务的问题，采用他们的关于美国的长期性时间序列数据证实了这一理论。

1. 介绍

折点回归模型由 Card 等人 (2012) 作为不连续性模型的修正并进行推广。在折点回归模型中，回归函数是连续的但斜率在门限点未知处存在不连续的情况，因此被称为一个“折点”。这一模型已经引起了许多经验主义者的注意，包括 Landais (2015)，Ganong 与 Jager (2014) 的应用。传统的不连续回归模型假定门限已知，但是在一些情况下（例如 Card, Mas 与 Rothstein 在 2008 年的研究）门限是未知的且必须要被估计出来。我们的文章关心后者的这一未知门限的情况。

不连续回归模型与门限回归模型相类似。后者在 Tong (1983, 1990) 的非线性自回归模型的文章中被引入，但是该模型可以在许多非线性回归模型的背景中被应用。大多数文章与方法着重关注不连续（不受约束的）门限模型，在这之中回归模型基于门限指示器被分为两类（或者更多类）。值得注意的是由 Chan 与 Tsay (1998) 引入的连续门限模型，这与带有分段的线性回归片段的折点回归模型形式相同。一个连续门限模型的经济应用是 Cox, Hansen 与 Jimenez (2004) 的研究。在门限效应集中于单变量并且没有理由去期望得到门限位置处的不连续回归因变量的情况下，折点回归模型也许对于这类经验应用具有吸引力。

这篇文章延伸了 Chan 与 Tsay (1998) 的理论，考虑了对于门限效应，回归参数的推断以及回归函数的推断的问题。正如 Chan 与 Tsay (1998) 的文章中所呈现的，我们将注意力集中于回归片段是线性而非非参数情形。这在非参数方法中也许不合适，但中等样本量背景下是合适的。

关于不连续门限回归的研究有大量的文献。对于门限效应的检验问题，相关的贡献包括 Chan (1990, 1991)，Chan 与 Tong (1990)，Hansen (1996)，与 Lee, Seo, Shin (2011) 的文章。对于回归参数的推断，相关的研究包括 Chan (1993)，Hansen (2000) 以及 Seo 与 Linton (2007) 的文献。面板数据方法已经在 Hansen (1999) 与 Ramirez-Rondan (2013) 的

文章中被提出。起作用的变量方法已经在 Caner 与 Hansen (2004) 的文章中被提出。而对于带有未知门限的不连续性回归的估计与推断理论已经在 Porter 与 Yu (2015) 的文章中被提出。

在统计学文章中，相关的模型种类包括两阶段回归模型，分段式回归模型，分割线段模型与弯曲角模型。对于这一领域的重要贡献包括 Hinkley (1969, 1971) 与 Feder (1975) 的工作。也可以在 Chiu, Lockhart 与 Routledge (2006) 的文献中看到他们的研究。

为了图解该模型，我们将折点回归模型应用于由 Reinhart 与 Rogoff (2010) 提出的著名的经济增长与债务问题。他们认为，总债务对经济增长存在非线性影响，特别是当债务占 GDP 的比例超过某个阈值时，总经济增长将趋于放缓。这可以形式化为一个折点回归模型，其中 GDP 增长是因变量，债务/GDP 比率是关键回归变量和门限变量。一些作者主张使用门限回归工具进行计量经济分析，这其中包括 Caner, Grennes 与 Koehler-Geib (2010), Cecchetti, Mohanty 与 Zampolli (2011), Lin (2014) 以及 Lee 等人 (2014) 的研究。这些文章使用各种截面、面板和时间序列回归来研究 Reinhart-Rogoff 命题，但都集中在不连续门限回归模型上。我们通过一个使用长跨度美国时间序列数据的小型调查来补充这一文献中的问题。

在我们的调查中，我们遇到了一个新的技术问题。尽管折点回归模型中的参数估计是渐近正态的（如 Chan 和 Tsay (1998) 所示），但是回归函数本身的估计不是渐近正态的，因为回归函数是一个参数估计的不可微函数。因此，传统的推断方法无法应用于回归函数。为了解决这一问题我们使用了 Fang 与 Santos (2014) 和 Hong 与 Li (2015) 近期提出的推断方法，呈现了一种新的用于回归函数估计的非正态分布理论，并展示构造回归函数的置信区间的数字化 Delta 方法。我们的文章结构如下所示。第二部分介绍

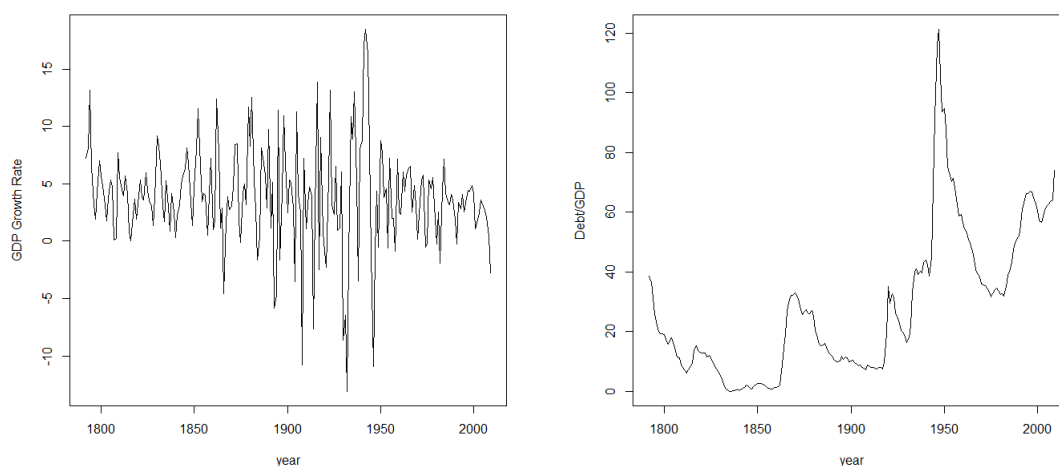


图 C-1 1791-2009 的年度美国真实 GDP 增长率与 GDP/债务比

了模型。第三部分描述了模型参数的最小二乘估计。第四部分讨论了模型背景下的门限效应的检验。第五部分展示了参数估计的渐近分布理论并且讨论了置信区间的构建。第六部分讨论了回归函数的推断。定理C.2的证明被呈现在附录中。

本文的实证分析、模拟的数据和 R 语言代码可以在作者的网页 <http://www.ssc.wisc.edu/bhansen1> 上找到。

2. 模型

我们的折点回归模型具有以下形式：

$$y_t = \beta_1(x_t - \gamma)_- + \beta_2(x_t - \gamma)_+ + \beta_3'z_t + e_t \quad (\text{附录 C-1})$$

其中 y_t , x_t 与 e_t 是标量, 并且 z_t 是包含截距项的 l 维向量。变量 (y_t, x_t, z_t) 在 $t = 1, \dots, n$ 的情形下被观测得到。参数是 $\beta_1, \beta_2, \beta_3$ 与 γ 。我们使用 $(a)_- = \min[a, 0]$ 与 $(a)_+ = \max[a, 0]$ 表示一个实数 a 的“负项部分”与“正项部分”。在模型(附录 C-1)中, 对于变量 x_t , 当 x_t 小于 γ 时, 斜率等于 β_1 , 当 x_t 大于 γ 时, 斜率等于 β_2 , 但回归函数在所有变量中都是连续的。

模型(附录 C-1)是一个折点回归模型是因为回归函数在变量 x 与 z 的位置是连续的, 但是关于 x 的斜率在 $x = \gamma$ 处不连续 (存在一个折点)。模型(附录 C-1)指定回归部分是线性的, 但这可以修改为任何形式 (如多项式)。传统的回归折点设计假定门限点 γ 已知 (这在许多门限已知并且由政策决定的应用中是合适的)。相反, 我们将参数 γ 看成一个需要被估计的未知量。因此, 当门限不是由政策决定的, 或者当研究人员希望调查该假设的稳健性时, 我们的方法是合适的。

模型(附录 C-1)有 $k = 3 + l$ 个参数。 $\beta = (\beta_1, \beta_2, \beta_3)$ 是回归斜率且总体而言不受约束因此 $\beta \in \mathbb{R}^{k-1}$ 。参数 γ 被称为“门限”, “结”或是“折点”。模型(附录 C-1)只有在门限位于门限变量 x_t 的支撑集的内部时才有意义。我们因此假定 $\gamma \in \Gamma$ 其中 Γ 是一个紧集且严格位于 x_t 的支撑集的内部。

折点回归模型被嵌套在不连续门限模型的内部。因此, 我们可以考虑在门限模型类之内检验连续性的假定。这是一个很困难的问题, 我们不知道如何解决这个问题, 因此本文不讨论这个问题。我们简单地假设回归函数是连续的, 并且没有探讨检验连续性假设的问题。

作为一个实证例子, 考虑 Reinhart 和 Rogoff (2010) 的增长债务比的回归问题。他们认为, 当政府债务相对于 GDP 的水平超过一个阈值时, 经济增长往往会放缓。为了将这一现象写成一个回归的形式, 我们设定 y_t 为第 t 年中的真实 GDP 增长率, x_t 为前一年中的债务相对于 GDP 的占比 (因此那似乎是预先被决定的)。我们设定 $z_t = (y_{t-1}, 1)'$, 因此回归包含一个滞后的因变量, 以确保误差 e_t 是近似序列不相关的。我们将重点放在美国并且使用由 Reinhart 和 Rogoff 收集并发布在他们网站上的 1792-2009 年的长跨度时间序列, 因此总共有 $n = 218$ 个观测值。我们在图C-1中显示了这两个序列的时间序列图。我们沿用 Lin (2014) 的方法, 关注单个国家的时间序列估计而不是跨国或面板估计, 因此不施加参数同质性假设。

在图C-2中, 我们显示了 (y_t, x_t) 的散点图以及拟合的回归线和逐点的 90% 置信区

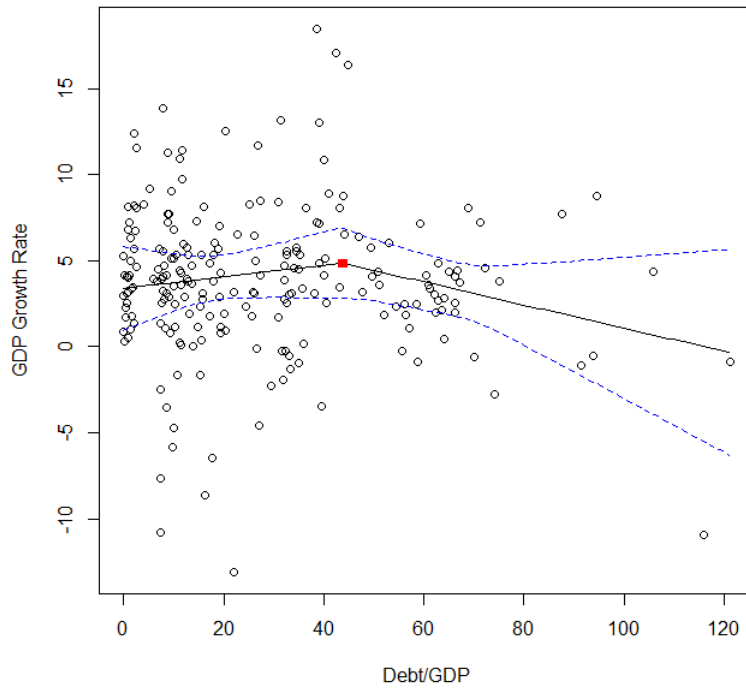


图 C-2 带有估计的折点回归模型与 90% 置信区间的真实 GDP 增长与债务/GDP 的散点图

间（我们将在下一节讨论估计，并在第 6 节讨论置信区间）。我们可以看到拟合回归显示，对于低债务比率情形有一个小的正斜率，折点（门限）大约在 44% 左右（显示为红色方块），而对于高于该值的债务比率则转变为一个负斜率。

我们想要在这一模型的背景下考虑统计推断，特别关注以下的问题：(1) 门限回归在统计上与线性回归不同吗？(2) 参数估计的渐近分布是什么，我们如何建立参数的置信区间？(3) 我们可以建立回归线的置信区间吗？

3. 估计

如果模型(附录 C-1)被解释为一个条件均值，那么参数的自然估计是最小二乘。这样写起来很方便：

$$x_t(\gamma) = \begin{pmatrix} (x_t - \gamma)_- \\ (x_t - \gamma)_+ \\ z_t \end{pmatrix} \quad (\text{附录 C-2})$$

因此(附录 C-1)式可以被写为 $y_t = \beta' x_t(\gamma) + e_t$ 。最小二乘的准则为：

$$S_n(\beta, \gamma) = \frac{1}{n} \sum_{t=1}^n (y_t - \beta' x_t(\gamma))^2 \quad (\text{附录 C-3})$$

最小二乘估计 $(\hat{\beta}, \hat{\gamma})$ 是 $S_n(\beta, \gamma)$ 的联合最小化形式:

$$(\hat{\beta}, \hat{\gamma}) = \underset{\beta \in \mathbb{R}^{k-1}, \gamma \in \Gamma}{\operatorname{argmin}} S_n(\beta, \gamma) \quad (\text{附录 C-4})$$

准则函数 $S_n(\beta, \gamma)$ 是关于 β 二次的但是关于 γ 非凸。因此, 使用集中式搜索和网格搜索相结合的方法, 这在计算上应用是很方便的, 这在门限模型中是很典型的。具体来说, 注意到采用集中式算法我们可以写成

$$\hat{\gamma} = \underset{\gamma \in \Gamma}{\operatorname{argmin}} \min_{\beta \in \mathbb{R}^{k-1}} S_n(\beta, \gamma) = \underset{\gamma \in \Gamma}{\operatorname{argmin}} S_n^*(\gamma) \quad (\text{附录 C-5})$$

其中 $\hat{\beta}(\gamma)$ 是来自于 y_t 在固定 γ 情形下对于自变量 $x_t(\gamma)$ 的回归的最小二乘系数, 并且

$$S_n^*(\gamma) = S_n(\hat{\beta}(\gamma), \gamma) = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{\beta}(\gamma)' x_t(\gamma))^2$$

是集成的误差平方和函数。(附录 C-5)的解可以通过对 γ 进行网格搜索来求数值解。在找到 $\hat{\gamma}$ 后系数估计 $\hat{\beta}$ 可以由 y_t 关于 $x_t(\hat{\gamma})$ 的标准最小二乘获得。我们将拟合回归函数写成

$$y_t = \hat{\beta}' x_t(\hat{\gamma}) + \hat{e}_t. \quad (\text{附录 C-6})$$

在(附录 C-6)中, \hat{e}_t 是最小二乘残差。误差方差 $\sigma^2 = E e_i^2$ 的估计为

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n \hat{e}_t^2 = S_n^*(\hat{\gamma}).$$

为了印证这一估计, 我们考虑美国 1792-2009 年的经济增长回归数据。首先, 我们设定门限参数的参数空间 $\Gamma = [10, 70]$, 因此至少 5% 的样本与 10% 的门限变量的支撑集小于参数空间下界或是大于参数空间上界。接着我们通过计算 $S_n^*(\gamma)$ 在一个增量为 0.1 的离散网格上的值 (有 601 个格点) 的方法估计(附录 C-5)的最小值。对于 γ 的每一个格点, 我们估计最小二乘系数并且计算最小二乘准则 $S_n^*(\gamma)$ 。这一准则以 γ 的函数形式被绘制在图C-3中。我们观察发现函数是相当光滑的, 并有一个界限清楚的全局最小值, 但准则不能很好地描述为二次形式。函数绘制图像的相对平滑性表明, 我们对网格评估的选择足够好, 可以获得全局最小值。准则函数在 $\hat{\gamma} = 43.8$ 处达到最小值。有趣的是, 这一门限估计十分接近于 Lin(2014) 使用不同的数据窗口与完全不同的经验数据结构作用于美国的相关数据集时的发现, 并且也非常接近于 Lee 等人 (2014) 用中位数跨国数据回归得到的估计结果。

拟合折点回归模型的参数估计如下:

$$y_t = \underset{(0.026)}{0.033}(x_{t-1} - \underset{(12.1)}{43.8})_- - \underset{(0.048)}{0.067}(x_{t-1} - \underset{(12.1)}{43.8})_+ + \underset{(0.09)}{0.28} y_{t-1} + \underset{(0.69)}{3.78} + \hat{e}_t,$$

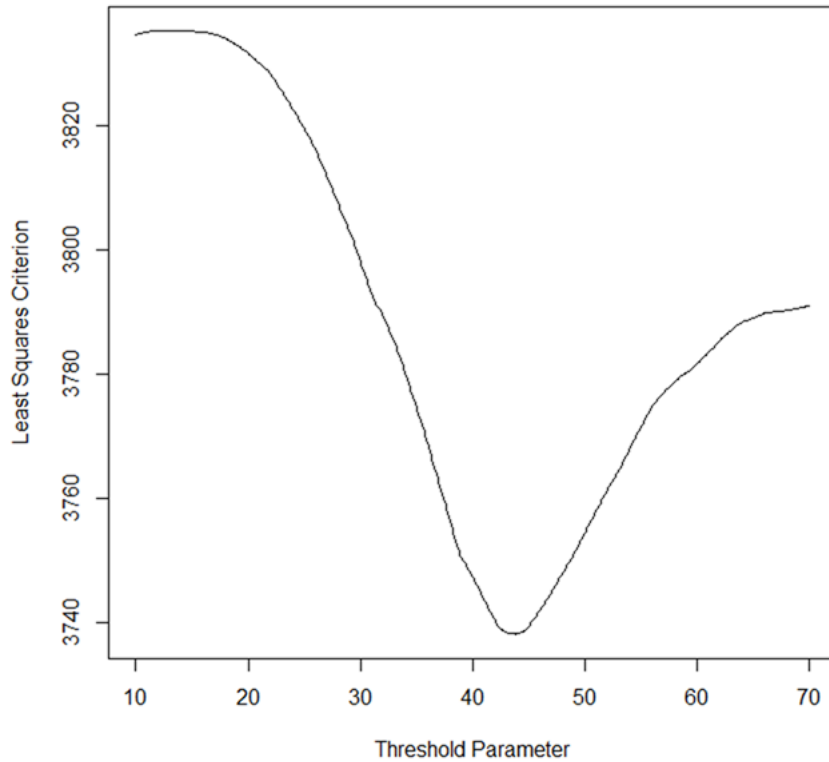


图 C-3 门限参数的集中的最小二乘准则函数

$$\hat{\sigma}^2 = 17.2.$$

标准误差将在第 5 节中讨论。这些估计显示了正的一阶自相关系数为 0.28，这对增长率的时间序列估计是标准的。这些估计表明，高负债率（高于 44%）会导致预期 GDP 增长率的适度放缓，这与 Reinhart-Rogoff 假说一致。

4. 对门限效应的检验

一个合理的问题是，门限模型(附录 C-1)相对于线性模型是否“显著”：

$$y_t = \beta_1 x_t + \beta_3' z_t + e_t. \quad (\text{附录 C-7})$$

线性模型(附录 C-7)被嵌入在门限模型(附录 C-1)中，因为(附录 C-7)满足限制条件 $\beta_1 = \beta_2$ （此处含有对截距项的重新定义）。在这一假设下门限 γ 消失并且在原假设下我们可以看到相类似的未经识别的参数问题。这需要非标准的检验方法。幸运的是，这些问题很容易被理解和应用。

对于线性模型的一个合理估计是最小二乘。我们可以把这些估计写成以下形式：

$$y_t = \tilde{\beta}_1 x_t + \tilde{\beta}_3' z_t + \tilde{e}_t, \quad (\text{附录 C-8})$$

误差项的方差估计为

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n \tilde{e}_t^2.$$

例如，对于我们的实证例子而言，线性模型的估计形式如下：

$$y_t = \underset{(0.013)}{-0.008} x_{t-1} + \underset{(0.09)}{0.30} y_{t-1} + \underset{(0.62)}{2.90} + \tilde{e}_t, \quad (\text{附录 C-9})$$

$$\tilde{\sigma}^2 = 17.6.$$

估计显示一阶自相关系数为正，债务比率的系数接近于零。

接受线性模型(附录 C-7)原假设拒绝门限模型(附录 C-1)的标准检验是对于大于临界值的 F 统计量

$$T_n = \frac{n(\tilde{\sigma}^2 - \hat{\sigma}^2)}{\hat{\sigma}^2}$$

拒绝原假设 $H_0 : \beta_1 = \beta_2$ 。例如，对于我们的实例可以计算得到 $T_n = 5.56$ 。

为了描述检验统计量的分布理论，我们需要精确地描述模型的随机假设。我们需要以下正规性条件。

假设 C.1 对于 $r > 1$,

1. 对于某些 $A > r/(r-1)$ ，带有混合系数 $\eta(m) = O(m^{-A})$ 的 (y_t, z_t, x_t) 是严平稳的，遍历性的且绝对正则的。

2. $E|y_t|^{4r} < \infty, E|x_t|^{4r} < \infty$, 且 $E||z_t||^{4r} < \infty$ 。

3. $\inf_{\gamma \in \Gamma} \det[Q(\gamma)] > 0$ ，其中 $Q(\gamma) = E[x_t(\gamma)x_t(\gamma)']$ 。

4. x_t 有密度函数 $f(x)$ 满足 $f(x) \leq \bar{f} < \infty$ 。

5. $\gamma \in \Gamma$ 其中 Γ 是紧集。□

假设C.1中的第 1、2 条满足中心极限定理的标准弱相关条件。 r 的选择涉及到允许的序列依赖程度和有限矩的数量之间的平衡。对于独立的观测，我们可以让 r 任意接近 1。假设C.1中的第三条是可识别条件，要求对参数空间 Γ 中所有 γ 的值都定义好投影系数。假设C.1中的第四条要求门限变量具有有界密度函数。

特别地，假设C.1允许回归量 x 和 z 包含滞后因变量。

以下定理隐含 Hansen (1996) 的定理 3。唯一的区别是后一个定理着重于异方差-稳健的 Wald 统计量，而为了简单起见，我们考虑检验统计量的更简单的同方差形式。

定理 C.1 假设满足假定 1 且 e_t 是一个鞅差序列。在原假设 $H_0 : \beta_1 = \beta_2$ 下

$$T_n \xrightarrow{d} \sup_{\gamma \in \Gamma} G(\gamma)' Q(\gamma)^{-1} G(\gamma) / \sigma^2, \quad (\text{附录 C-10})$$

其中 $G(\gamma)$ 是一个零均值高斯过程且具有协方差核为

$$E[G(\gamma_1)G(\gamma_2)] = E[x_t(\gamma_1)x_t(\gamma_2)'e_t^2]. \square \quad (\text{附录 C-11})$$

定理C.1表明，门限 F 统计量的渐近零分布可以写成一个随机过程的上界。定理C.1在假设C.1的基础上，增加了误差 e_t 是鞅差序列的正则性条件。这是一个充分条件，但不是必要的。重要的是折点回归模型中的最佳拟合近似是线性模型且中心极限定理适用于所有 γ ，并且因为回归分数是不相关的，这样协方差核具有(附录 C-11)的简单形式而非异方差性和自相关性一致 (HAC) 的形式。

极限分布(附录 C-10)是非标准的并且不能用列表显示出来。然而，正如 Hansen (1996) 所示，使用乘子引导来模拟近似分布(附录 C-10)的形式是简单的，因此可以计算出渐近有效的 p 值。以下是他推荐的算法 (Hansen (1996) 的定理 3 表明，在定理C.1的条件下，该算法产生渐近的一阶修正 p 值)。

- 1: 从正态分布 $N(0, 1)$ 生成 n 个独立同分布的变量 u_t 。
- 2: 设定 $y_t^* = \tilde{e}_t u_t$ 其中 \tilde{e}_t 是拟合的线性回归(附录 C-8)的最小二乘估计的残差。
- 3: 使用观测量 (y_t^*, z_t, x_t) 估计线性模型(附录 C-8)与折点回归模型(附录 C-6)并且计算误差方差估计 $\tilde{\sigma}^{*2}$ 与 $\hat{\sigma}^{*2}$ 与 F 统计量 $T_n^* = \frac{n(\tilde{\sigma}^{*2} - \hat{\sigma}^{*2})}{\hat{\sigma}^{*2}}$ 。
- 4: 重复这一过程 B 次，故可以获得模拟的 F 统计量样本序列 $T_n^*(1), \dots, T_n^*(B)$ 。
- 5: 计算 p 值，即模拟的 F 统计量大于真实值的百分比: $p_n = \frac{1}{B} \sum_{b=1}^B I(T_n^*(b) \geq T_n)$ 。
- 6: 如果需要的话，计算模拟的 F 统计量 $T_n^*(1), \dots, T_n^*(B)$ 的经验 $1 - \alpha$ 分位数作为置信水平 α 的临界值 c_α 。
- 7: 对于显著水平 α 若 $p_n < \alpha$ 或者等价地 $T_n > c_\alpha$ 时拒绝 H_0 接受 H_1 。

算法 C-1 对带有未知门限的回归折点的检验

F 统计量 T_n 的自举复制数量 B 应该设置得相当大，以确保 p 值 p_n 的准确性。例如，如果 $B = 10000$ ，那么 p_n 大约在其极限值（大值 B ）的 0.006 以内。幸运的是，这样做计算开销是较小的。例如，我的办公室计算机仅在 1 分钟内计算了本文中报告的所有实例计算，包括两个单独的自举模拟，其中每个模拟使用了 10000 个复制（计算是在 R 语言中执行的。代码利用自举复制中回归量是共同的这一事实，通过同时执行所有 10000 个回归来有效地实现算法乘子的自助程序）。

回到美国 GDP 的例子，我们之前说过，F 统计量的经验值是 $T_n = 5.66$ 。利用乘子自举计算得到的 p 值为 0.15，自举估计的 10% 临界值为 7.1。因此，在 10% 水平下，检验不会拒绝线性的原假设而接受折点回归模型。

乘子自举法不能解释观测量的时间序列性质，因此预期会出现一些有限的样本失真。特别是，假设C.1不考虑非平稳或接近非平稳的回归量。正如 Stambaugh (1999) 等人所论证的那样，高度持续的回归变量会扭曲正态分布理论。这与我们的例子特别相关，如图C-1所示，债务/GDP 水平显示出相当大的序列依赖性。

为了研究这种可能性，我们基于校正的实例提出了一个简单的模拟试验。我们的

数据生成过程是

$$y_t = \beta_1(x_t - \gamma)_- + \beta_2(x_t - \gamma)_+ + \beta_3 y_{t-1} + \beta_4 + e_t, \quad e_t \sim \mathbf{N}(0, \sigma^2).$$

为了评估样本量，我们设定 $\beta_1 = \beta_2 = 0$, $\beta_3 = 0.3$, $\beta_4 = 3$ 与 $\sigma^2 = 16$ 以匹配估计(附录 C-9)。我们设定 x_t 为美国的债务与 GDP 比值的实际值，并且如实例中那样设定 $n = 218$ 。

表 C-1 带有多重自助法模拟的门限 F 检验的势，名义上的检验水平为 10%

	β_2							
	-0.02	-0.04	-0.06	-0.08	-0.10	-0.12	-0.14	-0.16
检验的势	0.14	0.24	0.38	0.54	0.72	0.85	0.93	0.98

精确地设置 x_t 使其等于样本值，是为了使模拟在具有观察到的债务/GDP 序列的相关性的设定中评估性能。我们从这个过程中生成了 10000 个样本。为了加快计算速度，对于本文中的所有模拟，我们使用增量为 1 的网格来评估 $\Gamma = [10, 70]$ ，这样网格点数就减少到了 61 个。每个模拟样本的自举复制次数被设置为 $B = 1000$ 。

我们首先评估门限检验的 p 值的大小。在 10% 的名义显著性水平下，我们发现模拟大小为 10.6%。因此。检验没有显示有意义的检验程度的失真。

其次，我们评估了检验的势。我们使用了与上面相同的参数化方法，但现在设置 $\beta_1 = 0$ 和 $\gamma = 40$ ，并以 0.02 的步长使得 β_2 从 0 到 -0.16 变化。检验的势如表 C-1 所示。我们可以看到，该检验在 β_2 中势逐渐增加，并且具有检测小至 0.08 的变化的合适的能力。在检测门限模型的点估计中，我们可以看到，回归斜率的差异的估计为 $\hat{\beta}_1 - \hat{\beta}_2 = 0.10$ ，其中这次模拟表明，势应当在 70% 左右，这是合理但不太大的。由此可见，由于检验的势不太大，我们的经验 p 值为 15%。我们得出结论关于由于高债务率产生的是否在 GDP 增长中有回归折点效应的问题，门限检验是非决定性的。

5. 回归系数的推断

本节我们在假设门限已识别的情况下，考虑具有未知门限的折点回归模型的最小二乘估计的分布理论。

在理想情况下，我们可以将(附录 C-1)视为真正的回归函数，因此误差的条件均值为 0。为了允许额外的普遍性，我们不会加强这个条件。相反，我们将(附录 C-1)视为最佳近似，因为它使平方误差损失最小化。我们定义最佳的近似为真值 (β_0, γ_0) ，使平方损失

$$L(\beta, \gamma) = E[y_t - \beta' x_t(\gamma)]^2 \quad (\text{附录 C-12})$$

最小。

正如我们对于估计的分析，我们可以通过集中的方式定义极小化变量。设 $\beta(\gamma)$ 是对于固定的 γ 关于 β 的 $L(\beta, \gamma)$ 的极小化变量，即 $\beta(\gamma) = E[x_t(\gamma)x_t(\gamma)']^{-1}E[x_t(\gamma)y_t]$ 。在假定 1.3 的条件下，对于所有的 $\gamma \in \Gamma$ ， $\beta(\gamma)$ 是唯一定义的。集中化损失函数为

$$L^*(\gamma) = L(\beta(\gamma), \gamma) = E[y_t - \beta(\gamma)'x_t(\gamma)]^2$$

通过集中变量的方法， γ_0 是 $L^*(\gamma)$ 的极小化变量并且 $\beta_0 = \beta(\gamma_0)$ 。

我们需要极小值 γ_0 是唯一的。这排除了最佳拟合线性模型的情况（在这种情况下 $L^*(\gamma)$ 是一个常数函数）和多重最佳拟合门限参数 γ 的情况。

为了简化一致性的证明，我们还要求 β 的参数空间是紧集，尽管这个假设可以通过一个更细节的论据被放宽。

假设 C.2 对于损失函数 $L^*(\gamma)$ 及参数 β 而言

1. $\gamma_0 = \underset{\gamma \in \Gamma}{\operatorname{argmin}} L^*(\gamma)$ 是唯一解。
2. $\beta \in B \subset \mathbb{R}^{k-1}$ ，其中 B 为紧集。□

Chan 和 Tsay (1998) 证明在连续门限自回归模型中的最小二乘估计，同时包括斜率与门限系数的估计，是联合渐近正态的。我们将他们的分布理论拓展至带有未知门限的折点回归模型。设定 $\theta = (\beta, \gamma)$ ， $\theta_0 = (\beta_0, \gamma_0)$ ，

$$H_t(\theta) = -\frac{\partial}{\partial \theta}(y_t - \beta'x_t(\gamma)) = \begin{pmatrix} x_t(\gamma) \\ -\beta_1 I(x_t < \gamma) - \beta_2 I(x_t > \gamma) \end{pmatrix}$$

并且 $H_t = H_t(\theta_0)$ 。

定理 C.2 在满足假设 C.1 时，

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V)$$

其中

$$V = Q^{-1}SQ^{-1}, S = \sum_{j=-\infty}^{\infty} E(H_t H_{t+j}' e_t e_{t+j})$$

且

$$Q = E(H_t H_t') + E \begin{pmatrix} 0 & 0 & 0 & e_t I(x_t < \gamma_0) \\ 0 & 0 & 0 & e_t I(x_t > \gamma_0) \\ 0 & 0 & 0 & 0 \\ e_t I(x_t < \gamma_0) & e_t I(x_t > \gamma_0) & 0 & 0 \end{pmatrix}. \square$$

定理C.2的证明被呈现在附录中。

注意到，斜率和门限估计联合是渐近正态的，具有 \sqrt{n} 的收敛速度，并且斜率和门限估计具有一个非零的渐近协方差。而恰恰相反的是，在传统的非连续门限模型中，门限估计 $\hat{\gamma}$ 以速率 n 一致收敛于一个非标准渐近分布，且斜率系数估计是渐近独立于门限估计的。折点回归模型的存在差异是因为回归函数是连续的。因此，最小二乘准则函数在参数上是连续的（尽管不可微），并且是渐近局部二次形式的。

由于门限估计只有 \sqrt{n} 的收敛速度，我们应该预料它的精度低于非连续情况下的门限估计。因此在构造置信区间时，考虑其抽样分布是很重要的。

渐近分布不要求模型被准确指定，所以误差 e_t 不需要是鞅差序列。因此（通常来说），协方差矩阵 S 具有 HAC 形式。当回归被很好地动态指定（通过适当地包含滞后变量），那么矩阵将简化为 $S = E(H_t H_t' e_t^2)$ 。因此我们的应用包含一个滞后自变量。我们实际上用这种方式简化运算。

当门限模型被正确构建时， Q 的定义式的第二项为 0 故 $E(e_t|x_t) = 0$ 。然而，在模型无法指定的情况下（以至于回归是最佳近似时），第二项可以是非零的。

我们认为渐近协方差矩阵的估计如下所示（假定不相关性前提）。设定

$$\hat{H}_t = \begin{pmatrix} x_t(\hat{\gamma}) \\ -\hat{\beta}_1 I(x_t < \hat{\gamma}) - \hat{\beta}_2 I(x_t > \hat{\gamma}) \end{pmatrix}$$

$$\hat{V} = \hat{Q}^{-1} \hat{S} \hat{Q}^{-1}, \hat{S} = \frac{1}{n-k} \sum_{t=1}^n \hat{H}_t \hat{H}_t' \hat{e}_t^2$$

且

$$\hat{Q} = \frac{1}{n} \sum_{t=1}^n \left(\hat{H}_t \hat{H}_t' + \begin{pmatrix} 0 & 0 & 0 & \hat{e}_t I(x_t < \hat{\gamma}) \\ 0 & 0 & 0 & \hat{e}_t I(x_t > \hat{\gamma}) \\ 0 & 0 & 0 & 0 \\ \hat{e}_t I(x_t < \hat{\gamma}) & \hat{e}_t I(x_t > \hat{\gamma}) & 0 & 0 \end{pmatrix} \right)$$

对于 \hat{S} 的定义式我们除以 $n-k$ 而不是 n 。这是一个特别的自由度调整。给定 \hat{V} ，系数估计的标准误差是通过取 $n^{-1} \hat{V}$ 的对角元素的平方根得到的。如果动态地指定了错误的回归（例如，如果没有包含滞后的因变量），那么可以使用标准的 HAC 估计量来形成 \hat{S} 。

可以使用传统规则形成系数的渐近置信区间，例如，对于 β_1 的 95% 区间为 $\hat{\beta}_1 \pm 1.96s(\hat{\beta}_1)$ 。定理C.2证明了这样的置信区间有渐近的正确覆盖率。

然而，在小样本情形下，渐近置信区间可能覆盖率较差。如图C-2所示，最小二乘准则对于门限参数 γ 来说是非二次的，这意味着二次逼近（例如正态逼近）可能效果很差，除非样本容量相当大。在这种情况下，通过反转检验置信集可以获得更好的覆盖

率。这对于门限参数 γ 特别方便，因为反转检验置信区域是最小二乘最小化计算的天然副产品。具体来说，为了检验原假设 $H_0: \gamma = \gamma_0$ 与备择假设 $H_1: \gamma \neq \gamma_0$ ，基于标准的检验是拒绝 F 统计量 $F_n(\gamma_0)$ 的较大值，其中

$$F_n(\gamma) = \frac{n(\hat{\sigma}^2(\gamma) - \hat{\sigma}^2)}{\hat{\sigma}^2} \quad (\text{附录 C-13})$$

且当 γ 固定时 $\hat{\sigma}^2(\gamma) = S_n^*(\gamma)$ 是残差方差的估计。根据定理 C.2 的渐近正态性，该检验在 H_0 下有一个渐近的 χ_1^2 分布。因此，对于一个名义上的置信水平为 α 的检验我们可以从 χ_1^2 分布获取临界值 $c_{1-\alpha}$ 。通过反演检验可以得到 γ 的名义上的 $1 - \alpha$ 渐近置信区间： $F_n(\gamma)$ 小于 χ_1^2 分布的临界值的 γ 的集合：

$$C_\gamma = \{\gamma : F_n(\gamma) \leq c_{1-\alpha}\}$$

假定 $\hat{\sigma}^2(\gamma)$ 已经在网格上被计算出来（对于估计而言），在同一网格算法中我们有 $F_n(\gamma)$ 作为副产品。区间 C_γ 可以作为满足 $F_n(\gamma)$ 小于 $c_{1-\alpha}$ 的 γ 的格点的集合被计算出来。

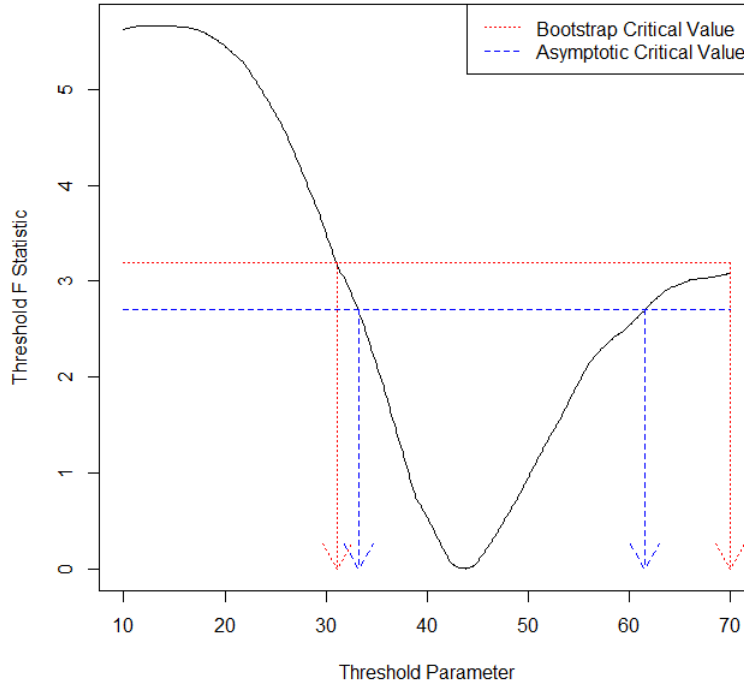


图 C-4 门限的置信区间的构建

为了图解说明，观察图 C-4。这里我们从(附录 C-13)式绘制出了统计量 $F_n(\gamma)$ 作为 γ 函数的图像。通过构建这一函数，发现统计量非负并且在 $\gamma = \hat{\gamma}$ 处达到 0 值。我们已经用蓝色的虚线画出了渐近 90% 临界值 $c_{.9} = 2.7$ 。交叉点表示渐近置信区间的范围，通过蓝色虚线箭头指向横轴在图像上表示出来。

可以通过一个辅助程序进一步提高覆盖精度。因为回归元 x ，显示出相当大的序

列依赖性，我们建议使用一个作用于 x_t 的值的粗糙自助方法。这比其他自助方法如更适合轻度序列依赖情况的块状自助方法更好地保有时间序列特性。也可以使用基于模型的自助方法，但债务/GDP 序列需要一个明确的模型。为了避免这些挑战，我们采取了一种粗糙自助的算法。以下是步骤。

- 1: 从正态分布 $N(0,1)$ 生成 n 个独立同分布变量 u_t 。
- 2: 设定 $e_t^* = \hat{e}_t u_t$ 其中 \hat{e}_t 是来自拟合的折点回归模型(附录 C-6)的最小二乘残差。
- 3: 设定 $y_t^* = \hat{\beta}' x_t(\hat{\gamma}) + e_t^*$ ，其中 (z_t, x_t) 是样本观测值，且 $(\hat{\beta}, \hat{\gamma})$ 是最小二乘估计。
- 4: 使用观测值 (y_t^*, z_t, x_t) 估计折点回归模型(附录 C-6)，得到参数估计 $(\hat{\beta}^*, \hat{\gamma}^*)$ ，并且 $\hat{\sigma}^{*2} = n^{-1} \sum_{t=1}^n \hat{e}_t^{*2}$ ，其中 $\hat{e}_t^* = y_t^* - \hat{\beta}^{*'} x_t(\hat{\gamma}^*)$ 。
- 5: 计算 γ 的 F 统计量 $F_n^* = \frac{n(\hat{\sigma}^{*2}(\hat{\gamma}) - \hat{\sigma}^{*2})}{\hat{\sigma}^{*2}}$ ，其中 $\hat{\sigma}^{*2} = n^{-1} \sum_{t=1}^n \hat{e}_t^*(\hat{\gamma})^2$ 并且 $\hat{e}_t^*(\hat{\gamma}) = y_t^* - \hat{\beta}^*(\hat{\gamma})' x_t(\hat{\gamma})$ 。
- 6: 重复 B 次，以至于得到模拟系数估计 $(\hat{\beta}^*, \hat{\gamma}^*)$ 的样本与 F 统计量 F_n^* 。
- 7: 通过对称分位数的方法构建系数 $\beta_1, \beta_2, \beta_3$ 的 $1 - \alpha$ 自助法置信区间: 系数估计加减绝对中心自助法估计的 $1 - \alpha$ 分位数。例如，对于 β_1 而言区间为 $\hat{\beta}_1 \pm q_1^*$ ，其中 q_1^* 为 $|\hat{\beta}_1^* - \hat{\beta}_1|$ 的 $1 - \alpha$ 分位数。
- 8: 计算模拟的 F 统计量 F_n^* 的 $1 - \alpha$ 分位数 $c_{1-\alpha}^*$ 。
- 9: 以满足 γ 的经验 F 统计量 $F_n(\gamma)$ (附录 C-13) 小于自助法临界值 $c_{1-\alpha}^*$ 的 γ 的集合作为 γ 的 $1 - \alpha$ 自助置信区间: $C_\gamma^* = \{\gamma : F_n(\gamma) \leq c_{1-\alpha}^*\}$ 。

算法 C-2 参数的粗糙自助置信区间

这种粗糙自助算法可以与用于门限检验的自助法乘子并行计算，因此计算效率高。我们没有一个正式分布理论来证明使用这种（或任何其他）自助方法构建置信区间的正确性。然而，在给出定理C.2的渐近正态性的基础上，我们没有理由认定自助算法是失败的。

再次通过图C-4对置信区间构建进行图解，其中统计量 $F_n(\gamma)$ 作为 γ 的函数被绘制出来了。自助算法 90% 临界值 3.3（通过 $B = 10000$ 次自助法复制计算）用红色虚线被绘制出来。交点表示置信区间的范围，在图中以红色虚线箭头指向水平轴表示出来。由于自助临界值大于渐近临界值，因此渐近区间是自助置信区间的子集。

在呈现经验结果之前，我们报告用于评估方法性能的模拟试验的结果。数据生成过程与前一节中使用的过程相同。与前面一样， β_2 是控制门限效应强度的关键自由参数，其余参数和变量的设置与实例数据相匹配。特别地，我们固定了回归变量 x ，让它等于美国的债务/GDP 的真实值，这是为了精确地保持其强序列依赖性。与前面一样，我们生成了 10000 个模拟样本，并为每个样本使用了 $B = 1000$ 个自助法重复进行模拟。在表C-2和表C-3中，我们报告了参数 β_2 和 γ 的名义上的 90% 区间的覆盖率。

表C-2给出了 β_2 的置信区间。第一行报告了渐近“加或减”标准差置信区间的覆盖率。我们看到置信区间的覆盖率，大约达到了 81%。第二行报告了百分位数置信区间的覆盖率。这些覆盖率比较好，特别是对于 β_2 的中间值。在第三行中，我们报告了

“反向百分位自助法”的覆盖率，它们的端点是系数估计的两倍，小于模拟系数估计 $\hat{\beta}_2^*$ 的 $1 - \alpha/2$ 分位数大于 $\alpha/2$ 分位数。它的表现并没有优于百分位数区间。在第四行中，我们报告了“对称百分位数自助法”的覆盖率，其端点等于系数估计加减模拟统计量 $|\hat{\beta}_2^* - \hat{\beta}_2|$ 的 $1 - \alpha$ 分位数。这些区间仅仅轻微地没有覆盖真值，覆盖率达到了大约 86%-88%。这些是我们对于斜率系数推荐的置信区间。

表 C-2 对于 β_2 的名义上 90% 置信区间的覆盖率

	β_2							
	-0.02	-0.04	-0.06	-0.08	-0.10	-0.12	-0.14	-0.16
$\hat{\beta}_2 \pm 1.645s(\hat{\beta}_2)$	0.81	0.80	0.81	0.81	0.81	0.80	0.81	0.81
分位数自助法	0.81	0.83	0.85	0.87	0.86	0.86	0.86	0.86
倒转分位数自助法	0.84	0.84	0.85	0.83	0.82	0.82	0.82	0.83
对称分位数自助法	0.86	0.86	0.87	0.88	0.87	0.87	0.86	0.86

在表C-3中，我们报告了参数 γ 的名义上 90% 置信区间的覆盖率。第一行报告了渐近“加或减”标准误差置信区间的覆盖率。这些区间没有覆盖真值的情况较为严重，覆盖率低至 67%。第二行和第三行报告传统百分位数区间和逆百分位数区间的覆盖率。两者表现都不好，对于较小的 β_2 ，百分位数区间基本上过度覆盖真值，而逆百分位数区间没有覆盖真值的情况很严重。第四行报告对称百分位数间隔的覆盖率。其性能优于其他百分位区间，但覆盖率对 β_2 的值很敏感。第 5 行和第 6 行报告了反转检验置信区间 C_γ 和 C_γ^* 的覆盖率。这两个都有很好的覆盖率，特别是使用了自助临界值的 C_γ^* 。区间只有轻微的没有覆盖不够的情况，对 β_2 而言有大致一致的覆盖。区间 C_γ^* 是我们对于 γ 的推荐置信区间。

我们的模拟结果表明，推荐的粗糙自助置信区间工作性能良好，但有一些轻微的欠覆盖情况出现。这可能是因为我们的模拟的数据是一个时间序列自回归（因此具有时间序列依赖性）数据，而粗糙自助无法解释时间序列依赖性。我们推荐粗糙自助算法（而不是基于模型的自助算法）的原因是，粗糙自助能够根据观察到的回归过程来设置条件，并且（最重要的是）可以非参数化地处理条件异方差情形。这些都是重要的优势和便利之处。

我们现在在表C-4中给出了经验回归的系数估计的置信区间，用我们推荐的自助算法设定 $B = 10000$ 次自助重复次数进行计算。在我们的评估中，两个最重要的系数是 β_2 （高负债率下债务对 GDP 增长的斜率效应）和 γ （门限水平）。 β_2 的置信区间为 $[-0.18, -0.01]$ 。这个范围足够大以至于我们无法准确推断债务对预期 GDP 增长的影响程度。门限的 90% 置信区间为 31% - 70%，也表明存在很大的不确定性。

门限的宽置信区间可能是门限估计以 \sqrt{n} 为收敛速度的特征所导致的结果，这也可能是由于样本量小所导致的。

表 C-3 对于 γ 的名义上 90% 置信区间的覆盖率

	β_2							
	-0.02	-0.04	-0.06	-0.08	-0.10	-0.12	-0.14	-0.16
$\hat{\gamma} \pm 1.645s(\hat{\gamma})$	0.67	0.68	0.69	0.70	0.71	0.73	0.74	0.75
分位数自助法	0.99	0.99	0.97	0.96	0.94	0.93	0.92	0.91
倒转分位数自助法	0.33	0.40	0.50	0.58	0.67	0.74	0.77	0.81
对称分位数自助法	0.96	0.94	0.91	0.89	0.87	0.86	0.85	0.86
C_γ	0.90	0.88	0.87	0.86	0.84	0.85	0.84	0.85
C_γ^*	0.91	0.89	0.88	0.88	0.87	0.88	0.87	0.88

6. 对折点回归函数的推断

在这部分，我们考虑对折点回归函数

$$g(\theta) = \beta'x(\gamma) \quad (\text{附录 C-14})$$

的统计推断，其中

$$x(\gamma) = \begin{pmatrix} (x - \gamma)_- \\ (x - \gamma)_+ \\ z \end{pmatrix}.$$

我们的理论将 (x, z) 视为固定值，尽管我们将把估计视为 x 的函数。（因此，我们关注回归函数的逐点置信区间）。

$g(\theta)$ 的插值估计为 $g(\hat{\theta}) = \hat{\beta}'x(\hat{\gamma})$ 。在我们的实例中，我们在图C-2中展示了在 z 取固定的 z_t 的样本均值情况下作为 x 的函数的 $g(\hat{\theta})$ 。

因为定理C.2表明 $\hat{\theta}$ 是渐近正态的，所以可以推测 $g(\hat{\theta})$ 也是。这个猜想后来证明是错误的。尽管 $g(\theta)$ 是 θ 的连续函数，但它在 $\gamma = x$ 处是不可微的。正如最近一系列文章所讨论的（Hirano 与 Porter (2012); Woutersen 与 Ham (2013); Fang 与 Santos (2014); Fang (2014); Hong 与 Li (2015)），不可微性意味着 $g(\hat{\theta})$ 在 $\gamma_0 = x$ 处不会渐近正态，且渐近正态可能是对于 γ_0 接近 x 情况下的糟糕近似。

尽管回归折点函数 $g(\theta)$ 在 $\gamma = x$ 处不可微，但它在所有点上都是方向可微的，这意味着左右导数都是有定义的。函数 $\phi(\theta) : \mathbb{R}^k \rightarrow \mathbb{R}$ 在方向 $h \in \mathbb{R}^k$ 上的方向导数为

$$\phi_\theta(h) = \lim_{\varepsilon \rightarrow 0} \frac{\phi(\theta + h\varepsilon) - \phi(\theta)}{\varepsilon} \quad (\text{附录 C-15})$$

（可参见 Shapiro (1990)）。与传统的导数概念的主要区别是允许 $\phi_\theta(h)$ 依赖于求导的方向 h 。对于连续可微函数，其方向导数在方向 h 上是线性的。

例如，若 $\phi(\theta) = x'\theta$ 是线性的，那么 $\phi_\theta(h) = x'h$ 。若 $\phi(\theta)$ 是关于 θ 连续可微的那

表 C-4 系数估计与自助法的 90% 置信区间

	估计	标准差	置信区间
β_1	0.033	0.035	$[-0.002, 0.136]$
β_2	-0.067	0.037	$[-0.178, -0.006]$
y_{t-1}	0.28	0.088	$[0.14, 0.43]$
截距	3.78	0.79	$[2.58, 4.94]$
γ	43.8	12.3	$[30.8, 70.0]$

么 $\phi_\theta(h) = (\frac{\partial}{\partial \theta} \phi(\theta))'h$ 。然而，若 $\phi(\theta)$ 是连续但不是连续可微的，方向导数将在 h 方向上非线性。

在(附录 C-14)式的案例下，我们可以计算 $g(\theta)$ 在方向 $h = (h_\beta, h_\gamma)$ 上的方向导数为

$$g_\theta(h) = x(\gamma)'h_\beta + g(\gamma)(h_\gamma) \quad (\text{附录 C-16})$$

其中

$$g_\gamma(h_\gamma) = \begin{cases} -\beta_1 h_\gamma, & x < \gamma \\ -\beta_1 h_\gamma I(h_\gamma < 0) - \beta_2 h_\gamma I(h_\gamma \geq 0), & x = \gamma \\ -\beta_2 h_\gamma, & x > \gamma \end{cases} \quad (\text{附录 C-17})$$

方向导数 $g_\gamma(h_\gamma)$ 在 $x \neq \gamma$ 处线性但在 $x = \gamma$ 处非线性，在左侧具有斜率 $-\beta_1$ 而在右侧具有斜率 $-\beta_2$ 。

函数 $\phi(\hat{\theta})$ 的传统渐近定理需要函数 $\phi(\theta)$ 连续可微。Shapiro (1991, 定理 2.1), Fang, Santos (2014, 推论 2.1) 已经将这总结为方向可微性的案例。他们已经建立当 $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} Z \sim \mathbf{N}(0, V)$ 且 $\phi_\theta(h)$ 在方向 h 上连续则有 $\sqrt{n}(\phi(\hat{\theta}) - \phi(\theta_0)) \xrightarrow{d} \phi_\theta(Z)$ 的相关理论。在经典案例中当 $\phi(\theta)$ 是连续可微的，极限分布变为 $\phi_\theta(Z) = (\frac{\partial}{\partial \theta} \phi(\theta))'Z \sim \mathbf{N}(0, (\frac{\partial}{\partial \theta} \phi(\theta))'V(\frac{\partial}{\partial \theta} \phi(\theta)))$ ，故他们的结果是 Delta 方法的严格推广。

因为(附录 C-16)式与(附录 C-17)式在 h 方向上连续，我们可以立马推断回归估计的渐近分布。

定理 C.3 当 $Z \sim \mathbf{N}(0, V)$ 且 $g_\theta(h)$ 为(附录 C-16)式与(附录 C-17)式中的定义形式时， $\sqrt{n}(g(\hat{\theta}) - g(\theta_0)) \xrightarrow{d} g_\theta(Z)$ 。□

对于 $x \neq \gamma_0$ 的情况定理C.3中的渐近分布为正态分布，但对于 $x = \gamma_0$ 的情况渐近分布为正态随机向量的非线性变换。在 $x = \gamma_0$ 处渐近分布是有偏的，偏的方向取决于 β_1 与 β_2 的相对大小。例如，若 $\beta_1 = 1$ 且 $\beta_2 = -1$ ，则 $g_\gamma(h_\gamma) = |h_\gamma|$ 时尽管系数的符号相反但是定理C.3中的渐近分布将有正的均值，而 $g_\gamma(h_\gamma) = -|h_\gamma|$ 时定理C.3中的渐近分布将存在负的均值。

这种非正态性意味着经典置信区间将具有不正确的渐近覆盖。我们在图C-5中通过

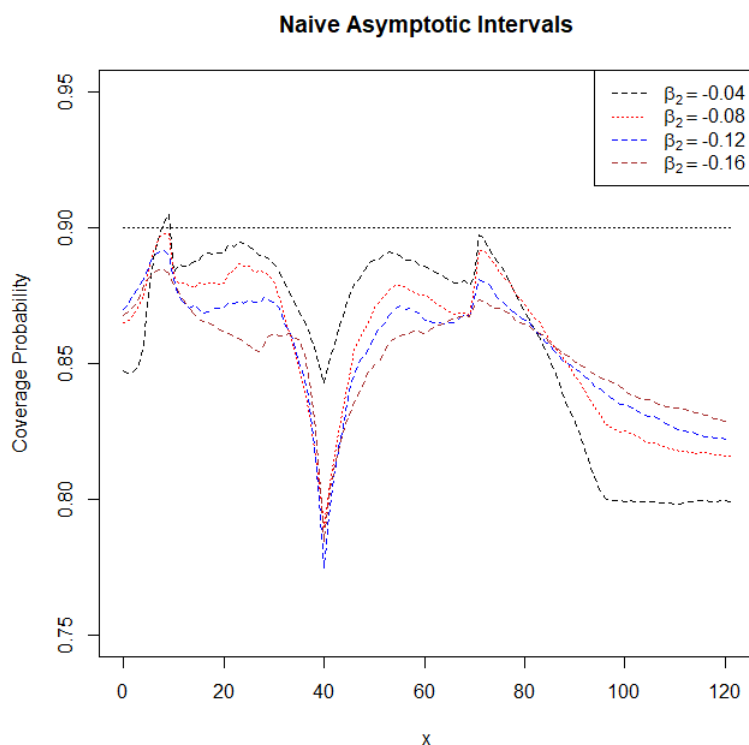


图 C-5 回归折点函数的名义上的 90% 原始渐近置信区间的覆盖概率

继续进行模拟试验来说明这一点。使用与前一节相同的模拟试验设计，以 0.04 为步长将 β_2 从 -0.04 变化到 0.16 ，我们绘制经典（朴素）90% 的逐点的置信区间的覆盖概率，将覆盖率作为 x 的函数进行绘制。渐近理论预测置信区间将对 $x \neq \gamma_0$ 具有渐近正确的 90% 的覆盖率，但对 $x = \gamma_0$ 没有，所以我们应该期望对 x 远离 $\gamma_0 = 40$ 的覆盖率更好，但 x 在 $\gamma_0 = 40$ 附近覆盖率会恶化。事实上，我们看到覆盖率从较小的 x 处的接近 90% 到 $x = 40$ 时的大约 77% 不等。我们还可以看到渐近理论没有预测出来的，对较大的 x 的值，覆盖率低于名义上的 90%，对于取较小值的 β_2 这些扭曲现象更严重。这似乎是一个小样本导致的问题（一个有限样本回归估计的偏差；当 $x_t \geq 80$ 时只有 7 个观测值），因此与定理 C.3 的渐近非正态性无关。

我们可能希望自助方法可以改善覆盖概率，但这是错误。如 Fang 和 Santos (2014, 推论 3.1) 所示，定理 C.3 的非正态性意味着传统的自助方法将是不一致的。事实上，我们研究了使用百分位数和逆百分位数方法构造的置信区间的覆盖概率，它们的覆盖率与图 C-5 所示的相似（尽管对于较大的 x 失真度较小），所以我们不在此显示。

Fang 和 Santos (2014) 建设性地提出了一种可选择性的与渐近分布相一致的自助方法。他们的方法是通过 $\hat{g}_\theta(\sqrt{n}(\hat{\theta}^* - \hat{\theta}))$ 估计 $\sqrt{n}(g(\hat{\theta}) - g(\theta_0))$ 的分布，其中 $\hat{\theta}^*$ 是 $\hat{\theta}$ 的自助分布且 $\hat{g}_\theta(h)$ 是 $g_\theta(h)$ 的估计。置信区间可以通过这种备选的自助分布构建出来。

Fang-Santos 备选自助算法需要方向导数 $g_\theta(h)$ 是已知的或是一致估计。幸运的是，方向导数的估计问题已经在 Hong 和 Li (2015) 的一篇近期的文章中得到解决。他们提议用离散模拟估计(附录 C-15)式。特别地，对于一些满足 $\varepsilon_n \rightarrow 0$ 且 $\sqrt{n}\varepsilon_n \rightarrow \infty$ 的大于

0 的序列 ε_n ，用

$$\hat{\phi}_\theta(h) = \frac{\phi(\hat{\theta} + h\varepsilon_n) - \phi(\hat{\theta})}{\varepsilon_n}$$

估计 $\phi(\theta)$ 的方向导数 $\phi_\theta(h)$ 。

Hong 与 Li (2015) 证明了当 $\phi(\theta)$ 是 Lipschitz 连续时会产生 $\phi_\theta(h)$ 的一致估计。他们称 $\hat{\phi}_\theta(h)$ 为“数字化的 Delta 方法”。

在我们的案例中，回归函数 $g(\theta)$ 关于 θ 这一参数 Lipschitz 连续且方向导数 $g_\theta(h)$ 具有(附录 C-16)式的半线性形式，所以我们只需要估计(附录 C-17)式的非线性部分。Hong-Li 估计为

$$\hat{g}_\gamma(h_\gamma) = \frac{g(\hat{\beta}, \hat{\gamma} + h_\gamma\varepsilon_n) - g(\hat{\beta}, \hat{\gamma})}{\varepsilon_n}$$

其中我们已经使 $g(\hat{\beta}, \hat{\gamma}) = g(\hat{\theta})$ 。因此，我们的方向全导数的估计为

$$\hat{g}_\theta(h) = x(\hat{\gamma})'h_\beta + \frac{g(\hat{\beta}, \hat{\gamma} + h_\gamma\varepsilon_n) - g(\hat{\beta}, \hat{\gamma})}{\varepsilon_n}.$$

在 $h = (h_\beta, h_\gamma) = (\sqrt{n}(\hat{\beta}^* - \hat{\beta}), \sqrt{n}(\hat{\gamma}^* - \hat{\gamma}))$ 处进行分析， $g(\hat{\theta}) - g(\theta_0)$ 的分布的自助估计为

$$r^* = x(\hat{\gamma})'(\hat{\beta}^* - \hat{\beta}) + \frac{g(\hat{\beta}, \hat{\gamma} + \sqrt{n}\varepsilon_n(\hat{\gamma}^* - \hat{\gamma})) - g(\hat{\beta}, \hat{\gamma})}{\sqrt{n}\varepsilon_n}.$$

Hong 和 Li (2015) 证明如果函数 $\phi(\theta)$ 是一个 θ 的凸函数，则用 r^* 作为控制参数以构造 θ 的上单边置信区间，但下单边置信区间则无法构建出来，且当 $\phi(\theta)$ 是一个凹函数时则逆命题成立。在我们的例子中，如果 $\beta_1 \leq \beta_2$ ，则函数 $g(\theta)$ 是 γ 的凸函数，但当 $\beta_1 \geq \beta_2$ 时，函数 $g(\theta)$ 是凹函数。因此，没有一个单边置信区间将一致控制大小。他们的建议是使用双侧对称置信区间，因为它们是渐近保守的。

设 $q_{1-\alpha}^*$ 表示变量 $|r^*|$ 的 $1 - \alpha$ 分位数。 $g(\theta_0)$ 的对称自助置信区间为 $g(\hat{\theta}) \pm q_{1-\alpha}^*$ 。我们将其称之为对称 Delta 方法的自助区间。

一个重要的问题是设置数值导数的增量 ε_n 。虽然 Hong 和 Li (2015) 的一般理论要求 $\sqrt{n}\varepsilon_n \rightarrow \infty$ ，但在该模型背景下他们推荐取 $\varepsilon_n = cn^{-1/2}$ ，虽然他们没有提供如何选择 c 的指导。我们遵循他们的建议，设置 $\varepsilon_n = cn^{-1/2}$ ，并通过仿真模拟评估常数 c 的选择。我们实现了数值 Delta 方法的自助算法，并在与图C-5相同的模拟中计算了 $c = 0.5$ ， $c = 1$ ， $c = 2$ 与 $c = 4$ 时的名义上的 90% 置信区间的逐点覆盖概率。覆盖概率绘制在图C-6中，其中每个面板对应一个 β_2 的值。

图C-6说明了覆盖概率关于 c 单调增加，因此一个更保守的（或更大的）置信区间可以通过增加 c 的方式找到。而关于回归参数 x ，覆盖概率也是不均匀的，当 x 小时会

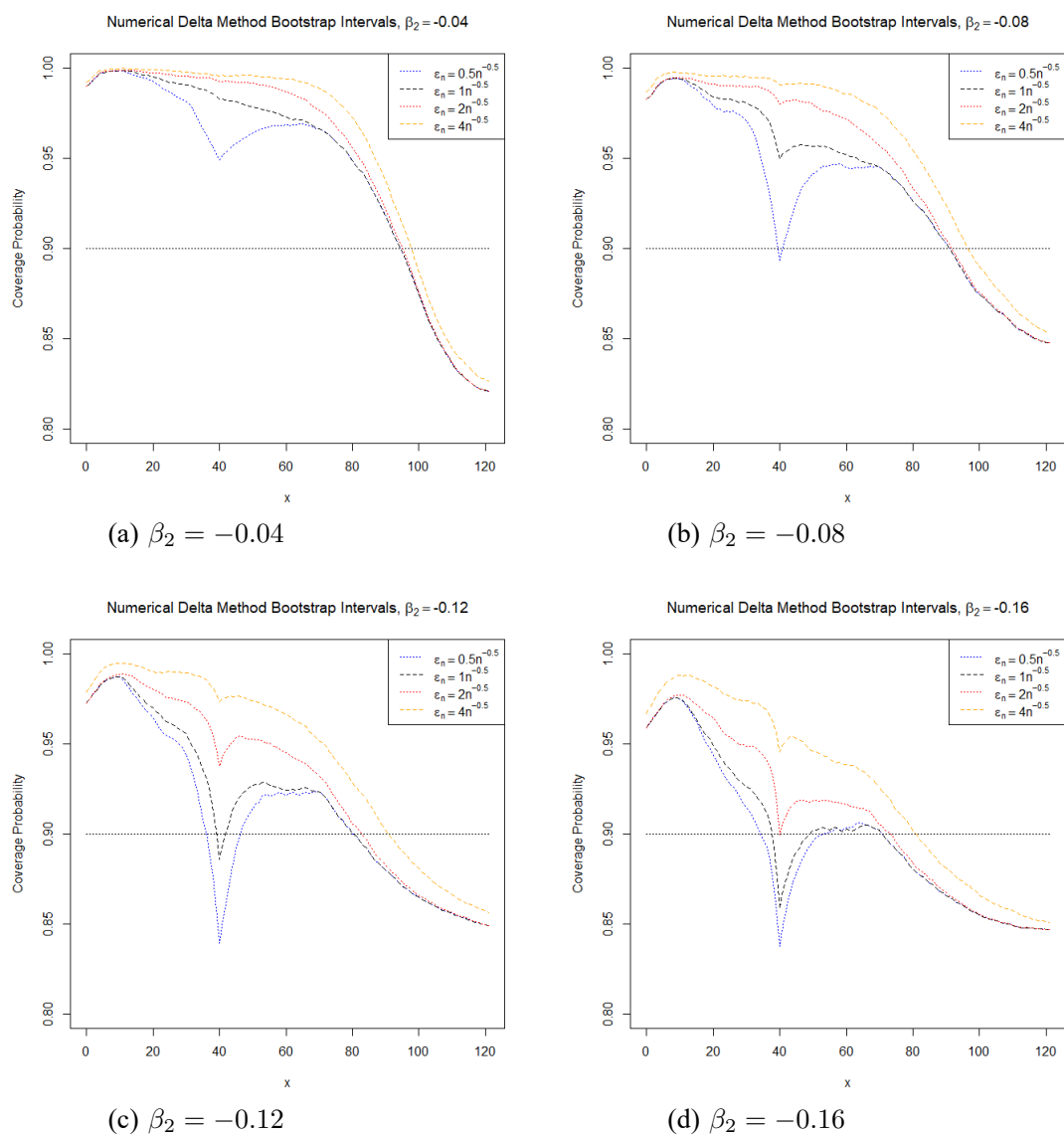


图 C-6 名义上的 90% 数字 Delta 方法构造的自助区间的覆盖概率

过度覆盖，而 x 较大时则欠覆盖。在 $x = 40$ 时，覆盖概率也出现一个下降高峰，这是真实门限的位置且符合非正态分布理论。这个向下的尖峰随着 c 的增大而减小，但也随着 β_2 的增大而增大。模拟试验表明 c 的正确选择依赖于 β_2 ，因此这使得对 c 的一般选择变得困难。在我们的应用中，我们对 $\beta_2 - \beta_1$ 的点估计是（这是合适的模拟） -0.10 ，这大致介于 (b) 和 (c) 面板之间。如果我们专注于在 $x = 40$ 时获得正确的覆盖范围，结果表明，取 $c = 1$ 是合理的，在 $\beta_2 = -0.08$ 时会过度覆盖而 $\beta_2 = -0.12$ 时会轻度欠覆盖。因此我们令 $c = 1$ ，因而 $\varepsilon_n = n^{-1/2}$ 。对于其他应用，我们建议通过模拟试验评估覆盖。

数值 Delta 法自助算法的理论论据是它对所有的 x 值都有渐近保守的覆盖，即使是在 $x = \gamma_0$ 处。然而，对于距离 γ_0 较远的 x 值，经典的自助法可能会有更好的覆盖（此时不那么保守）。然而，没有明确的混合使用自助方法的算法，并且在实际情况中 γ_0 的

值是未知的，所以我们建议对所有 x 值的置信区间使用数值 Delta 方法的自助算法。

现在我们简要概括数值 Delta 方法的自助算法的步骤。

- 1: 遵照算法C-2的粗糙自助方法的步骤 1-3。
- 2: 设定 $x(\hat{\gamma}) = ((x - \hat{\gamma})_-, (x - \hat{\gamma})_+, z')'$ 。
- 3: 设定 $\varepsilon_n = cn^{-1/2}$ 。
- 4: 设定 $r^* = x(\hat{\gamma})'(\hat{\beta}^* - \hat{\beta}) + \frac{g(\hat{\beta}, \hat{\gamma} + \sqrt{n}\varepsilon_n(\hat{\gamma}^* - \hat{\gamma})) - g(\hat{\beta}, \hat{\gamma})}{\sqrt{n}\varepsilon_n}$ 。
- 5: 重复 B 次，因而获得模拟估计量 r^* 的一组样本。
- 6: 计算 $|r^*|$ 的 $1 - \alpha$ 分位数 $q_{1-\alpha}^*$ 。
- 7: 设定 g_{θ_0} 的 $1 - \alpha$ 置信区间为 $[g(\hat{\theta}) - q_{1-\alpha}^*, g(\hat{\theta}) + q_{1-\alpha}^*]$ 。

算法 C-3 在一 (x, z) 的固定值处的折点回归函数的数值 Delta 方法的自助算法的置信区间

这在数值上实现起来非常简单，尽管第 4 步对于自助法过程来说并不常见。为了得到置信区间带，步骤 2 到 7 需要对每个考虑到的 (x, z) 值进行重复。

理论上，对于序列相关的数据，Hong-Li 数值 Delta 自助方法与这种特定的粗糙自助方法的合并还没有被研究。我们欢迎更多的可以有助于理解它的特性的工作。

对于 $c = 0.5$, $c = 1$, $c = 2$ 与 $c = 4$ ，我们应用使用 $\varepsilon_n = cn^{-1/2}$ 的数值 Delta 方法的自助算法于美国的 GDP 增长回归上，并且在图C-7中绘制出了逐点的 90% 置信区间。在数值上，这意味着算法C-3的应用是将 z 设置在 z_t 的样本均值处，并在网格上对 x 从 1 到 120 进行评估。除 $x = \hat{\gamma}$ 外，置信区间随着 ε_n 的增加而变宽。

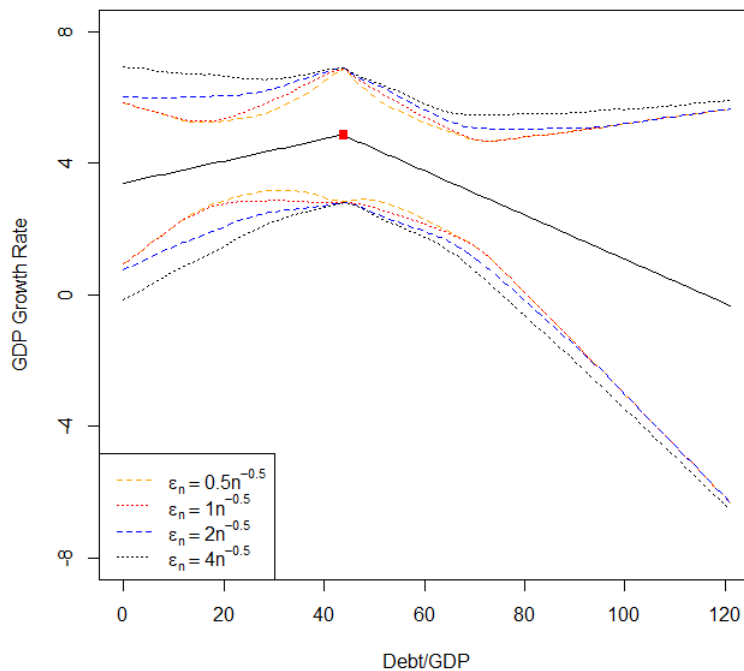


图 C-7 估计的回归折点函数与 90% 数字化 Delta 方法构造的置信区间

重要的是将这些置信区间解释为关于 x 的逐点形式（不均匀）。最近 Belloni 等人（文章在印刷中）对于发展非参回归的均匀推断方法有一些进展，但是这些方法不适用于数值 Delta 方法的自助算法。

根据我们模拟试验的建议，我们将第二个估计量 ($\epsilon_n = n^{-1/2}$) 作为我们的首选，这在图C-2中被用蓝色虚线绘制出来。置信区间足够宽，以至于不清楚真正的回归函数是平坦的还是向下倾斜的。置信区间显示，仅使用美国长期数据，折点回归模型的估计不够精确，不足以得出关于债务水平对 GDP 增长率是否有负面影响的强有力结论。

7. 结论

本文对门限未知的折点回归模型提出了估计与推断理论，并将其应用于研究 Reinhart 与 Rogoff (2010) 的 GDP 增长和债务门限问题。一个有趣的理论贡献是发现回归函数的估计由于其不可微性从而是非正态的，且利用 Fang 与 Santos (2014) 和 Hong 与 Li (2015) 最新推断方法可以形成置信区间。

我们将该方法应用于 Reinhart 和 Rogoff (2010) 搜集的关于美国的长跨度时间序列数据。我们对点的估计与 Reinhart-Rogoff 的假设一致，即当债务水平超过一个门限时，经济增长将放缓。然而，门限效应存在的正式证据是不确定的，且我们的回归函数的置信区间足够宽，以至于债务对经济增长的影响是难以被侦测出来。

一个重要的警告是，我们的经验结果只基于一个单一的时间序列（以美国这一国家为例），因此忽略了其他国家的经验信息。这样做的好处是不强加同质性，但也因此减少了准确性与权威性。将这里的结果拓展到面板分析将是比较有用的。

附录：数学证明

证明 定理C.2的证明：通过定义(附录 C-4)， $\hat{\theta}$ 最小化(附录 C-3)中定义的 $S_n(\theta)$ ，我们可以将 (3) 式写成 $S_n(\theta) = \frac{1}{n} \sum_{t=1}^n e_t(\theta)^2$ 而 $e_t(\theta) = y_t - \beta'x_t(\gamma)$ 。因此 $\hat{\theta}$ 满足一阶条件 $\frac{1}{n} \sum_{t=1}^n m_t(\hat{\theta}) = 0$ 其中 $m_t(\theta) = H_t(\theta)e_t(\theta)$ 。伪真值 θ_0 最小化 $L(\theta) = L(\beta, \gamma)$ 因此求解 $m(\theta_0)$ 其中 $m(\theta) = \frac{\partial}{\partial \theta} L(\theta) = E m_t(\theta)$ 。定义 $Q(\theta) = -\frac{\partial}{\partial \theta'} m(\theta)$ 。

Andrews (1994)，章节 3.2，证明定理C.2应满足以下条件：

假设 .3 对于参数 θ 有

1. $\hat{\theta} \xrightarrow{p} \theta_0$ 。
2. $\frac{1}{\sqrt{n}} \sum_{t=1}^n H_t e_t \xrightarrow{d} \mathcal{N}(0, S)$ 。
3. $Q(\theta)$ 关于 θ 连续且 $Q(\theta_0) = Q$ 。
4. $v_n(\theta) = \frac{1}{\sqrt{n}} \sum_{t=1}^n (m_t(\theta) - m(\theta))$ 是随机等度连续的。□

我们首先建立条件 1。因为 $x_t(\gamma)$ 关于 γ 连续， $e_t(\theta)$ 与 $e_t(\theta)^2$ 关于 θ 连续。回顾定理C.2我们有简单的上界 $\|x_t(\gamma)\|^2 = z_t' z_t + (x_t - \gamma)^2 \leq \|z_t\|^2 + x_t^2 + \gamma^2$ 。则通过 C_r 与

Cauchy-Schwarz 不等式,

$$e_t(\theta)^2 \leq 2y_t^2 + 2|\beta'x_t(\gamma)|^2 \leq 2y_t^2 + 2\bar{\beta}^2(\|z_t\|^2 + x_t^2 + \bar{\gamma}^2), (A.1)$$

其中 $\bar{\beta} = \sup\{\|\beta\| : \beta \in B\}$ 且 $\bar{\gamma} = \sup\{|\gamma| : \gamma \in \Gamma\}$, 二者在假设C.2的第二条下有限。在假设C.1的第二条下 (A.1) 式中的界有有限的期望。接着通过 Newey 和 McFadden (1994) 的引理 2.4 (基于 Tauchen (1985) 的引理 1), $L(\theta) = Ee_t(\theta)^2$ 是连续的且随着 $n \rightarrow \infty$, $\sup_{\theta \in B \times \Gamma} |S_n(\theta) - L(\theta)| \xrightarrow{P} 0$ 。Newey 与 McFadden 的引理 2.4 是对于独立同分布观测量陈述的, 但是结果只需要应用一个弱大数律, 且该弱大数律在假设C.1下通过遍历性定理仍成立。

在给定集合 $B \times \Gamma$ 的紧致性与最小值 θ_0 的唯一性的条件下 (相对地, 通过假设C.2的第一条与第二条), Newey 与 McFadden (1994) 的定理 2.1 建立了条件随着 $n \rightarrow \infty$, $\hat{\theta} \xrightarrow{P} \theta_0$, 也就是条件 1。

条件 2 依据的是强混合过程的 Herrndorf (1984) 中心极限定理, 在假设C.2的第一条与第二条下成立。

下面的界将在建立条件 3 与条件 4 时非常有用。设 ω_t 为任一向量, 其元素为变量 (y_t, x_t, z_t) 的两两乘积, 且设定 $q = r/(r-1)$ 。注意到假定 1.2 意指对于 $C \leq \infty$ 有 $(E\|\omega_t\|^{2r})^{1/r} \leq C$ 成立。设 $F(\gamma)$ 表示 x_t 的分布函数, 在假设C.2的第四条下满足 $F(\gamma_2) - F(\gamma_1) \leq \bar{f}|\gamma_2 - \gamma_1|$ 。通过 Holder 不等式, 我们发现

$$E\|\omega_t I(\gamma_1 \leq x_t \leq \gamma_2)\|^2 \leq (E\|\omega_t\|^{2r})^{1/r} (E|I(\gamma_1 \leq x_t \leq \gamma_2)|^q)^{1/q} \quad (\text{附录-18})$$

$$\leq C(F(\gamma_2) - F(\gamma_1))^{1/q} \quad (\text{附录-19})$$

$$\leq C\bar{f}^{1/q}|\gamma_2 - \gamma_1|^{1/q}. \quad (A.2) \quad (\text{附录-20})$$

这边界条件暗指 $E(\omega_t I(x_t \leq \gamma))$ 关于 γ 一致连续。

现在我们建立条件 3。注意到

$$Q(\theta) = E(H_t(\theta)H_t(\theta)') + E \begin{pmatrix} 0 & 0 & 0 & e_t I(x_t < \gamma_0) \\ 0 & 0 & 0 & e_t I(x_t > \gamma_0) \\ 0 & 0 & 0 & 0 \\ e_t I(x_t < \gamma_0) & e_t I(x_t > \gamma_0) & 0 & 0 \end{pmatrix}$$

这个矩阵的元素是 β 的二次函数与通过形式 $E(\omega_t I(x_t \leq \gamma))$ 形式的矩建立的 γ 的函数, 其中 ω_t 是变量 (t_t, x_t, z_t) 的内积。因为 (A.2) 式成立我们可以看到 $Q(\theta)$ 是关于 θ 连续的。在 θ_0 处分析我们发现 $Q(\theta_0) = Q$ 。因此条件 3 成立。

我们现在通过定理C.1的应用 4 来建立条件 4, 应用为 Doukhan, Massart 与 Rio (1995) 的案例 (2.15)。注意到 $m_t(\theta)$ 是 β 的二次函数, 所以唯一的问题是建立关于 γ 的随机等度连续性。因此, 我们简化符号, 将 $m_t(\theta)$ 写成 $m_t(\gamma)$ 。注意到我们可以写

$m_t(\gamma) = \omega_t I(x_t \leq \gamma)$ 其中 ω_t 是其元素为变量 (t_t, x_t, z_t) 的内积的向量。在假设C.2的第二条下 $m_t(\gamma)$ 具有有界的 2 阶矩，且因此满足所需要的包络条件。

对于任意的 $\delta > 0$ 设定 $N(\delta) = \delta^{-2/q}$ 且设定 $\gamma_k, k = 1, \dots, N$ 为集合 Γ 上的等距的网格。注意到格点间的距离为 $O(N(\delta)^{-1})$ 。定义 $m_{tk}^* = \min[m_t(\gamma_{k-1}), m_t(\gamma_k)]$ 且 $m_{tk}^{**} = \max[m_t(\gamma_{k-1}), m_t(\gamma_k)]$ 。那么对于每一个 $\gamma \in \Gamma$ 有一个 $k \in [1, \dots, N]$ 满足 $m_{tk}^* \leq m_t(\gamma) \leq m_{tk}^{**}$ 。因此 $[m_{tk}^*, m_{tk}^{**}]$ 包含了 $m_t(\gamma)$ 。通过他们的构造，得到事实 $m_t(\gamma) = \omega_t I(x_t \leq \gamma)$ ，且应用 (A.2) 式，

$$\begin{aligned} E||m_{tk}^{**} - m_{tk}^*||^2 &= E||m_t(\gamma_k) - m_t(\gamma_{k-1})||^2 \\ &\leq E||\omega_t I(\gamma_{k-1} \leq x_t \leq \gamma_k)||^2 \\ &\leq C \bar{f}^{1/q} |\gamma_k - \gamma_{k-1}|^{1/q} \\ &\leq O(N(\delta)^{-q}) = O(\delta^2) \end{aligned}$$

这意味着 $N(\delta) = \delta^{-2/q}$ 是 L^2 括号数且 $H_2(\delta) = \log N(\delta) = |\log \delta|$ 是支撑类 $\{m_t(\gamma) : \gamma \in \Gamma\}$ 的测度熵。这和假设C.2的第一条在建立了 $v_n(\theta)$ 的随机等度连续性因而建立条件 4 的情况下暗示了 Doukhan, Massart 与 Rio (1995) 的推论 (2.15) 且因此得到了他们的定理C.1。

我们已经在假设C.1与C.2下建立了条件 1-4。正如上面讨论的，这对于建立定理C.2是充分的。

致谢

作者感谢美国国家科学基金会的支持。作者感谢副主编、两位推荐人 Zheng Fang, Han Hong, Jessie Li, Tzu-Chi Lin 和 Andres Santos, 感谢他们提供的非常有帮助的评论、建议和通信，这些对本文的研究结果至关重要。

参考文献

- Andrews, D. W. K. (1994), “Empirical Process Methods in Econometrics,” in Handbook of Econometrics (Vol IV), eds. R. F. Engle and D. L. McFadden, New York: Elsevier, pp. 2247–2294. [239]
- Belloni, A., Chernozhukov, V., Chetverikov, D., and Kata, K. (in press), “On the Asymptotic Theory for Least Squares Series: Pointwise and Uniform Results,” Journal of Econometrics. [238]
- Caner, M., Grennes, T., and Koehler-Geib, F. (2010), “Finding the Tipping Point-When Sovereign Debt Turns Bad,” Policy Research Working Paper Series 5391, The World Bank. [228]
- Caner, M., and Hansen, B. E. (2004), “Instrumental Variable Estimation of a Threshold

- Model,” *Econometric Theory*, 20, 813–843. [228]
- Card, D., Lee, D., Pei, Z., and Weber, W. (2012), “Nonlinear Policy Rules and the Identification and Estimation of Causal Effects in a Generalized Regression Kind Design,” NBER Working Paper 18564. [228]
- Card, D., Mas, A., and Rothstein, R. (2008), “Tipping and the Dynamics of Segregation,” *Quarterly Journal of Economics*, 123, 177–218. [228]
- Cecchetti, S. G., Mohanty, M. S., and Zampolli, F. (2011), “The Real Effects of Debt,” in *Economic Symposium Conference Proceedings: Federal Reserve Bank of Kansas City*, pp. 145–196. [228]
- Chan, K.-S. (1990), “Testing for Threshold Autoregression,” *The Annals of Statistics*, 18, 1886–1894. [228]
- (1991), “Percentage Points of Likelihood Ratio Tests for Threshold Autoregression,” *Journal of the Royal Statistical Society, Series B*, 53, 691–696. [228]
- (1993), “Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model,” *The Annals of Statistics*, 21, 520–533. [228]
- Chan, K.-S., and Tong, H. (1990), “On Likelihood Ratio Tests for Threshold Autoregression,” *Journal of the Royal Statistical Society, Series B*, 52, 469–476. [228]
- Chan, K.-S., and Tsay, R. S. (1998), “Limiting Properties of the Least Squares Estimator of a Continuous Threshold Autoregressive Model,” *Biometrika*, 45, 413–426. [228,233]
- Chiu, G., Lockhart, R., and Routledge, R. (2006), “Bent-Cable Regression Theory and Applications,” *Journal of the American Statistical Association*, 101, 542–553. [228]
- Cox, D., Hansen, B. E., and Jimenez, E. (2004), “How Responsive are Private Transfers to Income?” *Journal of the Public Economics*, 88, 2193–2219. [228]
- Doukhan, P., Massart, P., and Rio, E. (1995), “Invariance Principles for Absolutely Regular Empirical Processes,” *Annales de l’Institut H. Poincare, Probabilites et Statistiques*, 31, 393–427. [239]
- Fang, Z. (2014), “Optimal Plug-in Estimators of Directionally Differentiable Functionals,” Working Paper, UCSD. [235]
- Fang, Z., and Santos, A. (2014), “Inference on Directionally Differentiable Functions,” Working Paper, UCSD. [228,235,236,238]
- Feder, P. I. (1975), “On Asymptotic Distribution Theory in Segmented Regression Problems —Identified Case,” *Annals of Statistics*, 3, 49–83. [228]
- Ganong, P., and Jager, S. (2014), “A Permutation Test and Estimation Alternatives for the Regression Kink Design,” Working Paper, Harvard University. [228]
- Hansen, B. E. (1996), “Inference When a Nuisance Parameter is not Identified Under the Null Hypothesis,” *Econometrica*, 64, 413–430. [228,231]

- (1999), “Threshold Effects in Non-Dynamic Panels: Estimation, Testing and Inference,” *Journal of Econometrics*, 93, 345–368. [228]
- (2000), “Sample Splitting and Threshold Estimation,” *Econometrica*, 68, 575–603. [228]
- Herrndorf, N. (1984), “A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables,” *The Annals of Probability*, 12, 141–153. [239]
- Hinkley, D. V. (1969), “Inference About the Intersection in Two-Phase Regression,” *Biometrika*, 56, 495–504. [228]
- (1971), “Inference in Two-Phase Regression,” *Journal of the American Statistical Association*, 66, 736–743. [228]
- Hirano, K., and Porter, J. (2012), “Impossibility Results for Nondifferentiable Functionals,” *Econometrica*, 80, 1769–1790. [235]
- Hong, H., and Li, J. (2015), “The Numerical Directional Delta Method,” Working Paper, Stanford University. [228,235,236,237,238]
- Landais, C. (2015), “Assessing the Welfare Effects of Unemployment Benefits Using the Regression Kink Design,” *American Economic Review: Economic Policy*, 7, 243–278. [228]
- Lee, S., Park, H., Seo, M. H., and Shin, Y. (2014), “A Contribution to the Reinhart and Rogoff Debate: Not 90 Percent but Maybe 30 Percent,” CEMMAP Working Paper. [228,231]
- Lee, S., Seo, M. H., and Shin, Y. (2011), “Testing for Threshold Effects in Regression Models,” *Journal of the American Statistical Association*, 106, 220–231. [228]
- Lin, T.-C. (2014), “High-Dimensional Threshold Quantile Regression With an Application to Debt Overhang and Economic Growth,” Working Paper, University of Wisconsin. [228,229,230]
- Newey, W. K., and McFadden, D. (1994), “Large Sample Estimation and Hypothesis Testing,” in *Handbook of Econometrics (Vol IV)*, eds. R. F. Engle and D. L. McFadden, New York: Elsevier, pp. 2113–2245. [239]
- Porter, J., and Yu, P. (2015), “Regression Discontinuity Designs With Unknown Discontinuity Points: Testing and Estimation,” *Journal of Econometrics*, 189, 132–147. [228]
- Ramirez-Rondan, N. (2013), “Maximum Likelihood Estimation of a Dynamic Panel Threshold Model,” Working Paper, Central Bank of Peru, <https://sites.google.com/site/teltoramirezrondan/>. [228]
- Reinhart, C. M., and Rogoff, K. S. (2010), “Growth in a Time of Debt,” *American Economic Review: Papers and Proceedings*, 100, 573–578. [228,229,238]
- Seo, M. H., and Linton, O. (2007), “A Smoothed Least Squares Estimator for Threshold Regression Models,” *Journal of Econometrics*, 141, 704–735. [228]

- Shapiro, A. (1990), “On Concepts of Directional Differentiability,” *Journal of Optimization Theory and Applications*, 66, 477-487. [235]
- (1991), “Asymptotic Analysis of Stochastic Programs,” *Annals of Operations Research*, 30, 169-186. [236]
- Stambaugh, R. F. (1999), “Predictive Regressions,” *Journal of Financial Economics*, 54, 375-421. [232]
- Tauchen, G. (1985), “Diagnostic Testing and Evaluation of Maximum Likelihood Models,” *Journal of Econometrics*, 30, 415-443. [239]
- Tong, H. (1983), *Threshold Models in Non-linear Time Series Analysis. Lecture Notes in Statistics (Vol. 21)*, Berlin: Springer. [228]
- (1990), *Non-Linear Time Series: A Dynamical System Approach*, Oxford: Oxford University Press. [228]
- Woutersen, T. and Ham, J. C. (2013), “Calculating Confidence Intervals for Continuous and Discontinuous Functions of Parameters,” Working Paper, University of Arizona. [235]