

分类号 _____
U D C _____

密级 _____
编号 10741

兰州财经大学

LANZHOU UNIVERSITY OF FINANCE AND ECONOMICS

硕士学位论文

论文题目 分层非参数混合效应模型的分位回归

研究生姓名: 曹苏周

指导教师姓名、职称: 田茂再 教授

学科、专业名称: 统计学 数理统计学

研究方向: 复杂数据分析

提交日期: 2022年5月30日

独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名： 曹安周 签字日期： 2022.5.30

导师签名： 田茂再 签字日期： 2022.5.30

关于论文使用授权的说明

本人完全了解学校关于保留、使用学位论文的各项规定，同意（选择“同意”/“不同意”）以下事项：

1.学校有权保留本论文的复印件和磁盘，允许论文被查阅和借阅，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文；

2.学校有权将本人的学位论文提交至清华大学“中国学术期刊（光盘版）电子杂志社”用于出版和编入CNKI《中国知识资源总库》或其他同类数据库，传播本学位论文的全部或部分内容。

学位论文作者签名： 曹安周 签字日期： 2022.5.30

导师签名： 田茂再 签字日期： 2022.5.30

Quantile Regression for Hierarchical Nonparametria with Mixed Effects Model

Candidate : Suzhou Cao

Supervisor: Maozai Tian

摘 要

具有分层结构的数据在生活中是十分常见的,这类数据使用范围广,常用于增长性研究、机构效应和综合研究。为了处理分层数据提出了分层线性模型,该模型有两个基本假定:①组间误差项独立同分布且均值为零、方差有界;②各层模型均为线性。然而,在实际研究中,数据经常会存在异方差或者重尾尖峰的情况,协变量和响应变量之间也经常存在非线性的关系,所以,上面假定不能满足实际对数据分析的需要。另外,分层线性模型在处理数据时采用的是均值回归,它在给定协变量时只能体现响应变量的平均变化情况,不能刻画出响应变量的整体条件分布,那么能够很好地刻画响应变量整体条件分布的方法就是分位数回归。为了更好地解释协变量与响应变量之间可能存在的非线性关系,非参数分位数回归模型的结合为解决这一问题提供了很好的思路,因此将非参数回归理论和分位回归理论结合到分层模型将会很好的处理这类问题。

本文首先将非参数回归理论和分位回归理论结合到分层模型中,同时加入混合效应,建立分层非参数混合效应模型,并使用基于检验函数的非参数分位数回归对其参数进行估计,在估计的过程中遇到了核函数和窗宽选择的挑战。研究表明,当数据样本容量足够大,不论选取哪种核函数,都可以在一定的正则条件下,保证估计量具有相合性这一最基本的要求,所以本文选择了高斯核函数,并通过最优渐近理论给出了最优窗宽与均值窗宽的关系。同时由于估计结果很难给出解析解,所以将分位数回归与 EM 算法结合形成 EQ 算法,并进行迭代达到了估计的目的。其次对于参数估计量的渐近性质进行了理论推导。接着通过 Monte Carlo 模拟,比较了随机误差项服从不同分布下的模型参数估计结果,可以发现,在样本量增大的情况下,本文所提出的模型的估计量都基本收敛到一个固定值,说明所用的估计方法在该模型下具有稳健性。最后通过实际数据说明了模型和方法的有效性和实用性。

关键词: 分层模型 非参数分位回归 混合效应 窗宽选择 EQ 算法

Abstract

Data with hierarchical structure is very common in daily life. This kind of data is widely used in growth studies, institutional effects and synthesis studies. In order to deal with hierarchical data, a hierarchical linear model is proposed, which has two basic assumptions: (1) the error terms between groups are independent and identically distributed with zero mean and bounded variance; (2) The models of each layer are linear. However, in actual studies, data often have heteroscedasticity or heavy tail spikes, and there is often a non-linear relationship between covariates and response variables. Therefore, the above assumption cannot meet the actual needs of data analysis. In addition, hierarchical linear model adopts mean regression in data processing, which can only reflect the average change of response variables when given covariables, but cannot depict the overall conditional distribution of response variables. Therefore, quantile regression is a good method to depict the overall conditional distribution of response variables. In order to better explain the possible nonlinear relationship between the explained variables and explanatory variables, the combination of non-parametric quantile regression model provides a good idea to solve this problem. Therefore, the combination of non-parametric regression theory and quantile regression theory into hierarchical model will be a good solution to this problem.

In this thesis, nonparametric regression theory and quantile

regression theory are combined into the layered model, and the mixed effect is added to establish the layered nonparametric mixed effect model, and the nonparametric quantile regression based on the test function is used to estimate its parameters. In the process of estimation, the challenges of kernel function and window width selection are encountered. The research shows that when the data sample size is large enough, no matter which kernel function is selected, the consistency of the estimator can be guaranteed under certain regular conditions, so the Gaussian kernel function is selected in this thesis, and the relationship between the optimal window width and the mean window width is given by the optimal asymptotic theory. At the same time, since it is difficult to give analytical solutions to the estimation results, quantile regression is combined with EM algorithm to form EQ two steps, and the purpose of estimation is achieved through iteration. Secondly, the asymptotic properties of parameter estimators are deduced theoretically. Then, the estimation results of model parameters with different distributions of error terms are compared by Monte Carlo simulation. It can be found that the estimators proposed in this thesis converge to a true value when the sample size increases, which proves the robustness of the estimation method. Finally, the effectiveness and practicability of the method are illustrated by practical data.

Keywords: Hierarchical model; Nonparametric quantile regression;
Mixed effects; Bandwidth selection; EQ algorithm

目录

1 引言	1
1.1 研究背景及意义.....	1
1.2 国内外研究现状.....	4
1.3 文献述评.....	6
1.4 本文研究内容.....	6
1.5 本文的创新之处.....	9
2 理论基础	10
2.1 分层线性分位回归模型.....	10
2.2 非参数回归理论.....	11
2.2.1 非参数回归模型.....	11
2.2.2 核函数.....	11
2.2.3 窗宽选择.....	12
2.3 分位回归理论.....	14
3 分层非参数混合效应模型	16
3.1 模型设定.....	16
3.2 参数估计.....	17
3.2.1 Q步.....	18
3.2.2 E步.....	19
3.2.3 迭代.....	20
3.3 窗宽选择.....	21
3.4 参数估计的渐近性质.....	23
4 数值模拟和实例分析	28
4.1 Monte Carlo 模拟.....	28
4.2 实例分析.....	32
4.2.1 汽车品牌数据.....	32
4.2.2 数据分析.....	33

5 总结与展望	36
5.1 结论.....	36
5.2 展望.....	36
参考文献	39
后记	40

1 引言

1.1 研究背景及意义

大数据时代给人们的生活带来了极大的便利，与此同时，也产生了各种各样的复杂数据。这种复杂数据几乎存在于所有的科学研究领域，例如经济金融、天文学、流行病学以及生物学等。数据复杂意味着它有高维、高频、多元、分层等特征，根据这些特征的统计建模覆盖了当代统计学的很多重要分支，通过建模可以清晰明了的显示数据类型和特征，以便我们在日常生活中可以更快速方便的储存、获取和计算，但是，在建模过程中也遇到了不少的挑战，其中包括在高维空间中直接进行系统收缩变得非常困难、一般高维数据的精确逼近非常棘手、高维函数积分目前不可能实现、对高维多元条件随机变量分布的全面刻画尚无例可循、如果没有考虑普遍存在的复杂分层数据特征，传统的统计方法往往无法满足实际数据处理的需求。

分层数据是复杂数据的一种类型，这类数据存在于特定的时间和空间中，是日常生活中最普遍的数据类型。比如学校在制定政策以便提高学生的成绩时，显然学生和学校都是需要分析的对象，对这两个层次的变化都必须进行考虑，这种处理就是针对学生嵌套在学校里的分层数据的分析；或者研究市里不同学校的升学率、一个省不同地级市的经济发展状况都是分层结构：学校嵌套在市里，地级市嵌套在省里，甚至还有更复杂的嵌套模式，这些情况都可以根据不同学校、不同市进行分层。针对分层数据的分析，一般要结合时间和空间两个方面，然而还有一种情况，就是离开时空性这一特征，考虑同一层的个体内部不同观测值之间的相依性以及个体之间的相依关系也是分层结构的考量。Lindly 和 Smith (1972) 及 Smith (1973) 首次提出了分层线性模型的概念。在早期的应用中，人们关注的是响应变量是如何受不同协变量影响的，并且通过线性均值回归模型来刻画它们之间的关系。均值回归模型体现的是协变量对响应变量的平均的影响，进行参数估计时最常用的估计方法是最小二乘估计方法，在回归模型随机误差项服从高斯分布的假设下，该方法得到的估计量是最小方差无偏估计，这体现了回归模型和最小二乘法的优越性。然而，在现实数据中，均值回归模型有以下几点不足。

第一，误差项的经典假设和高斯分布这些条件基本无法满足，样本数据常常会有厚尾尖峰等特点，这种情况下，模型中响应变量的条件均值很容易受到离群点的影响，由最小二乘估计方法得到的估计量基本都是有偏的；第二，均值回归无法解释给定协变量时响应变量的整体条件分布，尤其是分布两端且含有异常值时的分布状况。然而学者在研究现实问题时，更关注协变量对响应变量全面的影响，比如在研究生育率时，生育率的影响因素有年龄、学历、收入、地区等，这些因素对高生育率和低生育率的影响可能不同，学者更关心两个群体之间的差距，如果使用均值回归就将二者平均化，无法体现高低生育率之间的差距；第三，均值回归模型不能准确刻画协变量如何影响响应变量条件分布的峰度和偏度等形态。为了解决均值回归的不足，Koenker 和 Bassett（1978）提出了分位数回归理论，分位数回归的思想是基于最小化绝对残差来估计所有的条件分位数函数，从而可以更全面地描述协变量对响应变量在不同分位数水平上的影响。基于此，分位数回归正好弥补了上述均值回归模型中存在的不足。第一，分位数回归全面地刻画了协变量对响应变量整体的影响，尤其能够解释响应变量在高低分位数水平上的信息；第二，在均值回归中通常需要假定随机误差项服从某个已知分布，但是在分位数回归模型中没有假设分布的前提；第三，对均值回归模型的参数估计是通过最小化残差平方和来得到的，而分位数回归采用的是对估计值和真实值的差进行非对称形式的加权绝对值求和，权重是 τ 和 $1-\tau$ ，这种方法的好处是不易受数据中异常值的影响，在评价过程中更稳健。将估计方法改进之后的分层线性分位数回归模型中，假定模型的每层都是线性的。通过该模型进行建模时，为了更好地处理纵向和重复数据，人们通常会考虑混合线性模型。混合线性模型的优点是包含了固定效应和随机效应，充分考虑了数据聚集性的问题。但是，模型中的线性假设面对真实数据时是很苛刻的，甚至很多时候协变量与响应变量之间可能是非线性的关系，同时非线性回归模型的具体表达式也很难给出。考虑到这个问题，学者提出了非参数分位数回归模型。该模型一方面不设定具体的回归形式，另一方面也很少限制对变量的分布，具有很强的适应性和灵活性。Samanta（1989）首次结合了分位数回归与非参数理论，利用非参数中的加权核估计对条件分位数进行了估计。至今，对于非参数分位回归模型的理论研究还是很新颖的话题。分位数回归和混合效应模型有很多的优点，但在具体的参数估计过程中也遇到了困

难, 即很难求出解析解。在此基础上, Dempster (1977) 提出了 EM 算法, 该算法是针对数据有缺失情况下参数估计的迭代算法, 混合效应模型中的随机效应可以看作缺失数据, 所以用 EM 算法迭代进行估计有效的解决了这一难题。基于上述研究背景, 本文考虑分层非参数混合效应模型, 并进行分位数回归。此研究的理论意义和现实意义如下:

(1) 理论意义: 很多学者对分层数据做了颇为丰富的研究。这种数据使用范围广, 常用于增长性研究、机构效应和综合研究。这种分层线性模型有两个基本假定: ①组间误差项独立同分布且均值为零、方差有界; ②各层模型均为线性。然而, 在实际研究中, 数据经常会存在异方差或者重尾尖峰的情况, 协变量和响应变量之间也经常存在非线性的关系, 所以, 上面假定是不能满足实际对数据分析的需要。另外, 分层线性模型在处理数据时采用的是均值回归, 它在给定协变量时只能体现响应变量的平均变化情况, 不能刻画出响应变量的整体条件分布, 那么能够很好地刻画响应变量整体条件分布的方法就是分位数回归。与均值回归不同的是, 分位数回归采用最小绝对偏差来估计参数, 即如果模型的随机误差 $(\varepsilon_{ij})_{\{i=1, \dots, n; j=1, \dots, n_i\}}$ 满足 $P(\varepsilon_{ij} < 0) = \tau$, 则 τ 分位数回归的参数估计是通过最小化函数 $\rho_{\tau}(\varepsilon) = \sum_{i=1}^n \sum_{j=1}^{n_i} \varepsilon_{ij} \left[\tau I_{\varepsilon_{ij} > 0} - (1-\tau) I_{\varepsilon_{ij} \leq 0} \right]$ 来获得的。因此, 在实际问题的研究中, 需要结合非参数理论和分位回归理论, 方便更好地处理现实问题。

本文将非参数理论、分位回归理论和分层模型相结合起来, 同时加入混合效应, 建立分层非参数混合效应模型, 该模型第一层为非参数模型, 第二层为混合线性模型, 为了将两个层次之间的参数联系起来, 先将第一层通过泰勒展开对其进行参数化得到参数 θ , 然后对 θ 进行第二层建模, 这是对分层线性模型等基础模型的合理推广。然后通过检验函数的非参数分位回归对其进行估计, 在这个过程中, 模型参数的估计精度与核函数和窗宽的选择有很大的关系。在非参数估计中, 当数据样本容量足够大时, 不论选取哪种形式的核函数, 都可以保证在一定的正则条件下, 相应估计量具有相合性这一最基本的要求, 因此, 本文选择了高斯核函数。窗宽控制着估计结果的平滑性, 在此使用局部线性分位数回归中的自适应方法对窗宽进行选择。在估计的过程中由于估计结果很难给出解析解, 所以将分位数回归与 EM 算法结合形成 EQ 算法, 并进行迭代达到估计的目的。随后证明了参数估计量的渐近性质, 最后利用 R 软件进行蒙特卡洛模拟, 分别对

误差项服从不同分布下的估计结果作比较,来说明方法在该模型估计过程中的稳定性和实用性。。

(2)现实意义:我们生活在很多自然形成的嵌套结构中,这就使得分层数据在日常生活中无处不在。个体观测嵌套在单元以及上一层结构,这种情况下看起来个体并不是独立的,处理不好分层结构和缺乏独立性的数据,经常会导致参数的有偏性。使用分层模型分析数据可以观测到不同层次下的样本信息,可以更方便的提出同一水平和不同水平的假设,进而评估不同水平上的信息。分析分层特征数据时,重点想要知道协变量与相应变量之间的关系,进而想知道协变量对响应变量整体的影响,由于数据的复杂性,将非参数回归理论和分位回归理论与分层模型相结合对于这类数据的处理具有一定的实际意义。

1.2 国内外研究现状

科技高速发展的今天,高维多元分层等特征的复杂数据分析受到了众多学者的关注。其中,针对嵌套数据的分析一直没有间断。20世纪70年代,人们开始针对分层结构的数据进行统计建模。Lindly 和 Smith (1972) 及 Smith (1973) 在研究线性模型的贝叶斯估计时首次提出了分层线性模型的概念,20世纪80年代中期,一些学者开始研究如何引入系统的方法对分层结构的数据进行建模和分析,随后分层模型被很多学者研究并应用在不同的领域。Laird 和 Ware (1982) 发现两阶段随机效应模型更容易适用于高度不平衡的数据,在嵌套结构中,一个族中会包含几个不同阶段响应变量服从的概率分布,但是一些随机效应参数在个体之间存在差异,所以在第二阶段为其假定一个分布来解释这种差异。Goldstein (1986) 针对分层模型的估计给出了一种迭代广义最小二乘估计方法,并证明了它在一般情况下等价于极大似然估计,并通过该方法对含有缺失的纵向数据进行建模估计。Willms (1989) 研究了一种纵向数据的分层模型,对有重复测量误差的不同水平的响应变量的真实变化进行了分析。Goldstein (1995) 对分层模型作了系统介绍,重点解决了重复测量数据的分析问题。Dempste (1981) 在估计分层模型时分了两种情况,一是当方差和协方差已知时,用贝叶斯方法来估计固定效应和随机效应;二是当方差和协方差未知时,它们的点估计使用极大似然估计然后通过 EM 算法迭代。贺建风 (2019) 将广义线性模型引入到分层贝叶斯方法

中, 构建了基本的理论机制和分类数据的估计模型。对于模型来说, 最重要的问题是估计出模型中的参数。杨新平 (2019) 利用极大似然方法估计了两水平线性模型的参数, 并给出相应模型的 **Bootstap** 估计, 提高了估计的精度。周梦雨 (2021) 推广了分层线性模型的估计方法, 在 EM 算法的基础上提出了针对缺失数据的 **Guass-Seidel** 型迭代算法。

在分层线性模型技术被广泛应用之后, 有学者意识到在实际应用中, 并不是那么简单的。需要考虑数据的重复性和误差项结构, 以及由于误差项分布和离群点导致的点估计和方差估计可能存在的偏差, 例如, 离群点很容易影响分布的均值, 均值回归也不能在给定协变量时刻画出响应变量的整体条件分布, 而条件分位数则可以克服这个缺点。分位数回归是 **Koenker** 和 **Bassett** (1978) 提出的, 他们利用线性和非线性模型来估计条件分位数, 目前被认为是对经典均值回归最理想的改进, 参见 **Hendricks** (1992), **Lejeune** (2003)。分位数回归 (**Koenker**, 2005) 更加详细地描述了分布的条件分位数。**Koenker** (2004) 在处理面板数据时, 发现使用分位回归模型拟合效果更好。**Lee** (2003) 研究了部分线性假设的条件分位数函数的估计问题, 给出了条件分位数参数的简单估计, 并证明了在正则条件下估计量的渐进性质。**田茂再和陈歌迈** (2006) 将分位回归引入分层模型, 提出了分层线性分位回归模型, 并在估计参数时基于 EM 算法提出了一种新的算法 **EQ** 算法, 用分位数回归 **Q** 步替代最大步 **M** 步, 这样极大地利用了分位数回归和分层模型的优点。**Tian** (2009) 在分位回归引入分层模型的基础上, 放宽了误差项正态性的假设, 还证明了通过 **EQ** 算法得到的估计量的渐近性质。**Martins** (2004) 在研究教育对不同工资水平的影响时建立了分位数回归模型, 发现收入在不同分位点上, 教育回报率是不同的, 这否定了以前基于均值回归得到的参数是一个常数的假设。**张玉臣** (2013) 也构建了分位数回归模型, 研究了不同分位数上协变量对响应变量的影响。同时, 他还通过最小二乘估计构建了均值回归模型, 发现大多数参数的拟合程度没有条件分位数拟合的好, 充分说明了均值回归不足以解释协变量对响应变量的影响。

以上的模型是基于线性回归建立的线性分位回归模型。但是, 线性分位回归模型不能真实的描述各个协变量和响应变量之间可能存在的非线性关系。所以, 学者将非参数理论引入到分位数回归分析中, 并使用混合效应模型来分析重复数

据和纵向数据。在研究非参数分位回归模型时，条件分位数都是通过运用局部加权非对称的非参数估计方法来估计的，具体参见 Jones 和 Hall(1990), Fan(1994), Honda (2000), Jones 和 Yu (2007), Spoloiny (2003)。Tao (1999) 建立了一种半参数混合效应模型，通过放宽高斯随机效应的常见假设，分析了具有连续、顺序以及二元结果的聚类或纵向数据。国内，徐志科和平根建 (2014) 对非参数分位回归估计方法做了一个综述。谢晶晶 (2019) 基于不同的损失函数，比如检验函数、条件累计分布函数、条件分位数对模型的参数进行了估计，并模拟了这些方法的合理性。田茂再 (2019) 将非参数理论引入分层模型的第一层，提出了分层半参数分位回归模型，利用 EQ 方法讨论了估计的问题并给出了其渐近性质。刘婷婷等 (2020) 给出了基于非参数分位数估计的众数回归模型，通过分布函数与分位函数的互逆性估计了模型，模拟结果显示该方法优于线性分位数估计的众数回归模型。张敏 (2020) 研究了面板数据的非参数分位回归模型，并将贝叶斯方法与非参数理论相结合建立了贝叶斯分层分位回归模型，给出了参数估计的抽样算法，结果表明引入非参数理论在无偏性和鲁棒性上都优于传统模型。

1.3 文献述评

对国内相关文献梳理可以发现，目前国内主要有以下三个方面：(1) 将分位回归引入分层模型进行研究，但是缺乏非参数理论的结合；(2) 现有的研究只考虑了固定效应，很少考虑随机效应的影响；(3) 对非参数分位回归的估计，大多数文献使用的还是传统的估计方法。

国外现有的文献主要集中于以下几个方面：(1) 大量的文献研究主要在分层线性分位数回归模型和非参数分位数回归模型上，将非参数理论和分层模型结合的情况研究较少；(2) 使用迭代算法估计非参数分位回归参数的方法研究也较为欠缺；(3) 在进行非参数分位数回归时，窗宽的选择方法需要改进。

1.4 本文研究内容

本文的研究内容主要分为五个方面：

第一部分主要包括关于分层模型、分位数回归理论、非参数回归理论的提出背景和目前相关研究的发展现状以及本文可能的创新之处，并给出了文章的结构

和内容框架。

第二部分是对不同模型理论相关知识的梳理与总结,主要涉及分层线性回归模型、非参数理论和分位回归理论,其中非参数理论中包括核函数和窗宽的选择。这些基础理论为文章的后续研究做了铺垫。

第三部分是本文研究的创新点和重点,将非参数理论、分位回归理论和分层模型结合起来,同时加入混合效应,建立分层非参数混合效应模型,该模型第一层为非参数模型,第二层为混合线性模型,为了将两个层次之间的参数联系起来,先将第一层通过泰勒展开对其进行参数化,然后进行第二层建模。接着通过检验函数的非参数分位回归对其进行估计,在估计的过程中由于估计结果很难给出解析解,所以将分位数回归与 EM 算法结合形成 EQ 算法,并进行迭代达到估计的目的。随后证明了参数估计量的渐近性质。

第四部分利用 R 软件进行 Monte Carlo 模拟研究和实例分析,分别对误差项服从不同分布下的估计结果作比较,并利用实际数据分析,来说明方法在该模型估计过程中的稳定性和实用性。

第五部分是对文章的总结和展望。

文章内容结构的流程框架如下所示：

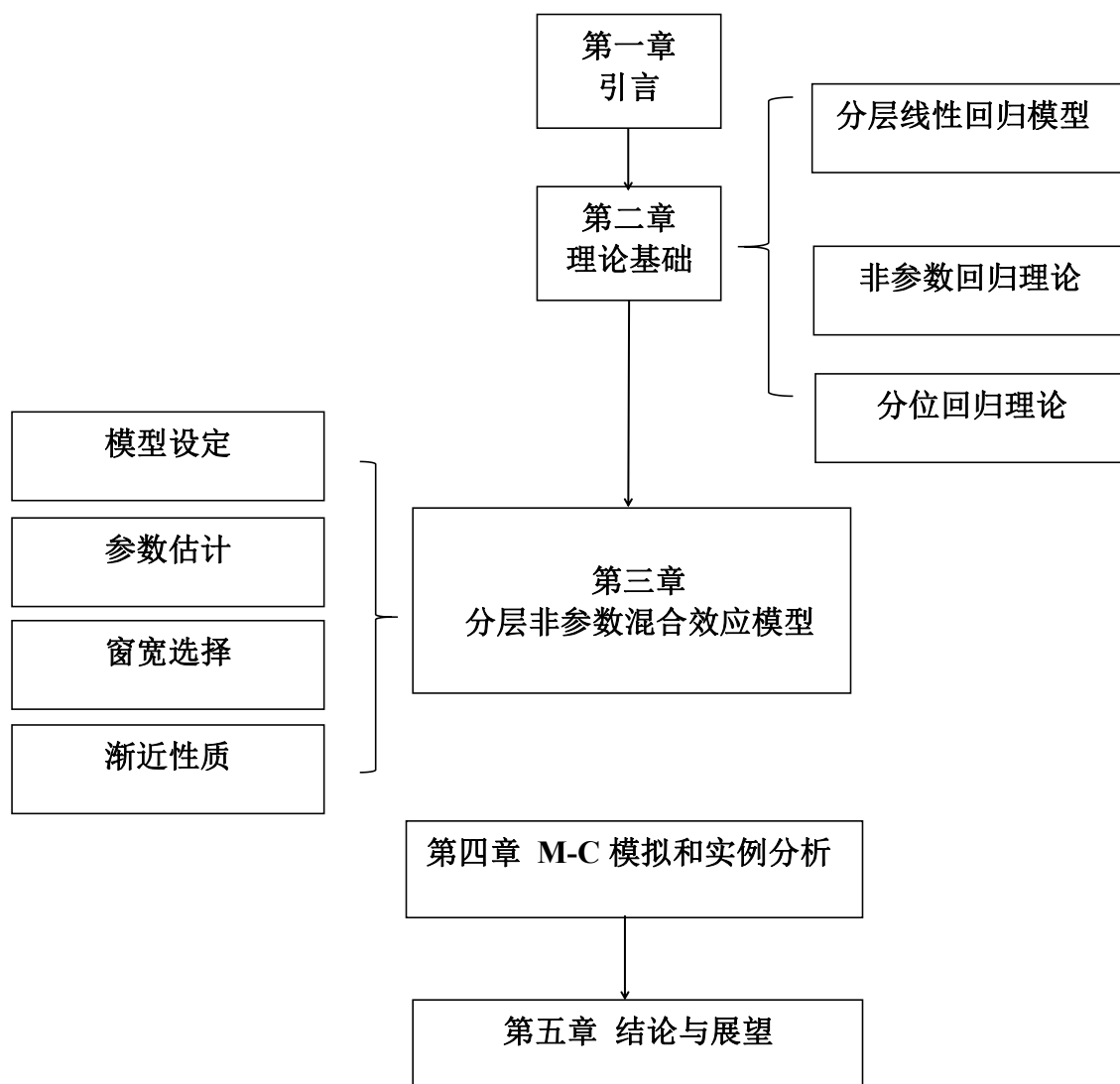


图 1.1 文章结构框架图

1.5 本文的创新之处

①将非参数理论、分位回归理论和分层模型相结合，同时加入混合效应，建立分层非参数混合效应模型，然后将分位回归和 EM 算法结合对模型参数进行估计。

②证明了参数估计量的大样本性质，得到固定效应渐近服从多元正态分布。

③在估计过程中使用局部线性分位数回归中的自适应方法对窗宽进行了选择。

2 理论基础

2.1 分层线性回归模型

分层模型具有嵌套结构，具有很多不同的层次，田茂再（2006）系统给出了分层线性分位回归模型。为了体现分层模型的特点又能够方便研究，本文给出两水平的情况。假设有来自 $(\mathbf{X}, \mathbf{W}, Y)$ 的独立同分布观测值 $\{(\mathbf{X}_1, \mathbf{W}_1, Y_1), \dots, (\mathbf{X}_n, \mathbf{W}_n, Y_n)\}$ ，其中 Y_i 是实数响应变量的值， \mathbf{x}_i 是已知的 $1 \times d$ 维第一层预测值向量， \mathbf{w}_i 是已知的 $d \times f$ 第二层预测矩阵，满足第一层模型：

$$Y_i = \mathbf{X}_i \boldsymbol{\beta}_i + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2) \quad (2.1)$$

其中 $\boldsymbol{\beta}_i$ 是第一层 $d \times 1$ 维待估计的参数向量， ε_i 是随机效应变量，不可直接观测得到，但它有两个假设：① ε_i 与 \mathbf{X}_i 相互独立；② 服从正态分布且 $E(\varepsilon_i) = 0$ ， $\text{var}(\varepsilon_i) = \sigma^2$ 。

在第二层模型中，第一层模型的系数向量 $\boldsymbol{\beta}_i$ 为响应变量：

$$\boldsymbol{\beta}_i = \mathbf{W}_i \boldsymbol{\gamma} + \mathbf{u}_i, \mathbf{u}_i \sim N(\mathbf{0}, \mathbf{T}) \quad (2.2)$$

其中 $\boldsymbol{\gamma}$ 是 $f \times 1$ 维固定效应向量， \mathbf{u}_i 是 $d \times 1$ 维第二层随机效应向量。同样假设 \mathbf{u}_i 同时与 \mathbf{w}_i 和 ε_i 相互独立，且服从多元正态分布， $E(\mathbf{u}_i) = \mathbf{0}$ ， $\text{Cov}(\mathbf{u}_i) = \mathbf{T}_{d \times d}$ 。

将第二层模型代入到第一层模型，得到以下组合模型：

$$Y_i = \mathbf{X}_i \mathbf{W}_i \boldsymbol{\gamma} + \mathbf{X}_i \mathbf{u}_i + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2), \mathbf{u}_i \sim N(\mathbf{0}, \mathbf{T}). \quad (2.3)$$

通过 Y 的分位函数，来全面刻画给定预测变量 $(\mathbf{X}, \mathbf{W}) = (x, w)$ 下的条件响应变量 Y 的条件分布 $F(y | \mathbf{x}, \mathbf{w})$ 。假定 $F(y | \mathbf{x}, \mathbf{w})$ 是 y 的增函数且在 \mathbf{x} 和 \mathbf{w} 连续，那么在给定 $\mathbf{X}=\mathbf{x}$ 和 $\mathbf{W}=\mathbf{w}$ 的条件下， Y 的 τ 分位可定义为 $q_\tau(\mathbf{x}, \mathbf{w})$ ，具体表达式如下：

$$q_\tau(\mathbf{x}, \mathbf{w}) = \arg \inf \{t \in \mathbb{R} : F(t | \mathbf{x}, \mathbf{w}) \geq \tau\}, 0 < \tau < 1. \quad (2.4)$$

在组合模型（1.3）下， Y 的 τ 分位可具体表示为

$$q_\tau(\mathbf{x}, \mathbf{w}) = \mathbf{xw}\boldsymbol{\gamma} + (\mathbf{xT}\mathbf{x}^T + \sigma^2)^{1/2} \Phi^{-1}(\tau) \quad (2.5)$$

其中 $\Phi(\cdot)$ 是标准正态分布函数。

2.2 非参数回归理论

2.2.1 非参数回归模型

考虑非参数回归模型：

$$Y_i = m(\mathbf{X}_i) + u_i \quad (2.6)$$

其中， $(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)$ 是随机样本，假设样本间独立同分布， $m(\cdot)$ 的具体函数形式和随机误差项的分布都是未知的。

参数模型是最常用的模型，在具体操作过程中容易估计出参数和解释其具体的含义。在理论建模或者统计推断过程中，不管协变量和响应变量之间是哪种关系，都要假设它们之间的回归函数形式是已知的。但在分析实际数据时，很多协变量和响应变量的关系往往无法确定或者根本未知，如果不管变量之间的这种关系直接设定函数形式，一旦模型的设定出现错误，那么后续的所有研究都将没有意义。因此，当协变量和响应变量之间存在比较复杂的非线性关系，或者无法知道它们之间的关系，这种关系就需要考虑新的理论来支撑研究。

非参数理论在 20 世纪 30 年代开始被研究，延续近几十年以来，它还是比较新颖和值得研究的话题。与参数回归模型不同的是，非参数回归模型有很强的适应性和灵活性，这种灵活性是基于模型自身所含有的特点。最实用的是，非参数回归模型在建模之前不用提前假设它具体是哪种函数形式，这使得它几乎可以应用于所有数据。其次，非参数回归模型不假设随机误差项服从特定分布，更不设定误差项的特征函数。同时，在这么灵活条件下，模型还能充分地包含样本数据的特点。

在非参数回归模型的参数估计问题中，不提前假设回归函数的具体形式，直接利用样本数据来回归每个点的函数值，即在 $\mathbf{x} = \mathbf{X}_i$ 处， $\hat{m}(\mathbf{X}_i)$ 的值由样本 Y_i 的加权值决定。

2.2.2 核函数

在非参数回归模型的估计过程中，由于回归函数的具体形式是未知的，密度函数的估计精度取决于权重，而权重由核函数和窗宽来决定，因此，非参数回归模型估计最关键的就是核函数和窗宽的选取。

设有随机变量 \mathbf{x} ，密度函数为 $f(\mathbf{x})$ ，则密度函数 $f(\mathbf{x})$ 的核估计为 $\hat{f}(\mathbf{x}) = \frac{1}{n \det(H)} \sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})$ $K(\cdot)$ 为核函数， \mathbf{H} 为窗宽。如果该估计满足以下性质，则 $\hat{f}(\mathbf{x})$ 是 $f(\mathbf{x})$ 的一致估计量：

$$\int K(z) dz = 1 ;$$

$$\int z K(z) dz = 0 ;$$

$$\int z^2 K(z) dz = \mu_2 < \infty .$$

从上面核估计与核函数的性质可以看出，核估计的思想就是在估计过程中，通过核函数对每个样本点赋予不同的权重，样本点在节点 \mathbf{x} 处所占的权重通过点 \mathbf{X}_i 与 \mathbf{x} 之间的距离 $\mathbf{X}_i - \mathbf{x}$ 来确定。具体来说，离 \mathbf{x} 越近的数据点所赋予的权重越大，离 \mathbf{x} 越远的点所赋予的权重越小。

均匀核函数、高斯核函数等都是可供选择的核函数。然而，当数据样本容量足够大，不论选取哪种核函数，都可以在一定的正则条件下，保证估计量具有相合性这一最基本的要求，因此，本文选取高斯核函数作为非参数估计的权重。

2.2.3 窗宽选择

在非参数估计中，窗宽 \mathbf{H} 控制着估计结果的平滑性，因此窗宽的选择非常重要。窗宽过大，会使估计结果更加稳定，但会忽视部分局部信息，使得估计结果过于平滑，进而使偏差过大；反之，窗宽太小，估计偏差也会变小，但会过度依赖现有观测值来体现局部特征，从而使得估计结果波动变大，进而使方差变大。窗宽实际上和核函数一样，对估计起到惩罚的作用，但是窗宽是在平衡估计的偏差与方差，因此，需要选择合适的窗宽使得偏差和方差达到平衡。目前，在理论研究中，常见的窗宽选择方法有三种（谢晶晶，2019），分别是理论窗宽最佳选

择法、最小二乘交叉验证法、经验法则。

(1) 理论窗宽最佳选择法

有随机变量 \mathbf{X} ，令 $v_0 = \int K^2(z)dz$ ， $\mu_2 = \int z^2 K(z)dz$ ，在渐近条件下可得模型估计的均方误差为 $MSE(\hat{q}_\tau(s)) \approx \frac{1}{4}(\det(H))^4 \mu_2(K)^2 q_\tau''(\mathbf{x})^2 + \frac{R(K) \tau(1-\tau)}{n \det(H) g(\mathbf{x}) f(q_\tau(\mathbf{x}) | \mathbf{x})^2}$ ，由此得到渐近偏差为 $\frac{(\det(H))^2}{2} (m''(\mathbf{x}) + \frac{2m'(\mathbf{x})f'(\mathbf{x})}{f(\mathbf{x})}) \mu_2$ ，渐近方差为 $\frac{\sigma^2(\mathbf{x})}{n \det(H) f(\mathbf{x})} v_0$ ，其中 $\sigma^2(\mathbf{x})$ 是响应变量的条件方差。

在非参数回归模型中，回归模型 $m(x)$ 估计的渐近偏差和渐近方差会根据窗宽的变化而变化，即窗宽变大时渐近偏差会变大，渐近方差会变小，窗宽变小时，渐近偏差会变小，渐近方差会变大。因此，合适的窗宽就是让渐近偏差和渐近方差达到平衡。

一维情况下，最优渐近理论的窗宽窗宽为 $h_n = cn^{-1/5}$ ，其中 c 为一个常数，与 n 无关，但与核函数以及协变量的密度函数有关，在具体的窗宽选择中可以通过不断调整 c 的值来确定最优窗宽。

(2) 最小二乘交叉验证法

这种方法来是通过最小化相应估计的积分平方误差来选择窗宽，即该方法给出了对应于 $f(\mathbf{x})$ 的支撑集中任一点 \mathbf{x} 处的最优窗宽。它的窗宽过程可以总结为：在某个局部观测点 $\mathbf{x} = \mathbf{X}_i$ ，首先在样本中剔除掉这一样本点 (\mathbf{X}_i, Y_i) ，然后在点 $\mathbf{x} = \mathbf{X}_i$ 处，对剩余的 $n-1$ 个样本观测点进行核估计：

$$\hat{m}_{(-i)} = \frac{\sum_{j=1, j \neq i}^n Y_j K_H(\mathbf{X}_j - \mathbf{x})}{\sum_{j=1, j \neq i}^n K_H(\mathbf{X}_j - \mathbf{x})}$$

最后，通过比较平方拟合误差：

$$CV(H) = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{m}_{(-i)}(\mathbf{x}))^2 \frac{K_H(\mathbf{X}_i - \mathbf{x})}{\sum_{i=1}^n K_H(\mathbf{X}_i - \mathbf{x})}$$

从而挑选出窗宽的最优值，即

$$H = \arg \min_h CV(H).$$

(3) 经验法则

当核函数 $K(\cdot)$ 在 $[-1,1]$ 上满足像高斯核函数这样对称、单峰的概率密度时，非

参数模型中回归函数的估计 $\hat{m}(\mathbf{x}) = \sum_{i=1}^n w_i Y_i$ 是集中在 \mathbf{x} 附近的加权平均, 由于 \mathbf{x} 是对称的, 则取的窗宽恰好是该取值区域的宽度。当窗宽逐渐增大时, 参与平均的样本观测点就很多, 能够提高估计的精度, 但是在这一过程中估计的偏差就会增大。反之, 当窗宽逐渐减小, 偏差就会变小, 但是方差就会增大。所以, 可以根据该特征, 绘制样本观测点的散点图, 并根据散点图的特征来选择窗宽, 使得观测点 Y_i 的局部区域的加权平均能够达到预期的估计效果。

2.3 分位回归理论

设有随机变量 Y , 则 $q(\tau) = \inf\{y: F(y) \geq \tau\}$ 为 Y 的第 τ 分位数, 其中 τ 是 $(0,1)$ 之间的任意一个数, 分位数可以描述随机变量的整体特征, 它将 Y 的概率分布范围分成了不等的两个部分, 比例为 τ 和 $1-\tau$ 。

Koenker 和 Bassett 在 1978 年开创性的提出了线性分位回归模型, 而后学者对其做了大量的研究并取得了丰富的成果, 后面又将该理论与分层模型、非参数理论、贝叶斯等相结合, 使得分位数回归理论得到了广泛的科学研究。

设随机变量 \mathbf{x} 和 Y , 有条件概率密度函数 $f(y|\mathbf{x})$ 和条件累计分布函数 $F(y|\mathbf{x})$, 假设给定 \mathbf{x} 下 Y 的 τ 条件分位数为:

$$q_{\tau}(y) = \inf\{y: F(y|\mathbf{x}) \geq \tau\} = F^{-1}(\tau|\mathbf{x})$$

在线性条件下, 给定 $\mathbf{X} = \mathbf{x}$ 下, Y 的分位数函数为:

$$q_{\tau}(y|\mathbf{X}=\mathbf{x}) = \mathbf{x}^T \boldsymbol{\beta}(\tau)$$

其中 $\boldsymbol{\beta}(\tau)$ 为未知参数, 且令 $q_{\varepsilon, \tau}$ 是 ε 的第 τ 分位数, 则随机误差项满足分位数约束 $q_{\varepsilon, \tau} = 0$ 。就可以通过最小化以下目标得到参数 $\boldsymbol{\beta}(\tau)$ 的估计值:

$$\hat{\boldsymbol{\beta}}(\tau) = \arg \min_{\boldsymbol{\beta}} E[\rho_{\tau}(Y - \mathbf{X}^T \boldsymbol{\beta})]$$

其中 $\rho_{\tau}(z) = z(\tau - I(z < 0))$, 称为损失函数, $I(\cdot)$ 示性函数, 则上式就是的最小一乘回归。如果有随机样本 $\{\mathbf{X}_i, Y_i\}_1^n$, 且样本间是独立同分布的, 则模型中参数 $\boldsymbol{\beta}(\tau)$ 的估计可以通过下列式子给出:

$$\begin{aligned}\hat{\boldsymbol{\beta}}(\tau) &= \arg \min_{\boldsymbol{\beta}} \sum_{i=1}^n \rho_{\tau}(Y_i - \mathbf{X}_i^T \boldsymbol{\beta}) \\ &= \arg \min_{\boldsymbol{\beta}} \left[\sum_{i \in \{i: Y_i \geq \mathbf{X}_i^T \boldsymbol{\beta}\}} \tau(Y_i - \mathbf{X}_i^T \boldsymbol{\beta}) + \sum_{i \in \{i: Y_i < \mathbf{X}_i^T \boldsymbol{\beta}\}} (\tau-1)(\mathbf{X}_i^T \boldsymbol{\beta} - Y_i) \right]\end{aligned}\quad (2.7)$$

(2.7) 式即为分位数回归的基本表达式，它表示对最小化加权离差的绝对值进行求和，绝对值离差的权重分别为 τ 和 $1-\tau$ ，将满足上式的 $\boldsymbol{\beta}(\tau)$ 参数估计结果代入到分位函数中，便可以得到 Y 的条件分位数回归。随着不同分位点的变化，参数 $\boldsymbol{\beta}(\tau)$ 的分位数回归估计结果就不同，这正是分位数回归最大的优势。当分位数 τ 在 0 到 1 之间取不同的值时，可以得到一组响应变量关于协变量的分位数回归曲线，这些曲线在不同的应用场景中就有着不同的实际含义。

在对数据进行回归建模时，目的就是利用协变量中的信息，通过建立合理的回归模型并对其进行估计之后，来描述协变量对响应变量的影响，那么，评价拟合估计的好坏也是需要重点关注的。在评价过程中，使用的原则就是使拟合值与真实值之间的差距越小越好。分位数回归采用最小化误差的绝对值之和，即绝对离差法 (LAD)。在对回归模型参数的解释上，分位数回归模型和均值回归也不同，均值回归估计主要是协变量对被响应变量的平均影响，而分位数回归反映的是在给定某个分位点下协变量对响应变量的影响情况，然后通过改变分位点来刻画协变量对响应变量整体的影响。因此，分位数回归在估计过程中不易受极端值的影响，从而使得估计结果就更为稳健。

分位数回归可以在所有分位点下求参数估计的问题，在理论上看似非常的简洁，但是分位数回归采用的是最小化加权绝对离差法，这使得分位数回归在求解的过程中变得非常困难。在目前的研究中，已知的方法主要有单纯形法、内点算法、平滑算法和自适应算法等，也有学者提出了迭代算法，这些算法逐步地提高了运算效率。同时，许多统计类的软件也已经安装了很多分位数回归算法的程序，例如 R 官网已创建了 `quantreg` 包，为学者们提供了研究分位数回归理论的一系列开放平台。

3 分层非参数混合效应模型

3.1 模型设定

在传统的分层线性模型中，模型假定是线性的。在这种情况下，它的条件均值函数为 $\mu(\mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x})$ ，它的一阶偏导向量如下定义：

$$\nabla\mu(\mathbf{x}) = (\partial\mu(\mathbf{x})/\partial x_1, \dots, \partial\mu(\mathbf{x})/\partial x_d)^T = (\beta_1, \dots, \beta_d)^T$$

其中 $\beta_k (k=1, \dots, d)$ 是当其他协变量不变时，第 k 个协变量的平均响应。在计量经济学中， $\nabla\mu(\mathbf{x})$ 通常被称为边际效应。例如，若 $\mu(\mathbf{x}) = \mathbf{x}^T \beta$ ，则 $\frac{\partial\mu}{\partial x_k} = \beta_k$ ， β_k 即为边际效应，类似地，非参数模型关于某个变量的求导也可以称之为边际效应，同样，在分层模型中，这个边际效应将由第二层解释变量来解释。

然而，模型中的线性假设在实际应用中是不现实的。本文将分层线性模型的第一层推广至非参数模型，并在第二层加入混合效应，建立一种分层非参数混合效应模型。第一层的表达式为：

$$Y_{ij} = m_i(\mathbf{X}_{ij}) + \varepsilon_{ij}, i=1, \dots, n, j=1, \dots, n_i \quad (3.1)$$

其中 $m_i(\cdot)$ 是每一层控制个体内部行为的光滑回归函数，具体形式是未知的， ε_{ij} 是不可观测的随机误差变量，假设它的方差为 σ_i^2 。考虑模式 (3.1) 中的非参数模型，假设 $m_i(\cdot)$ 的一阶偏导在 \mathbf{x} 点出连续，对于 \mathbf{x} 点周围的样本点 \mathbf{X}_{ij} ，运用非参数估计中基于局部拟合的思想，通过泰勒展开估计 $m_i(\mathbf{X}_{ij})$ ，有

$$\begin{aligned} m_i(\mathbf{X}_{ij}) &\approx m_i(\mathbf{x}) + (\mathbf{X}_{ij} - \mathbf{x})^T \nabla m_i(\mathbf{x}) \\ &= (1, \mathbf{X}_{ij} - \mathbf{x}) \begin{pmatrix} m_i(\mathbf{x}) \\ \nabla m_i(\mathbf{x}) \end{pmatrix} \\ &\stackrel{\Delta}{=} \tilde{\mathbf{X}}_{ij} \boldsymbol{\theta}_i \end{aligned} \quad (3.2)$$

$$\boldsymbol{\theta}_i = (m_i(\mathbf{x}), \nabla m_i(\mathbf{x}))^T \stackrel{\Delta}{=} (\theta_{i0}, \theta_{i1}, \dots, \theta_{id})^T。$$

线性混合模型在处理实际数据时具有较强的灵活性，它可以分析面板数据、纵向数据、组数据等重复测量的数据，突破了传统线性模型模型只能处理观测值是独立和等方差的局限性，对观测值的协方差矩阵有了更灵活的设定，从而能够

更好的观测数据。同时，相对于一般线性回归模型，线性混合效应模型针对多水平的随机效应部分给出了更合理的假设。所以在此基础上给出分层非参数混合效应模型的第二层：

$$\boldsymbol{\theta}_i = \mathbf{W}_i \boldsymbol{\gamma}(x) + \mathbf{Z}_i \mathbf{U}_i \quad (3.3)$$

其中， \mathbf{W}_i 是第二层 $(d+1) \times f$ 维解释变量， $\boldsymbol{\gamma}(x)$ 是 $f \times 1$ 维的固定效应， \mathbf{U}_i 是 $f \times 1$ 维的随机效应向量，和 ε_{ij} 相互独立，服从均值为 $\mathbf{0}$ ，方差为 \mathbf{T} 的正态分布， \mathbf{Z}_i 是 $(d+1) \times f$ 维随机变量，其构成方式与 \mathbf{W}_i 相同。

将第二层代入第一层得到组合模型为

$$Y_{ij} \approx \tilde{\mathbf{X}}_{ij} \mathbf{W}_i \boldsymbol{\gamma}(x) + \tilde{\mathbf{X}}_{ij} \mathbf{Z}_i \mathbf{U}_i + \varepsilon_{ij} \quad (3.4)$$

3.2 参数估计

利用分位回归方法估计参数时，需要量化模型（3.4）中 Y_{ij} 的 τ 阶分位数和协变量之间的关系，下面根据2.3节给出的条件分位数的定义来描述二者之间的关系。

模型（3.4）中，令 $\eta_{ij} = \tilde{\mathbf{X}}_{ij} \mathbf{Z}_i \mathbf{U}_i + \varepsilon_{ij}$ ，且 $\eta_{ij} \sim G_{ij}$ ， $F(y)$ 为 Y_{ij} 的分布函数，于是有

$$\begin{aligned} \tau &= P(Y_{ij} \leq F_{ij}^{-1}(\tau) | \mathbf{X}_{ij}, \mathbf{W}_i, \mathbf{Z}_i) \\ &\approx P(\tilde{\mathbf{X}}_{ij} \mathbf{W}_i \boldsymbol{\gamma}_\tau(x) + \tilde{\mathbf{X}}_{ij} \mathbf{Z}_i \mathbf{U}_i + \varepsilon_{ij} \leq F_{ij}^{-1}(\tau)) \\ &= P(\tilde{\mathbf{X}}_{ij} \mathbf{Z}_i \mathbf{U}_i + \varepsilon_{ij} \leq F_{ij}^{-1}(\tau) - \tilde{\mathbf{X}}_{ij} \mathbf{W}_i \boldsymbol{\gamma}_\tau(x)) \\ &= P(\eta_{ij} \leq F_{ij}^{-1}(\tau) - \tilde{\mathbf{X}}_{ij} \mathbf{W}_i \boldsymbol{\gamma}_\tau(x)) \end{aligned}$$

令 $\xi_{ij}(\tau)$ 为 G_{ij} 的 τ 阶分位数，则

$$\xi_{ij}(\tau) \approx F_{ij}^{-1}(\tau) - \tilde{\mathbf{X}}_{ij} \mathbf{W}_i \boldsymbol{\gamma}_\tau(x) \quad (3.5)$$

从而可以得到对所有定义域中的 y_{ij} 服从 $F(y)$ ，在给定 $(\mathbf{X}_{ij}, \mathbf{W}_i, \mathbf{Z}_i)$ 后 Y_{ij} 的 τ 阶分位数是

$$F_{ij}^{-1}(\tau) = \tilde{\mathbf{X}}_{ij} \mathbf{W}_i \boldsymbol{\gamma}_\tau(x) + \xi_{ij}(\tau) \quad (3.6)$$

其中 $\gamma_\tau(x)$ 是随 τ 变化的固定效应函数。

特殊地, 假设 ε_{ij} 独立同分布于均值为 0, 方差为 σ^2 的正态分布, \mathbf{U}_i 和 ε_{ij} 关于所有的 $i=1, \dots, n$, $j=1, \dots, n_i$ 都相互独立, 则可以根据误差项的分布得出 Y_{ij} 的分布, 即 $Y_{ij} \sim N(\tilde{\mathbf{X}}_{ij} \mathbf{W}_i \gamma(x), \tilde{\mathbf{X}}_{ij} \mathbf{Z}_i \mathbf{T} \mathbf{Z}_i^T \tilde{\mathbf{X}}_{ij}^T + \sigma^2)$, 进而 $\xi_{ij}(\tau) = (\tilde{\mathbf{X}}_{ij} \mathbf{Z}_i \mathbf{T} \mathbf{Z}_i^T \tilde{\mathbf{X}}_{ij}^T + \sigma^2)^{1/2} \Phi^{-1}(\tau)$, $\Phi(\cdot)$ 为标准正态分布。这种假设在后面估计方差和协方差时会用到。

对于模型 (3.4), 接下来需要对任意给定 \mathbf{x} 的三个参数进行估计, 分别是 $\gamma(x)$, σ_i^2 和 \mathbf{T} 。模型中含有随机效应, 一般情况下, 参数的估计没有解析解, 只能通过迭代算法求解。本文选取 EQ 算法进行迭代, EQ 算法是在 EM 算法的基础上, 用分位回归替代最大步。EM 算法是由 Dempster 等人 (1977) 提出, 主要用于求不完全数据模型中参数的最大似然估计, 它的基本原理是在每一次迭代后都可以使后验密度函数值增加。在模型 (3.4) 中, 随机效应无法观测, 相当于缺失数据, 因此可以把 EQ 算法应用到分层非参数混合效应模型中。下面给出具体的 EQ 算法。

3.2.1 Q 步

对于 $\gamma(x)$ 的估计, 本文使用基于检验函数的非参数分位回归来估计。针对模型 (3.4), 通过最小化以下目标函数可以估计出固定效应 $\gamma(x)$, 记为 $\hat{\gamma}_\tau(x)$:

$$\hat{\gamma}_\tau(x) = \arg \min_{\gamma} \sum_{i=1}^n \sum_{j=1}^{n_i} \rho_\tau(Y_{ij} - \tilde{\mathbf{X}}_{ij} \mathbf{W}_i \gamma(x) - \tilde{\mathbf{X}}_{ij} \mathbf{Z}_i^T \mathbf{U}_i) \mathbf{K}_{\mathbf{H}} \{ \mathbf{H}^{-1}(\mathbf{X}_{ij} - \mathbf{x}) \} \quad (3.7)$$

其中, $\rho_\tau(\cdot)$ 是检验函数, $\rho_\tau(z) = \tau z I_{[0, \infty)}(z) - (1 - \tau) z I_{(-\infty, 0)}(z)$, $I(\cdot)$ 是示性函数。 $\mathbf{K}(\cdot)$ 为核函数, 是对称、单峰的密度函数, $\mathbf{K}_{\mathbf{H}} \{ \mathbf{H}^{-1}(\mathbf{X}_{ij} - \mathbf{x}) \} = \frac{1}{\det(\mathbf{H})} K(\mathbf{H}^{-1}(\mathbf{X}_{ij} - \mathbf{x}))$, \mathbf{H} 为窗宽, 令 $\mathbf{H} = h \mathbf{I}_d$, \mathbf{I}_d 是 $d \times d$ 维单位矩阵。

根据 Y_{ij} 的 τ 阶分位数和协变量之间的关系, (3.7) 式可以进一步写为

$$\hat{\gamma}_\tau(x) = \arg \min_{\gamma} \sum_{i=1}^n \sum_{j=1}^{n_i} \rho_\tau(Y_{ij} - \tilde{\mathbf{X}}_{ij} \mathbf{W}_i \gamma(x) - \xi_{ij}(\tau)) \mathbf{K}_{\mathbf{H}} \{ \mathbf{H}^{-1}(\mathbf{X}_{ij} - \mathbf{x}) \} \quad (3.8)$$

上式中 $K(\cdot)$ 的作用是通过给不同的样本点赋予不同的权重, 来控制第一层解

释变量的波动程度。有了这个权重，最直观的意义就是在估计 $\hat{\gamma}_\tau(x)$ 时，估计值的精度不易受异常点或者 Y_{ij} 的分布函数有重尾等尾部特征的影响；同时，可以增强 Y_{ij} 的分布函数在其定义域内的平滑性。

在非参数估计中，当数据样本容量足够大时，不论选取哪种形式的核函数，都可以保证在一定的正则条件下，相应估计量具有相合性这一最基本的要求（谢晶晶，2019），因此，本文在估计 $\gamma(x)$ 时选择了高斯核函数。

h 为窗宽，控制着估计结果的平滑性。窗宽过大，估计结果会比较稳定，但会忽视部分局部信息，使得估计结果过于平滑，偏差过大；反之，窗宽太小，估计偏差也会变小，但会过度依赖现有观测值来体现局部特征，从而使得估计结果波动变大，进而使方差变大。因此，需要选择合适的窗宽使得偏差和方差达到平衡。这里，我们假设窗宽相等均为 h ，在 EQ 算法迭代过程中，取 h 为简单情形下的理论窗宽 $h = n^{-1/5}$ ，具体的窗宽选择在 3.3 节给出。

解决（3.8）式中的最小化问题，就涉及到样本分位数的估计。由（3.5）式得，在给定样本时，样本 τ 阶分位数为

$$\hat{\xi}_{ij}(\tau) = Y_{ij} - \tilde{\mathbf{X}}_{ij} \mathbf{W}_i \hat{\gamma}_\tau(\mathbf{X}_{ij})$$

方便起见，将其进行取整，即

$$\hat{\xi}_{ij}(\tau) = (Y_{ij} - \tilde{\mathbf{X}}_{ij} \mathbf{W}_i \hat{\gamma}_\tau(\mathbf{X}_{ij}))_{([s])} \quad (3.9)$$

其中 $[s]$ 表示不大于真实值的最大整数，式（3.8）定义了分层数据中第 i 个单元 n_i 个个体的第 $[s]$ 个最小值。

对于 \mathbf{T} 和 σ_i^2 ，最大似然估计可以直接给出：

$$\hat{\mathbf{T}} = \frac{1}{n} \sum_{i=1}^n \mathbf{Z}_i \mathbf{U}_i \mathbf{U}_i^T \mathbf{Z}_i^T$$

$$\hat{\sigma}_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \tilde{\mathbf{X}}_{ij} \mathbf{W}_i \hat{\gamma}_\tau(x) - \tilde{\mathbf{X}}_{ij} \mathbf{Z}_i \mathbf{U}_i)^2$$

3.2.2 E 步

由于随机向量的不可观测性，Q 步给出的只是参数在形式上的表达式，不可

直接计算。具体的估计可以在给定数据 Y_{ij} 和参数初始值后，通过 Q 步得到的表达式 (3.7) 和 (3.8) 以及参数的条件期望来估计。

在给定数据 Y 和其他参数的情况下， σ_i^2 和 \mathbf{T} 的条件期望为

$$E(\hat{\mathbf{T}} | Y, \gamma_\tau(x), \sigma_i^2, \mathbf{T}) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{n_i} (\mathbf{Z}_i \mathbf{U}_i^* \mathbf{U}_i^{*T} \mathbf{Z}_i^T + \mathbf{T}_i^*)$$

$$E(\hat{\sigma}_i^2 | Y, \gamma_\tau(x), \sigma_i^2, \mathbf{T}) = \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \tilde{\mathbf{X}}_{ij} \mathbf{W}_i \gamma_\tau(x) - \tilde{\mathbf{X}}_{ij} \mathbf{Z}_i \mathbf{U}_i^*)^2 + \text{tr}(\tilde{\mathbf{X}}_{ij}^T \mathbf{T}_i^* \tilde{\mathbf{X}}_{ij}),$$

其中， $\mathbf{T}_i^* = \sigma_i^2 (\tilde{\mathbf{X}}_{ij}^T \tilde{\mathbf{X}}_{ij} + \sigma_i^2 \mathbf{T}^{-1})^{-1}$, $\mathbf{U}_i^* = \mathbf{T}_i^* \tilde{\mathbf{X}}_{ij} (Y_{ij} - \tilde{\mathbf{X}}_{ij} \mathbf{W}_i \gamma_\tau(x)) / \sigma_i^2$ 。

3.2.3 迭代

得到参数的条件期望和分位回归估计表达式后，就可以根据二者进行迭代。在迭代运算的过程中，迭代过程需要依据迭代判断准则停止，判断准则的依据是所有参数的绝对变化值小于一个预先设定的容忍度就停止迭代，这个容忍度可以主观给定。下面给出具体的迭代过程：

(1) 初始化参数值 $(\hat{\sigma}_i^{2(0)}, \hat{\mathbf{T}}^{(0)}, \hat{\gamma}_\tau^{(0)}(x))$

通过普通最小二乘估计和方差的定义得到参数的初始值，即

$$\hat{\gamma}_\tau^{(0)}(x) = ((\tilde{\mathbf{X}}_{ij} \mathbf{W}_i)^T (\tilde{\mathbf{X}}_{ij} \mathbf{W}_i))^{-1} \Sigma (\tilde{\mathbf{X}}_{ij} \mathbf{W}_i)^T (Y_{ij} - \tilde{\mathbf{X}}_{ij} \mathbf{Z}_i \mathbf{U}_i)$$

$$\hat{\sigma}_i^{2(0)} = \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \tilde{\mathbf{X}}_{ij} \mathbf{W}_i \hat{\gamma}_\tau^{(0)}(x) - \tilde{\mathbf{X}}_{ij} \mathbf{Z}_i \mathbf{U}_i)^2$$

$$\hat{\xi}_{ij}^{(0)}(\tau) = (Y_{ij} - \tilde{\mathbf{X}}_{ij} \mathbf{W}_i \hat{\gamma}_\tau^{(0)}(x))_{(n_i, \tau)}$$

$$\hat{\mathbf{T}}^{(0)} = \frac{1}{n} \sum_{i=1}^n \mathbf{Z}_i \mathbf{U}_i \mathbf{U}_i^T \mathbf{Z}_i^T$$

(2) 根据 Q 步的分位回归和 E 步的条件期望更新参数得到 $(\hat{\sigma}_i^{2(1)}, \hat{\mathbf{T}}^{(1)}, \hat{\gamma}_\tau^{(1)}(x))$

$$\hat{\gamma}_\tau^{(1)}(x) = \arg \min_{\gamma} \sum_{i=1}^n \sum_{j=1}^{n_i} \rho_\tau(Y_{ij} - \tilde{\mathbf{X}}_{ij} \mathbf{W}_i \gamma_\tau^{(0)}(x) - \hat{\xi}_{ij}^{(0)}(\tau)) \mathbf{K}_H \{ \mathbf{H}^{-1}(\mathbf{X}_{ij} - \mathbf{x}) \}$$

$$\hat{\xi}_{ij}^{(1)}(\tau) = \arg \min_a \sum_{i=1}^n \sum_{j=1}^{n_i} \rho_\tau(Y_{ij} - \tilde{\mathbf{X}}_{ij} \mathbf{W}_i \hat{\gamma}_\tau^{(1)}(x) - a)$$

$$\hat{\mathbf{T}}^{(1)} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{n_i} (\mathbf{Z}_i \mathbf{U}_i^{*(0)} \mathbf{U}_i^{*T(0)} \mathbf{Z}_i^T + \mathbf{T}_i^{*(0)})$$

其中， $\mathbf{T}_i^{*(0)} = \sigma_i^{2(0)} (\tilde{\mathbf{X}}_{ij}^T \tilde{\mathbf{X}}_{ij} + \sigma_i^{2(0)} \mathbf{T}^{-1(0)})^{-1}$ ， $\mathbf{U}_i^{*(0)} = \mathbf{T}_i^{*(0)} \tilde{\mathbf{X}}_{ij} (Y_{ij} - \tilde{\mathbf{X}}_{ij} \mathbf{W}_i \hat{\gamma}_\tau^{(1)}(x)) / \sigma_i^{2(0)}$ 。

$$\hat{\sigma}_i^{2(1)} = \frac{1}{n_i} \left\{ \sum_{j=1}^{n_i} (Y_{ij} - \tilde{\mathbf{X}}_{ij} \mathbf{W}_i \gamma_\tau^{(1)}(x) - \tilde{\mathbf{X}}_{ij} \mathbf{Z}_i \mathbf{U}_i^{*(1)})^2 + \text{tr}(\tilde{\mathbf{X}}_{ij}^T \mathbf{T}_i^{*(1)} \tilde{\mathbf{X}}_{ij}) \right\}$$

其中, $\mathbf{T}_i^{*(1)} = \sigma_i^{2(0)} (\tilde{\mathbf{X}}_{ij}^T \tilde{\mathbf{X}}_{ij} + \sigma_i^{2(0)} \mathbf{T}^{-1(1)})^{-1}$, $\mathbf{U}_i^{*(1)} = \mathbf{T}_i^{*(1)} \tilde{\mathbf{X}}_{ij} (Y_{ij} - \tilde{\mathbf{X}}_{ij} \mathbf{W}_i \gamma_\tau^{(1)}(x)) / \sigma_i^{2(0)}$ 。

(3) 重复第(2)步的过程, 不断更新产生 $(\hat{\gamma}_\tau^{(2)}(x) | \hat{\gamma}_\tau^{(1)}(x), \hat{\mathbf{T}}^{(1)}, \hat{\sigma}_i^{2(1)})$, $(\hat{\mathbf{T}}^{(2)} | \hat{\gamma}_\tau^{(2)}(x), \hat{\mathbf{T}}^{(1)}, \hat{\sigma}_i^{2(1)})$, $(\hat{\sigma}_i^{2(2)} | \hat{\gamma}_\tau^{(2)}(x), \hat{\mathbf{T}}^{(2)}, \hat{\sigma}_i^{2(1)})$, ..., 一直迭代下去, 直到每个参数的绝对变化量小于给定的容忍度为止。

3.3 窗宽选择

本文利用核权局部线性拟合非参数分位回归估计, 在这个过程中, 曲线的好坏依赖于窗宽 \mathbf{h} 的选择, 因此, 窗宽的选择非常重要。假设窗宽相等, (3.8) 式中窗宽的核函数可以表示为 $K_d(\mathbf{X}_{ij} - \mathbf{x}) = \frac{1}{h^d} K(\frac{X_{ij1} - x_1}{h}, \dots, \frac{X_{ijd} - x_d}{h})$ 。接下来使用理论窗宽最优选择方法来选择窗宽 (田茂再, 2019), 在等窗宽的假设下, 窗宽选择只考虑一维的情况。

理论窗宽最优选择法依赖于评价非参数估计量好坏的均方误差。首先根据 Fan (1994) 给出局部线性条件分位回归拟合的 $MSE(\hat{q}_\tau(x))$:

$$MSE(\hat{q}_\tau(x)) \approx \frac{1}{4} h^4 \mu_2(K)^2 q_\tau''(x)^2 + \frac{R(K)\tau(1-\tau)}{nhg(x)f(q_\tau(x)|x)^2} \quad (3.10)$$

其中 $\mu_2(K) = \int u^2 K(u) du$, $R(K) = \int K^2(u) du$, $g(x)$ 是 X 的边缘分布。3.2 节得到了 $\gamma(x)$ 的估计结果, 从 (3.2) 式第一层非参数模型的展开方式可以发现, 取 $\hat{\gamma}_\tau(x)$ 的第一列即为 $m_i(\mathbf{x})$ 的估计量 $\hat{q}_\tau(x)$ 。

接下来从渐近最优窗宽出发, 得到 τ 分位数下的窗宽:

$$h_\tau^5 = \frac{R(K)\tau(1-\tau)}{n\mu_2(K)^2 q_\tau''(x)^2 g(x)f(q_\tau(x)|x)^2} \quad (3.11)$$

上式给出了最优窗宽和不同分位点 τ 值之间的关系:

$$\left(\frac{h_{\tau_1}}{h_{\tau_2}}\right)^5 = \frac{\tau_1(1-\tau_1) q_{\tau_2}''(x)f(q_{\tau_2}(x)|x)}{\tau_2(1-\tau_2) q_{\tau_1}''(x)f(q_{\tau_1}(x)|x)} \quad (3.12)$$

通过逼近未知相关分位数来简化 h_{τ_1} 和 h_{τ_2} 之间的关系。为了使逼近更接近真

实, 先假定一个正态分布的条件。假设 f 服从于均值为 μ_x 方差为 σ_x^2 的正态分布, 则 $f(q_\tau(x)|x) = \sigma_x^{-1} \phi(\Phi^{-1}(\tau))$, ϕ 和 Φ 是标准正态分布的密度和分布函数。由此可以得到,

$$\frac{f(q_{\tau_2}(x)|x)}{f(q_{\tau_1}(x)|x)} = \frac{\phi(\Phi^{-1}(\tau_2))}{\phi(\Phi^{-1}(\tau_1))}$$

进一步逼近可得,

$$\left(\frac{h_{\tau_1}}{h_{\tau_2}}\right)^5 = \frac{\tau_1(1-\tau_1) \phi(\Phi^{-1}(\tau_2))^2}{\tau_2(1-\tau_2) \phi(\Phi^{-1}(\tau_1))^2} \quad (3.13)$$

公式 (3.13) 给出了一个明确实用的关于分位数 τ 的修正窗宽 h 的方法。

特别地, 取 $\tau_2 = \frac{1}{2}$ 时, 有 $h_\tau^5 = \frac{1}{\pi} \frac{2\tau(1-\tau)}{\phi(\Phi^{-1}(\tau))^2} h_{1/2}^5$ 。因此, 接下来对 $\tau_2 = \frac{1}{2}$ 时的窗宽寻找一个选择方法。Ruppert 等 (1995) 研究了利用均值回归估计中 h 的最优选择 h_{mean} 表示自动窗宽 $h_{1/2}$ 。Fan (1993) 给出了 h_{mean} ,

$$h_{mean}^5 = \frac{R(K)\sigma^2(x)}{n\mu_2(K)^2 \{m''(x)\}^2 g(x)} \quad (3.14)$$

$m(x)$ 和 $\sigma^2(x)$ 是条件均值和方差, 由此得出 h_{mean} 和 $h_{1/2}$ 的关系为:

$$\left(\frac{h_{mean}}{h_{1/2}}\right)^5 = \frac{4q_{1/2}''(x)^2 \sigma^2(x) f(q_{1/2}(x)|x)^2}{m''(x)^2}$$

上式中, $q_{1/2}''(x)$ 和 $m''(x)$ 近似, 在本文中设为相同, 且正态分布 $\sigma^2(x) f(q_{1/2}(x)|x)^2$ 应该替代为 $\phi(\Phi^{-1}(1/2))^2 = (2\pi)^{-1}$, 因此, h_{mean} 和 $h_{1/2}$ 的关系变为:

$$\left(\frac{h_{mean}}{h_{1/2}}\right)^5 = \frac{2}{\pi} \quad (3.15)$$

由以上过程总结条件分位数的自动窗宽选择步骤如下:

- (1) 使用已有的方法选择 h_{mean} , 例如 Ruppert 等 (1995);
- (2) 使用 $h_\tau = h_{mean} \left\{ \frac{\tau(1-\tau)}{\phi(\Phi^{-1}(\tau))^2} \right\}^{1/5}$, 从 h_{mean} 得到所有其他的窗宽 h_τ 。

为了更加清晰的表示窗宽是怎么随分位数的不同而变化的, 记

$$b(\tau) = \left\{ \frac{\tau(1-\tau)}{\phi(\Phi^{-1}(\tau))^2} \right\}^{1/5} \quad (3.16)$$

其中 $b(\tau)$ 随 τ 的变化而变化。可以发现，当 τ 趋于 $\frac{1}{2}$ 时变化很小；当 τ 接近于 0 或 1 时变化明显（事实上， $\lim_{\tau \rightarrow 0} b(\tau) = \lim_{\tau \rightarrow 1} b(\tau) = \infty$ ）。

3.4 参数估计的渐近性质

在这节中，重点研究分层非参数混合效应模型中关于固定效应估计量 $\sqrt{n \det(\mathbf{H})}(\hat{\gamma}_\tau(x) - \gamma_\tau(x))$ 的渐近性质。将模型 (3.4) 写成一般形式：

$$Y_{ij} = \mathbf{A}_{fj} \boldsymbol{\gamma}(x) + \mathbf{A}_{rj} \mathbf{U}_i + \varepsilon_{ij}$$

其中 $\mathbf{A}_{fj} = \tilde{\mathbf{X}}_{ij} \mathbf{W}_i$ ， $\mathbf{A}_{rj} = \tilde{\mathbf{X}}_{ij} \mathbf{Z}_i$ 。则 (3.8) 可以写为：

$$\hat{\gamma}_\tau(x) = \arg \min_{\boldsymbol{\gamma}} \sum_{i=1}^n \sum_{j=1}^{n_i} \rho_\tau(Y_{ij} - \mathbf{A}_{fj} \boldsymbol{\gamma}(x) - \xi_{ij}(\tau)) K_{\mathbf{H}}\{\mathbf{X}_{ij} - \mathbf{x}\} \quad (3.17)$$

为了方便后面的证明，先做出以下假设。

假设 1 令 $\{Y_{ij}\}(i=1, \dots, n; j=1, \dots, n_i)$ 是一系列随机变量且相互独立，有绝对连续的分布函数 $F_{ij}(\cdot)$ ，分布函数也具有绝对连续的、有限的和正的一阶导，对于所有的 s ，有 $0 < F_{ij} < 1$ ，且二阶偏导在 $q_\tau(Y_{ij})$ ， $\tau \in (0,1)$ 的领域内是有界的。

假设 2 设 $\mathbf{e}_k = (0, \dots, 0, 1, 0, \dots, 0)^T$ ， \mathbf{e}_k 是 $f \times 1$ 维向量，1 为第 k 个元素，其他元素都为 0，当 $n \rightarrow \infty$ 时，

$$\max_{1 \leq k \leq n} \sum_{i=1}^f |\mathbf{A}_{ij} \mathbf{e}_k| = O(1)。$$

假设 3 $\frac{1}{n} \sum_{i=1}^n |\mathbf{A}_{ij} \mathbf{e}_k|^4 = O(1), n \rightarrow \infty$ 。

假设 4 核函数 $K(\cdot)$ 是 d 维的有界密度函数，且它在定义域内有紧支撑 \mathbf{C}^d ，使得 $\int K(u) du = 1$ ， $\int u K(u) du = 0$ ， $\int uu^T K(u) du > 0$ 。

假设 5 当 $n \rightarrow \infty$ 时，核函数的窗宽 \mathbf{H} 满足 $\det(\mathbf{H}) \rightarrow 0$ 和 $n \det(\mathbf{H}) \rightarrow \infty$ 。

以上假设部分及下面证明中用到的部分公式来自 Tian (2009)，假设 1 是分位数回归的经典假设，假设 2 和假设 3 是估计方法的标准，假设 4 和假设 5 保证了窗宽的有效性以及核函数的正则性、对称性、有界性和相合性，同时假设 5 给出了窗宽和样本量的关系。

下面给出两个引理, 这两个引理表明了分位数回归估计在上述假设下的收敛性。更多引理的具体说明可以参考田茂再 (2019)。令

$$L(\mathbf{b}_n) = -\frac{1}{\sqrt{n \det(\mathbf{H})}} \sum_{i=1}^n \sum_{j=1}^{n_i} \psi_{\tau}(Y_{ij} - (n \det(\mathbf{H}))^{-1/2} \mathbf{A}_{\beta} \mathbf{b}_n - F_{ij}^{-1}(\tau)) K_{\mathbf{H}}(\mathbf{X}_{ij} - \mathbf{x}) \mathbf{A}_{\beta}^T,$$

引理 1 根据假设 1 ~ 假设 5, 有

$$E(L(\mathbf{b}_n)) = \left\{ \frac{1}{\sqrt{n \det(\mathbf{H})}} \sum_{i=1}^n \sum_{j=1}^{n_i} F_{ij}'(F_{ij}^{-1}(\tau)) \mathbf{A}_{\beta}^T K_{\mathbf{H}}(\mathbf{X}_{ij} - \mathbf{x}) \mathbf{A}_{\beta} \right\} \mathbf{b}_n + o(1)$$

对于任何随机向量使得 $\|\mathbf{b}_n\|_{\infty} = O_p(1)$ 。

证明: 根据 $L(\mathbf{b}_n)$ 的表达式及期望的定义得

$$\begin{aligned} E(L(\mathbf{b}_n)) &= -\frac{1}{\sqrt{n \det(\mathbf{H})}} \sum_{i=1}^n \sum_{j=1}^{n_i} \int_{-\infty}^{\infty} \psi_{\tau}(s - \frac{1}{\sqrt{n \det(\mathbf{H})}} \mathbf{A}_{\beta} \mathbf{b}_n - F_{ij}^{-1}(\tau)) K_{\mathbf{H}}(\mathbf{X}_{ij} - \mathbf{x}) \mathbf{A}_{\beta}^T dF_{ij}(s) \\ &= -\frac{1}{\sqrt{n \det(\mathbf{H})}} \sum_{i=1}^n \sum_{j=1}^{n_i} \int_{-\infty}^{\infty} \psi_{\tau}(v) K_{\mathbf{H}}(\mathbf{X}_{ij} - \mathbf{x}) \mathbf{A}_{\beta}^T dF_{ij}(v + \frac{1}{\sqrt{n \det(\mathbf{H})}} \mathbf{A}_{\beta} \mathbf{b}_n + F_{ij}^{-1}(\tau)) \\ &= -\frac{1}{\sqrt{n \det(\mathbf{H})}} \sum_{i=1}^n \sum_{j=1}^{n_i} \left(\int_{-\infty}^0 (\tau - 1) dF_{ij}(v + \frac{1}{\sqrt{n \det(\mathbf{H})}} \mathbf{A}_{\beta} \mathbf{b}_n + F_{ij}^{-1}(\tau)) \right. \\ &\quad \left. + \int_0^{\infty} \tau dF_{ij}(v + \frac{1}{\sqrt{n \det(\mathbf{H})}} \mathbf{A}_{\beta} \mathbf{b}_n + F_{ij}^{-1}(\tau)) \right) K_{\mathbf{H}}(\mathbf{X}_{ij} - \mathbf{x}) \mathbf{A}_{\beta}^T \\ &= -\frac{1}{\sqrt{n \det(\mathbf{H})}} \sum_{i=1}^n \sum_{j=1}^{n_i} \left\{ (\tau - 1) F_{ij} \left(-\frac{1}{\sqrt{n \det(\mathbf{H})}} \mathbf{A}_{\beta} \mathbf{b}_n + F_{ij}^{-1}(\tau) \right) \right. \\ &\quad \left. + \tau \left[1 - F_{ij} \left(\frac{1}{\sqrt{n \det(\mathbf{H})}} \mathbf{A}_{\beta} \mathbf{b}_n + F_{ij}^{-1}(\tau) \right) \right] \right\} K_{\mathbf{H}}(\mathbf{X}_{ij} - \mathbf{x}) \mathbf{A}_{\beta}^T \\ &= -\frac{1}{\sqrt{n \det(\mathbf{H})}} \sum_{i=1}^n \sum_{j=1}^{n_i} \left\{ \tau - F_{ij} \left(\frac{1}{\sqrt{n \det(\mathbf{H})}} \mathbf{A}_{\beta} \mathbf{b}_n + F_{ij}^{-1}(\tau) \right) \right\} K_{\mathbf{H}}(\mathbf{X}_{ij} - \mathbf{x}) \mathbf{A}_{\beta}^T \\ &= -\frac{1}{\sqrt{n \det(\mathbf{H})}} \sum_{i=1}^n \sum_{j=1}^{n_i} \left\{ \tau - F_{ij}(F_{ij}^{-1}(\tau)) - F_{ij}'(F_{ij}^{-1}(\tau)) \frac{1}{\sqrt{n \det(\mathbf{H})}} \mathbf{A}_{\beta} \mathbf{b}_n \right. \\ &\quad \left. - \frac{1}{2} \mathbf{b}_n^T \mathbf{A}_{\beta}^T \mathbf{A}_{\beta} \mathbf{b}_n F_{ij}''(F_{ij}^{-1}(\tau)) + \theta \frac{1}{\sqrt{n \det(\mathbf{H})}} \mathbf{A}_{\beta} \mathbf{b}_n \right\} K_{\mathbf{H}}(\mathbf{X}_{ij} - \mathbf{x}) \mathbf{A}_{\beta}^T \\ &= \left\{ \frac{1}{\sqrt{n \det(\mathbf{H})}} \sum_{i=1}^n \sum_{j=1}^{n_i} F_{ij}'(F_{ij}^{-1}(\tau)) \mathbf{A}_{\beta}^T K_{\mathbf{H}}(\mathbf{X}_{ij} - \mathbf{x}) \mathbf{A}_{\beta} \right\} \mathbf{b}_n \\ &\quad + \frac{1}{\sqrt{(n \det(\mathbf{H}))^3}} \sum_{i=1}^n \sum_{j=1}^{n_i} \left\{ F_{ij}''(F_{ij}^{-1}(\tau)) + \theta \frac{1}{\sqrt{n \det(\mathbf{H})}} \mathbf{A}_{\beta} \mathbf{b}_n \right\} \frac{1}{2} \mathbf{b}_n^T \mathbf{A}_{\beta}^T \mathbf{A}_{\beta} \mathbf{b}_n \\ &\quad \cdot K_{\mathbf{H}}(\mathbf{X}_{ij} - \mathbf{x}) \mathbf{A}_{\beta}^T \\ &= \Delta_n \mathbf{b}_n + \nu_n. \end{aligned}$$

其中 $\theta \in [0, 1]$ 。接着要证明, 对于某个固定的 $C > 0$, 有 $\sup_{\|\mathbf{b}_n\|_{\infty} \leq C} \|\nu_n\|_{\infty} = o(1)$ 。根据假设

1, 有 $|F_{ij}''(\cdot)| = O(1)$ 。再由假设 2、假设 3 和假设 5 可以得到, 对于某个固定的 $C > 0$, 当 $n \rightarrow \infty$ 时,

$$\sup_{\|\mathbf{b}_n\| \leq C} \|V_n\|_\infty \leq O\left(\frac{1}{\sqrt{n \det(\mathbf{H})}}\right) \|\mathbf{b}_n\|_\infty^2.$$

证明完毕。

引理 2 假设 F_{ij} 的二阶偏导在 $F_{ij}^{-1}(\tau)$ 的领域内有界, 则当 $n \rightarrow \infty$ 时, 有

$$\sup_{\|\mathbf{b}_n\| \leq C} \|L(\mathbf{b}_n) - L(\mathbf{0})\|_\infty = EL(\mathbf{b}_n) + O_p(1)$$

对于任何常数 C 和任何随机向量序列使得 $\|\mathbf{b}_n\|_\infty = O_p(1)$ 。

证明: 记 $\mathbf{U}, \mathbf{V} \in \mathbb{R}^f$, 如果 $\mathbf{U}^T \mathbf{e}_j \leq \mathbf{V}^T \mathbf{e}_j$, 则 $\mathbf{U} \leq \mathbf{V}$, 其中 $\mathbf{e}_j = (0, \dots, 1, \dots, 0)^T, j = 1, \dots, f$ 。

令 $L^*(\mathbf{b}_n) = L(\mathbf{b}_n) - EL(\mathbf{b}_n) - L(\mathbf{0})$, 由引理 1 可以得到 $EL(\mathbf{0}) = 0$ 。对于任何 $\mathbf{U}, \mathbf{V} \in \mathbb{R}^f$,

$\|\mathbf{U}\|_\infty = O(1), \|\mathbf{V}\|_\infty = O(1)$ 及 $\mathbf{U} \leq \mathbf{V}$, 有

$$\begin{aligned} & L^*(\mathbf{U}) - L^*(\mathbf{V}) \\ &= \left\{ -\frac{1}{\sqrt{n \det(\mathbf{H})}} \sum_{i=1}^n \sum_{j=1}^{n_i} \left[\psi_\tau(Y_{ij} - \frac{1}{\sqrt{n \det(\mathbf{H})}} \mathbf{A}_{ij} \mathbf{U} - F_{ij}^{-1}(\tau)) \right. \right. \\ & \quad \left. \left. - \psi_\tau(Y_{ij} - \frac{1}{\sqrt{n \det(\mathbf{H})}} \mathbf{A}_{ij} \mathbf{V} - F_{ij}^{-1}(\tau)) \right] K_{\mathbf{H}}(\mathbf{X}_{ij} - \mathbf{x}) \mathbf{A}_{ij}^T \right\} \\ & - \left\{ \frac{1}{\sqrt{n \det(\mathbf{H})}} \sum_{i=1}^n \sum_{j=1}^{n_i} \left[F_{ij} \left(\frac{1}{\sqrt{n \det(\mathbf{H})}} \mathbf{A}_{ij} \mathbf{U} + F_{ij}^{-1}(\tau) \right) \right. \right. \\ & \quad \left. \left. - F_{ij} \left(\frac{1}{\sqrt{n \det(\mathbf{H})}} \mathbf{A}_{ij} \mathbf{V} + F_{ij}^{-1}(\tau) \right) \right] K_{\mathbf{H}}(\mathbf{X}_{ij} - \mathbf{x}) \mathbf{A}_{ij}^T \right\} \end{aligned}$$

上面的过程用到了引理 1 的证明过程。在假设 2 和假设 3 下, 可以直接证明

$E \|L^*(\mathbf{U}) - L^*(\mathbf{V})\|_\infty^4 \leq O\left(\frac{1}{n \det(\mathbf{H})}\right) \|\mathbf{U} - \mathbf{V}\|_\infty^8$ 。根据施瓦茨不等式和切比雪夫不等式, 对于

$\mathbf{U}, \mathbf{V}, \mathbf{W} \in \mathbb{R}^f$, 有

$$\begin{aligned}
& P\left\{\|L^*(\mathbf{W}) - L^*(\mathbf{V})\|_\infty \geq \lambda, \|L^*(\mathbf{V}) - L^*(\mathbf{U})\|_\infty \geq \lambda\right\} \\
& \leq \sqrt{P\left\{\|L^*(\mathbf{W}) - L^*(\mathbf{V})\|_\infty \geq \lambda\right\}} \cdot \sqrt{P\left\{\|L^*(\mathbf{V}) - L^*(\mathbf{U})\|_\infty \geq \lambda\right\}} \\
& \leq \sqrt{\frac{1}{\lambda^4} E\|L^*(\mathbf{W}) - L^*(\mathbf{V})\|_\infty^4} \cdot \sqrt{\frac{1}{\lambda^4} E\|L^*(\mathbf{V}) - L^*(\mathbf{U})\|_\infty^4} \\
& \leq \frac{1}{\lambda^2} O\left(\frac{1}{\sqrt{n \det(\mathbf{H})}}\right) \|\mathbf{W} - \mathbf{V}\|_\infty^4 \cdot \frac{1}{\lambda^2} O\left(\frac{1}{\sqrt{n \det(H)}}\right) \|\mathbf{V} - \mathbf{U}\|_\infty^4 \\
& \leq \frac{1}{\lambda^4} O\left(\frac{1}{n \det(\mathbf{H})}\right) \|\mathbf{W} - \mathbf{V}\|_\infty^4.
\end{aligned}$$

证明完毕。

在上述假设下，可以得到分位回归估计的渐近性质。通过以下定理，可以证明分层非参数混合效应模型的分位数回归估计渐近服从于多元正态分布。

定理 1 在假设 1~假设 5 下，

$$\begin{aligned}
& \sqrt{n \det(\mathbf{H})}(\hat{\gamma}(x) - \gamma(x)) \\
& = \frac{1}{\sqrt{n \det(\mathbf{H})}} \Delta_n^{-1} \left\{ \sum_{i=1}^n \sum_{j=1}^{n_i} \psi_\tau(Y_{ij} - F_{ij}^{-1}(\tau)) K_{\mathbf{H}}(\mathbf{X}_{ij} - \mathbf{x}) \mathbf{A}_{ij}^T \right\} + o_p(1)
\end{aligned}$$

其中， $\Delta_n = \frac{1}{\sqrt{n \det(\mathbf{H})}} \sum_{i=1}^n \sum_{j=1}^{n_i} F_{ij}'(F_{ij}^{-1}(\tau)) \mathbf{A}_{ij}^T K_{\mathbf{H}}(\mathbf{X}_{ij} - \mathbf{x}) \mathbf{A}_{ij}$ 。

证明：由引理 2 可知，对于任何 $\|\mathbf{b}_n\|_\infty = O_p(1)$ ，有

$$\begin{aligned}
& \left\{ -\frac{1}{\sqrt{n \det(\mathbf{H})}} \sum_{i=1}^n \sum_{j=1}^{n_i} \psi_\tau(Y_{ij} - \frac{1}{\sqrt{n \det(\mathbf{H})}} \mathbf{A}_{ij} - F_{ij}^{-1}(\tau)) K_{\mathbf{H}}(\mathbf{X}_{ij} - \mathbf{x}) \mathbf{A}_{ij}^T \right\} \\
& - \left\{ -\frac{1}{\sqrt{n \det(\mathbf{H})}} \sum_{i=1}^n \sum_{j=1}^{n_i} \psi_\tau(Y_{ij} - F_{ij}^{-1}(\tau)) K_{\mathbf{H}}(\mathbf{X}_{ij} - \mathbf{x}) \mathbf{A}_{ij}^T \right\} \\
& - \Delta_n \mathbf{b}_n \\
& = o_p(1)
\end{aligned} \tag{3.18}$$

依据 Ruppert 和 Carroll (1980) 中的引理 A.2，对 (3.17) 式求一阶偏导得

$$\begin{aligned}
& \frac{1}{\sqrt{n \det(\mathbf{H})}} \nabla \left\{ \sum_{i=1}^n \sum_{j=1}^{n_i} \rho_\tau(Y_{ij} - \mathbf{A}_{ij} \gamma(x) - \xi_{ij}(\tau)) K_{\mathbf{H}}\{(\mathbf{X}_{ij} - \mathbf{x})\} \right\} \\
& = \frac{1}{\sqrt{n \det(\mathbf{H})}} \left\{ \sum_{i=1}^n \sum_{j=1}^{n_i} \psi_\tau(Y_{ij} - \mathbf{A}_{ij} \gamma(x) - \xi_{ij}(\tau)) K_{\mathbf{H}}\{(\mathbf{X}_{ij} - \mathbf{x})\} \mathbf{A}_{ij}^T \right\} \\
& = O_p(1)
\end{aligned} \tag{3.19}$$

所以，结合 (3.6) 式可得

$$\begin{aligned} & \frac{1}{\sqrt{n \det(\mathbf{H})}} \left\{ \sum_{i=1}^n \sum_{j=1}^{n_i} \psi_{\tau}(Y_{ij} - \mathbf{A}_{ij}(\mathbf{b}_n - \gamma_{\tau}(x)) - F_{ij}^{-1}(\tau)) K_{\mathbf{H}}(\mathbf{X}_{ij} - \mathbf{x}) \mathbf{A}_{ij}^T \right\} \\ & = O_p(1) \end{aligned}$$

取 $\mathbf{b}_n = \sqrt{n \det(\mathbf{H})}(\hat{\gamma}_{\tau}(x) - \gamma_{\tau}(x))$ ，有

$$\begin{aligned} & \sqrt{n \det(\mathbf{H})}(\hat{\gamma}(x) - \gamma(x)) \\ & = \frac{1}{\sqrt{n \det(\mathbf{H})}} \Delta_n^{-1} \left\{ \sum_{i=1}^n \sum_{j=1}^{n_i} \psi_{\tau}(Y_{ij} - F_{ij}^{-1}(\tau)) K_{\mathbf{H}}(\mathbf{X}_{ij} - \mathbf{x}) \mathbf{A}_{ij}^T \right\} + o_p(1) \end{aligned} \quad (3.20)$$

定理 1 得证。

定理 2 假设 Δ_n 和 Θ_n 是正定矩阵，在假设 1~假设 5 下， $\sqrt{n \det(\mathbf{H})}(\hat{\gamma}(x) - \gamma(x))$ 收敛于一个多元正态分布，具体表达为

$$\sqrt{n \det(\mathbf{H})}(\hat{\gamma}(x) - \gamma(x)) \rightarrow N(\mathbf{0}, \tau(1-\tau)\Delta_n^{-1}\Theta_n\Delta_n^{-1}).$$

其中， $\Theta_n = \frac{1}{n \det(\mathbf{H})} \sum_{i=1}^n \sum_{j=1}^{n_i} \mathbf{A}_{ij}^T K_{\mathbf{H}}^2(\mathbf{X}_{ij} - \mathbf{x}) \mathbf{A}_{ij}$ 。

证明：有了前面的假设、引理和计算过程，估计的渐近分布证明过程是比较容易的。根据 (3.6) 式的推导过程可以得到

$$\text{var}(\psi_{\tau}(Y_{ij} - F_{ij}^{-1}(\tau))) = \tau(1-\tau)$$

$$\text{记 } \Gamma_n = \frac{1}{\sqrt{n \det(\mathbf{H})}} \left\{ \sum_{i=1}^n \sum_{j=1}^{n_i} \psi_{\tau}(Y_{ij} - F_{ij}^{-1}(\tau)) K(\mathbf{X}_{ij} - \mathbf{x}) \mathbf{A}_{ij}^T \right\},$$

由方差定义计算得

$$\begin{aligned} & \text{var}(\Gamma_n) \\ & = \text{var} \left\{ \frac{1}{\sqrt{n \det(\mathbf{H})}} \left[\sum_{i=1}^n \sum_{j=1}^{n_i} \psi_{\tau}(Y_{ij} - F_{ij}^{-1}(\tau)) K_{\mathbf{H}}(\mathbf{X}_{ij} - \mathbf{x}) \mathbf{A}_{ij}^T \right] \right\} \\ & = \frac{1}{n \det(\mathbf{h})} \tau(1-\tau) \left\{ \sum_{i=1}^n \sum_{j=1}^{n_i} \mathbf{A}_{ij}^T K_{\mathbf{H}}^2(\mathbf{X}_{ij} - \mathbf{x}) \mathbf{A}_{ij} \right\} \\ & = \tau(1-\tau)\Theta_n \end{aligned} \quad (3.21)$$

在假设 4 和假设 5 的条件下可以看出， $\sqrt{n \det(\mathbf{H})}(\hat{\gamma}(x) - \gamma(x))$ 满足林德伯格条件，因此，它是一个渐近多元正态分布，协方差矩阵通过 (3.20) 式和 (3.21) 式得到。

定理 2 得证。

4 数值模拟和实例分析

4.1 Monte Carlo 模拟

基于第三章的讨论, 本节将在 R 语言环境下, 采用有限随机样本模拟的方法验证上文中提到的分层非参数混合效应模型分位回归估计方法的稳定性。在前面的分层非参数混合效应模型的分位回归中, 没有假设随机误差项的分布, 所以在此假设误差项服从不同的分布来验证模型和方法的有效性和稳定性。在模拟过程中, 随机误差项的分布依次设为标准正态分布, beta 分布, t 分布, 卡方分布, 指数分布以及泊松分布, 分别模拟出模型参数在不同的误差项分布下的估计结果。在模拟中, 会涉及到核函数和窗宽的选择, 方便起见, 模拟中核函数选择高斯核函数, 窗宽使用最优理论窗宽 $h = n^{-1/5}$ 。首先给出几种随机误差项分布的概率密度曲线图, 见图 4.1。

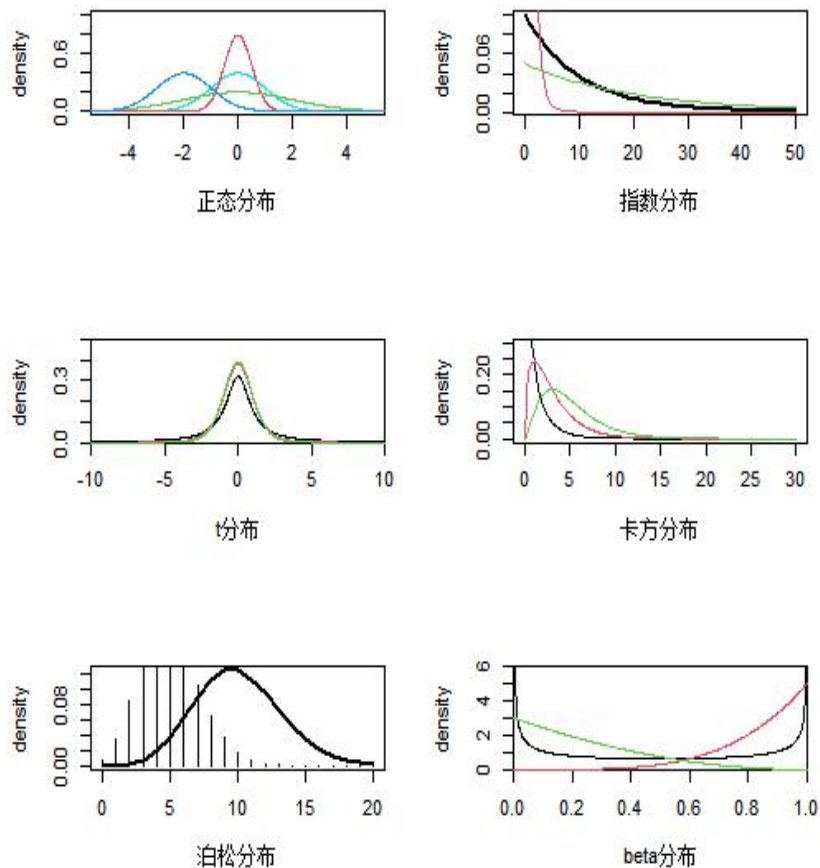


图 4.1 随机误差项概率密度曲线图

从图 4.1 可以看出，不同分布下的数据分布形态有很大的差异，这些分布几乎包含了所有数据可以分布的类型，所以，模拟比较这些分布下的估计结果能够体现模型和方法的实用性和稳定性。

接下来给出随机误差项在不同分布下的估计结果。设定迭代次数为 5, 10, 100, 500, 1000。取迭代算法估计的初始值 $\sigma_0^2=0.2$ ，固定效应初始值 $\hat{\gamma}_0=[1,1,1,1]^T$ ，协方差初始值 $\hat{T}_{00_0} = 0.2$ ， $\hat{T}_{01_0} = 0$ ， $\hat{T}_{10_0} = 0$ ， $\hat{T}_{11_0} = 0.2$ 。

表 4.1 误差项为正态分布的参数估计结果

迭代次数	σ_i^2	T_{00}	T_{11}	γ_{00}	γ_{01}	γ_{10}	γ_{11}
5	1.003	1.255	1.387	0.716	0.698	0.993	0.955
10	0.999	0.466	1.275	0.411	0.367	0.989	0.908
100	0.997	0.006	0.002	-0.053	-0.005	0.525	-0.004
500	0.991	0.001	0.001	-0.051	-0.004	0.525	-0.005
1000	0.991	0.001	0.001	-0.051	-0.004	0.525	-0.005

表 4.1 是误差项服从标准正态分布情况下的模型参数估计结果。可以看出，在迭代次数不断增多的过程中，估计值有明显的收敛趋势，在迭代 500 次后，参数的估计收敛到一个固定值，对应的参数估计结果为 $\sigma_i^2 = 0.991$ ，协方差估计值为 $[\hat{T}_{00}, \hat{T}_{01}, \hat{T}_{10}, \hat{T}_{11}] = [0.001, -0.002, -0.002, 0.001]$ ，相对应的固定效应的估计值为 $[\hat{\gamma}_{00}, \hat{\gamma}_{01}, \hat{\gamma}_{10}, \hat{\gamma}_{11}] = [-0.051, -0.004, 0.525, -0.005]$ 。

表 4.2 误差项为 exp(1)的参数估计结果

迭代次数	σ_i^2	T_{00}	T_{11}	γ_{00}	γ_{01}	γ_{10}	γ_{11}
5	1.088	0.336	1.505	1.050	0.462	0.958	0.956
10	1.064	0.086	1.361	1.071	0.000	0.920	0.914
100	1.053	0.014	0.002	1.000	-0.063	0.517	0.014
500	1.055	0.013	0.001	0.998	-0.064	0.517	0.014
1000	1.055	0.013	0.001	0.998	-0.064	0.517	0.014

表 4.2 是误差项服从参数为 1 的指数分布下的模型参数估计结果。可以看出，在迭代次数不断增多的过程中，估计值有明显的收敛趋势，在迭代 500 次后，参数的估计收敛到一个固定值，对应的参数估计结果为 $\sigma_i^2 = 1.055$ ，协方差估计值为 $[\hat{T}_{00}, \hat{T}_{01}, \hat{T}_{10}, \hat{T}_{11}] = [0.001, 0.008, 0.008, 0.001]$ ，相对应的固定效应的估计值为 $[\hat{\gamma}_{00}, \hat{\gamma}_{01}, \hat{\gamma}_{10}, \hat{\gamma}_{11}] = [0.998, -0.064, 0.517, 0.014]$ 。

表 4.3 误差项为 t(5)分布的参数估计结果

迭代次数	σ_i^2	T_{00}	T_{11}	γ_{00}	γ_{01}	γ_{10}	γ_{11}
5	1.509	0.440	0.915	0.516	0.481	0.976	0.942
10	1.465	0.086	0.780	0.043	0.030	0.942	0.870
100	1.463	0.006	0.001	0.025	-0.024	0.501	0.015
500	1.471	0.001	0.000	0.026	-0.025	0.501	0.015
1000	1.471	0.001	0.000	0.026	-0.025	0.501	0.014

表 4.3 是误差项服从参数为 5 的 t 分布下的模型参数估计结果。可以看出，在迭代次数不断增多的过程中，估计值有明显的收敛趋势，在迭代 500 次后，参数的估计收敛到一个固定值，对应的参数估计结果为 $\sigma_i^2 = 1.471$ ，协方差估计值为 $[\hat{T}_{00}, \hat{T}_{01}, \hat{T}_{10}, \hat{T}_{11}] = [0.001, 0.006, 0.006, 0.000]$ ，相对应的固定效应的估计值为 $[\gamma_{00}, \gamma_{01}, \gamma_{10}, \gamma_{11}] = [0.026, -0.025, 0.501, 0.014]$ 。

表 4.4 误差项为卡方分布的参数估计结果

迭代次数	σ_i^2	T_{00}	T_{11}	γ_{00}	γ_{01}	γ_{10}	γ_{11}
5	7.587	9.325	0.880	2.001	0.624	1.030	0.640
10	7.436	4.155	0.455	2.911	0.214	0.995	0.211
100	7.630	0.054	0.014	4.779	0.065	0.445	0.022
500	7.687	0.015	0.010	4.779	0.064	0.445	0.022
1000	7.694	0.008	0.009	4.779	0.064	0.445	0.022

表 4.4 是误差项服从参数为 5 的卡方分布下的模型参数估计结果。可以看出，在迭代次数不断增多的过程中，估计值有明显的收敛趋势，在迭代 500 次后，参数的估计收敛到一个固定值，对应的参数估计结果为 $\sigma_i^2 = 7.694$ ，协方差估计值为 $[\hat{T}_{00}, \hat{T}_{01}, \hat{T}_{10}, \hat{T}_{11}] = [0.008, 0.053, 0.053, 0.009]$ ，相对应的固定效应的估计值为 $[\hat{\gamma}_{00}, \hat{\gamma}_{01}, \hat{\gamma}_{10}, \hat{\gamma}_{11}] = [4.779, 0.064, 0.445, 0.022]$ 。

表 4.5 误差项为泊松分布的参数估计结果

迭代次数	σ_i^2	T_{00}	T_{11}	γ_{00}	γ_{01}	γ_{10}	γ_{11}
5	5.022	10.318	1.125	1.839	0.695	1.004	0.838
10	5.008	6.067	0.659	2.558	0.385	1.010	0.589
100	4.782	0.040	0.002	4.595	-0.098	0.560	0.008
500	4.806	0.015	0.001	4.960	-0.098	0.560	0.007
1000	4.811	0.011	0.001	4.960	-0.098	0.560	0.007

表 4.5 是误差项服从参数为 10 的泊松分布下的模型参数估计结果。可以看出，在迭代次数不断增多的过程中，估计值有明显的收敛趋势，在迭代 500 次后，参数的估计收敛到一个固定值，对应的参数估计结果为 $\sigma_i^2 = 4.811$ ，协方差估计值为 $[\hat{T}_{00}, \hat{T}_{01}, \hat{T}_{10}, \hat{T}_{11}] = [0.011, -0.001, -0.001, 0.001]$ ，相应的固定效应的估计值为 $[\hat{\gamma}_{00}, \hat{\gamma}_{01}, \hat{\gamma}_{10}, \hat{\gamma}_{11}] = [4.960, -0.098, 0.560, 0.007]$ 。

表 4.6 误差项为 beta 分布的参数估计结果

迭代次数	σ_i^2	T_{00}	T_{11}	γ_{00}	γ_{01}	γ_{10}	γ_{11}
5	0.050	0.748	1.906	0.947	0.879	0.995	0.990
10	0.050	0.669	1.901	0.937	0.847	0.995	0.989
100	0.048	0.001	1.829	0.507	0.013	0.987	0.970
500	0.048	0.001	1.477	0.505	0.011	0.939	0.872
1000	0.048	0.001	1.475	0.505	0.011	0.870	0.731

表 4.6 是误差项服从 beta(2, 5)下的模型参数估计结果。可以看出，在迭代

次数不断增多的过程中，估计值有明显的收敛趋势，在迭代 500 次后，参数的估计收敛到一个固定值，对应的参数估计结果为 $\hat{\sigma}_i^2 = 0.048$ ，协方差估计值为 $[\hat{T}_{00}, \hat{T}_{01}, \hat{T}_{10}, \hat{T}_{11}] = [0.001, 0.380, 0.380, 1.475]$ ，相应的固定效应的估计值为 $[\hat{\gamma}_{00}, \hat{\gamma}_{01}, \hat{\gamma}_{10}, \hat{\gamma}_{11}] = [0.505, 0.011, 0.870, 0.731]$ 。

根据模拟计算可以很直观的看出参数估计结果依据迭代次数 5, 10, 100, 500, 1000 的变化趋势，从表 4.1 到表 4.6 显示，不管误差项服从连续型分布还是离散型分布，所有参数在迭代到 500 次时，虽然选取的随机数不同导致最后估计出的结果不同，但是所有参数基本都收敛到固定值，这点很好地显示了该方法在模型下的稳定性，说明本文提出的分层非参数混合效应模型和估计方法是合理的。

4.2 实例分析

4.2.1 汽车品牌数据

随着经济的发展和科技的进步，人们的衣食住行在发生着天翻地覆的变化，其中，汽车的需求量也在快速的增长，基于此，全球各种品牌的汽车陆续问世。本文使用 7 个国家 32 种汽车品牌 352 条数据，该数据来自 mtcars 数据包。利用分层非参数混合效应模型对该组数据进行分析和解释。研究各变量与油耗之间的关系，国家和汽车品牌被认为是嵌套分类变量。

对于每辆汽车，有 11 个指标，以不同的符号表示如下：

Mpg: 油耗（单位是每加仑英里，1mpg=235.2156 升/百公里），每行驶 100 公里所消耗的油量。功能越强大、更重的汽车往往消耗更多的燃油；

Cyl: 汽缸数。功率越大的汽车通常具有更多的汽缸；

Disp: 排量（单位是立方英寸）。发动机汽缸的总容积；

Hp: 总马力。汽车产生的功率的量度；

Drat: 后轴比率。描述了驱动轴的转动与车轮的转动如何对应，较高的值会降低燃油效率；

Wt: 重量；

Qsec: 汽车的速度和加速度；

Vs: 发动机缸体。表示车辆的发动机形状是“V”型还是直型；

Am: 变速箱。表示汽车的变速箱是自动（0）还是手动（1）；

Gear: 前进挡的数量。跑车往往具有更多的档位；

Carb: 化油器数量。它与更强大的发动机相关；

4.2.2 数据分析

首先对数据进行初步简单的分析，这里选取部分数据进行嵌套建模。根据排量等变量，分析油耗的情况。图 4.2 先给出不同国家耗油量变化的箱线图：

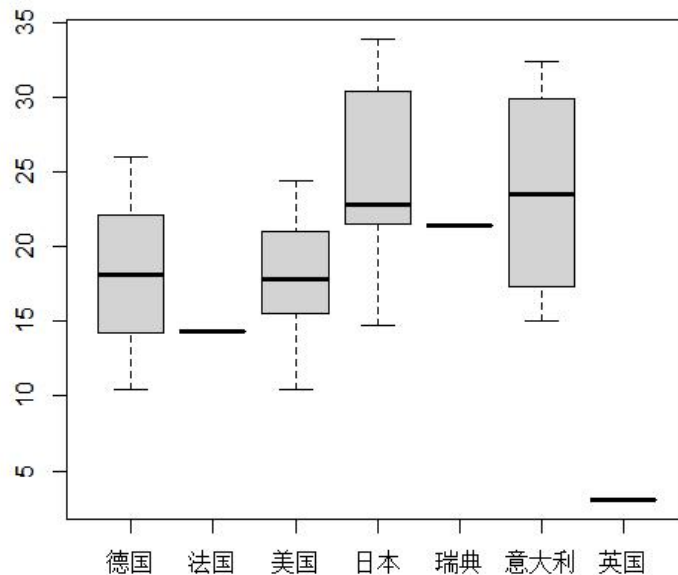


图 4.2 不同国家油耗箱线图

如图 4.2 所示，所有国家汽车的耗油量没有特别的极端值，但是可以看出，日本和意大利的耗油量比其他几个国家要高，英国的最低，德国和美国处于中间相差不大。接下来给出耗油量随不同变量的变化情况，如图 4.3 所示：

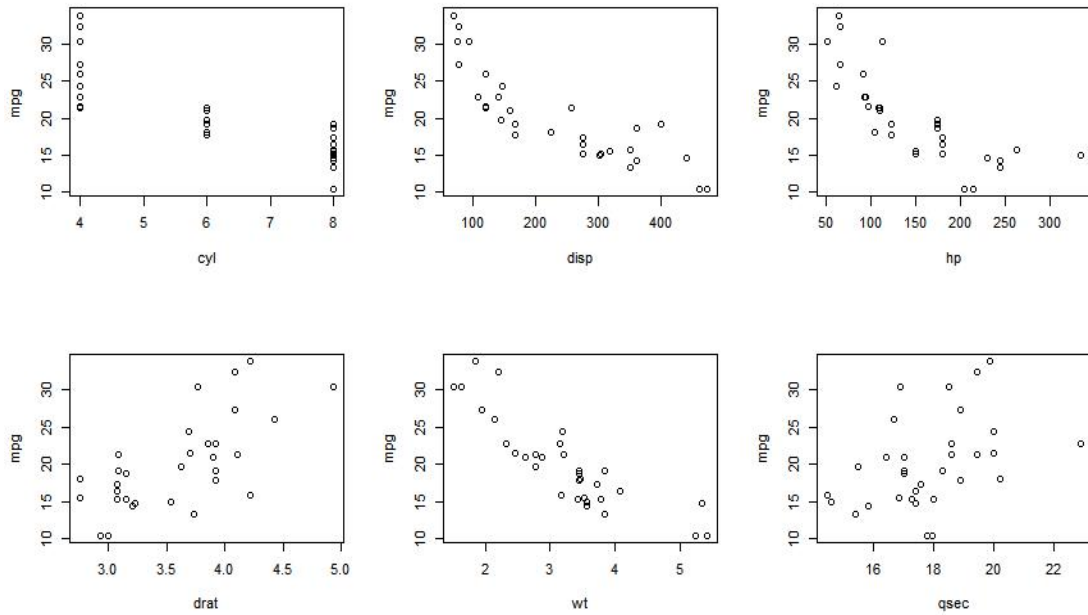


图 4.3 油耗随各种变量的变化图

在图 4.3 中，可以观察到油耗大体上随着气缸数、排量、马力、汽车重量的增大而减少，随着后轴比率和汽车的速度的增大而增多。另外，在这些图中可以看到，在大多数变量影响下其与 mpg 变量并非有较强的线性关系，且 mpg 变量的方差比较明显，这会促使我们考虑到分层非参数分位数模型，同时加入混合效应。也就是说，第一层模型被表示为 $Y_{ij} = m(X_{ij}) + \varepsilon_{ij}$ ， Y_{ij} 表示第 j 个国家第 i 个汽车品牌的油耗 mpg， X_{ij} 是指标，设 X_{ij} 的待估参数为 θ_i ，则第二层模型被表示为 $\theta_i = W_i \gamma(x) + Z_i U_i$ ， W_i 是第二层的自变量， $\gamma(x)$ 是待估的固定效应， Z_i 是随机效应， U_i 是第二层随机误差。

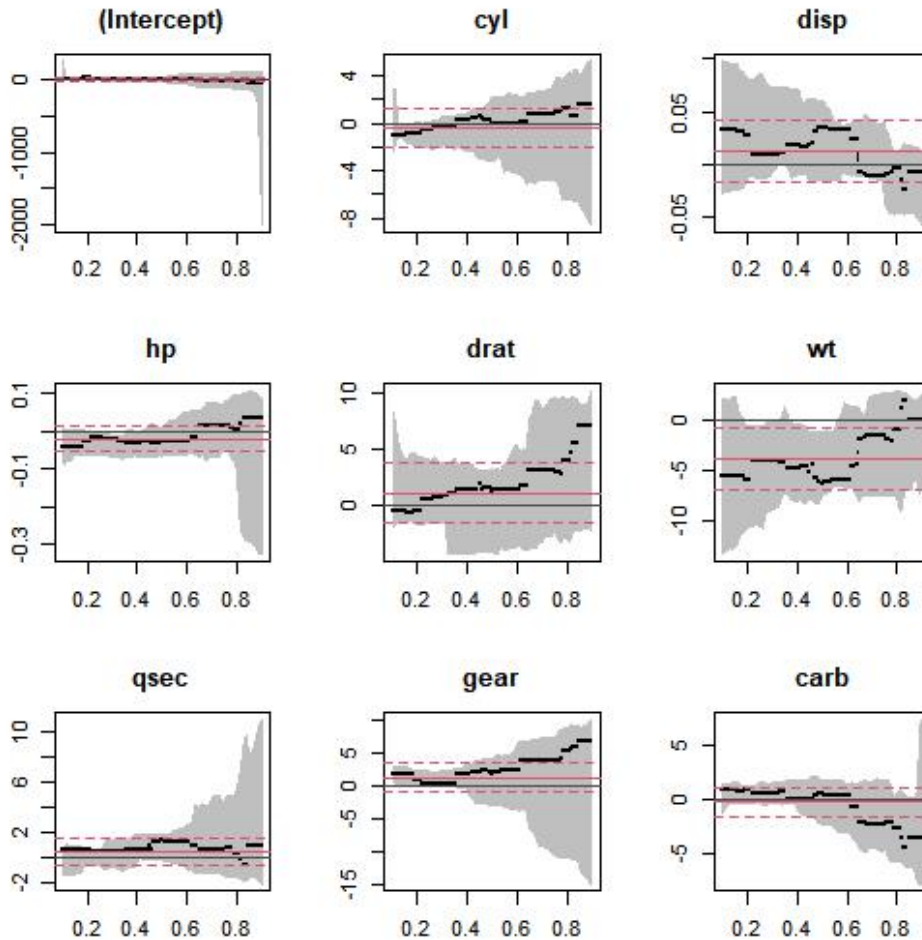


图 4.4 固定效应函数估计结果

图 4.4 展示了影响因素气缸数、排量、总马力、后轴比率、速度、前进挡数量以及化油器数量基于分层非参数混合效应模型分别在 0.1~0.9 分位数下固定效应的估计结果。可以看出，固定效应估计都是显著的，说明这几个指标在此模型下对油耗 mpg 的影响是显著的。具体来说，气缸数、排量、总马力的显著性更强，对于汽车的油耗影响较大。特别地，在通常情况下，基于普通的均值回归的结构应该接近于基于中位数的分位回归，但是在这里观察到是有差别的，这个差别可能是由线性假设所导致的。对于这个猜想，可以通过数据拟合线性模型和非线性模型，利用模型比较准则（AIC 和 BIC）来比较拟合结果。

5 总结与展望

5.1 结论

本文主要针对分层数据，并且这些数据中协变量与响应变量之间具有非线性关系或者具体关系未知的情况下，提出了分层非参数混合效应模型。研究的主要内容包括三部分，第一部分是基于分层线性模型、非参数理论和分位回归理论提出了分层非参数混合效应模型；第二部分是对模型中的参数进行估计，在估计的过程中，核函数和窗宽的选择是比较重要的内容，本文选择了高斯核函数，且从最优渐近窗宽出发得到了窗宽，估计之后给出了参数的渐近性质；第三部分是分析该模型在误差项服从不同分布下的参数估计，通过随机模拟和实例分析发现，模型和方法的稳健性很好。通过上述三部分理论研究，本文得出三点结论，第一，对于分层且协变量与响应变量关系不清晰的数据，使用分层非参数混合效应模型更加贴切，拟合得更好；第二点是通过比较误差项在连续型分布、离散型分布下参数的估计结果，在增大样本量后都达到收敛效果，说明估计该模型的方法具有稳定性；第三点是通过参数估计的渐近性证明，得出了模型参数的估计结果渐近服从于多元正态分布。通过上述结果可以看出理论上模型和估计方法比较可靠，并且通过随机模拟和实例分析结果更加证明了模型和方法的有效性。

5.2 展望

在本文的研究过程中，使用分位数回归估计固定效应不需假定随机误差项的分布，但是在估计方差和协方差矩阵时假定了误差项是服从正态分布的，接下来可以假设误差项是混合正态分布或其他分布来估计参数并证明其大样本性质；在连接第一层和第二层时，采用的是泰勒展开，在后续的研究中可以通过样条等更前沿的方法展开。在对模型参数进行估计时，采用的是基于检验函数的非参数分位回归来估计，并使用了已有的 EM 算法进行迭代，在以后可以通过结合贝叶斯估计方法对其估计。另外，在窗宽选择的研究上还有很大的空间，也是比较困难的，是后续可以重点探究的问题。

参考文献

- [1] Dempster A P, Laird N M, Rubin D B. Maximum likelihood from incomplete data via the EM algorithm [J]. Proceedings of the Royal Statistical Society, 1977, 39(1):1-22.
- [2] Dempster A P, Rubin D B, Tsutakawa R K. Estimation in covariance components model[J]. Journal of the American Statistical Association , 1981, 76(374): 341 – 353.
- [3] Fan J, Hu T C, Truong Y K. Robust nonparametric function estimation[J]. Scandinavian Journal of Statistics, 1994,21, 433–446 .
- [4] Goldstein H. Multilevel Statistical Models, volume 922[M]. New York: John Wiley & Sons,1995.
- [5] Goldstein H. Multilevel mixed linear model analysis using iterative generalized least squares[J]. Biometrika, 1986, 73(1):43-56.
- [6] Hendricks W, Koenker R. Hierarchical spine models for conditional quantiles and the demand for electricity[J]. Journal of the American Statistical Association, 1992,87(417):58-68.
- [7] Honda T. Nonparametric Estimation of a Conditional Quantile for α -Mixing Processes[J].Annals of the Institute of Statistical Mathematics, 2000, 52(3): 459-470.
- [8] Jones M C, Hall P. Mean Squared Error Properties of Kernel Estimates of Regression Quantiles[J]. Statistics & Probability Letters,1990,10(4): 283-289.
- [9] Kass R E, Steffey D. Approximate Bayesian in conditionally independent hierarchical models(parametric empirical bayes models)[J].Journal of the American Statistical Association,1989,84(407):717-726.
- [10]Koenker R, Bassett G. Regression quantiles[J]. Econometrica,1978, 46, 33 – 50. .
- [11]Koenker R. Quantile regression for longitudinal data[J]. Journal of Multivariate Analysis, 2004, 91(1):74-89.
- [12]Koenker R. Quantile Regression[M].Cambridge: Cambridge University Press,

- 2005.
- [13] Laird N M, Ware J H. Random-effects models for longitudinal data[J]. *Biometrics*, 1982,38(4):963-974.
- [14] Lee S. Efficient semiparametric estimation of partially linear quantile regression model[J]. *Econometric Theory*, 2003, 19(1):1-31.
- [15] Lejeune M G, Sarda P. Quantile regression: A nonparametric approach[J]. *Computational Statistics & Data Analysis*, 1988,6(3):229-239.
- [16] Lin H M, Willson, et al. Median filters with adaptive length[J]. *Circuits and Systems, IEEE Transactions on*, 1988, 35(6):675-690.
- [17] Lindley D V, Smith A F. Bayes estimates for the linear model[J]. *Journal of the Royal Statistical Society, Series B*, 1972,34, 1-41 .
- [18] Martins P S, Pereira P T. Does Education Reduce Wage Inequality? Quantile Regression Evidence from 16 Countries[J]. *Labour Economics*, 2004,11(3): 0-371.
- [19] Mason W M, Wong G Y, Entwisle B. Contextual analysis through the multilevel linear model[J]. *Sociological Methodology*, 1983,14:72-103.
- [20] Paul S. Nonparametric and Semiparametric Models[J]. *Publications of the American Statistical Association*, 2010, 101(475):1308-1309.
- [21] Rosenberg B. Linear regression with randomly dispersed parameters[J]. *Biometrika*, 1973,60(1): 65-72.
- [22] Samanta M . Non-parametric estimation of conditional quantiles[J]. *Statistics & Probability Letters*, 1989, 7(5):407-412.
- [23] Singer J D. Using sasproc mixed to fit multilevel models, hierarchical models, and individual growth models[J]. *Journal of Educational and Behavioral Statistics*, 1998,23(4):323-355.
- [24] Smith A F. A general Bayesian linear model[J]. *Journal of the Royal Statistical Society, Series B*, 1973,35,67-75 .
- [25] Spokoiny V, Wang W, Hardle W K. Local quantile regression[J]. *Journal of Statistical Planning & Inference*, 2013,143(7): 1109-1129.
- [26] Tao H, Palta M, Yandell B S, Newton M A. An estimation method for the

- semiparametric mixed effects model[J]. *Biometrics*,1999,55(1).
- [27] Tian M Z, Tang M L, Chan P S. Semiparametric quantile modelling of hierarchical data[J]. *Acta Mathematica Sinica*, 2009.
- [28] Willms J D, Raudenbush S W. A Longitudinal Hierarchical Linear Model for Estimating School Effects and Their Stability[J]. *Journal of Educational Measurement*, 1989, 26(3):209-232.
- [29] Yu K, Jones M C. Local linear quantile regression[J]. *Journal of the American statistical Association*, 1998,93,228–238 .
- [30] 贺建风,付永超,熊健.基于分层贝叶斯广义线性模型的小域估计方法研究[J]. *数理统计与管理*,2019,38(02):247-260.
- [31] 梁春燕. 线性混合模型中参数估计与检验问题的研究[D]. 青岛科技大学,2014.
- [32] 刘婷婷,杨联强,王学军.基于非参数分位数估计的众数回归模型[J].*应用概率统计*,2020,36(05):483-492.
- [33] 田茂再,陈歌迈.条件分位中的分层线性回归模型[J].*中国科学(A 辑:数学)*,2006(10):1103-1118.
- [34] 田茂再.现代非参数统计中的窗宽选择及应用[M].北京:科学出版社,2019:333-396.
- [35] 谢晶晶.非参数分位数回归模型估计方法的比较研究[D].厦门大学,2019.
- [36] 徐志科,平根建.非参数方法估计分位数模型的研究综述[J].*数学的实践与认识*,2014,44(01):151-156.
- [37] 杨新平,徐登国.两水平线性模型的参数估计及调整[J].*统计与决策*,2019,35(04):71-75.
- [38] 张敏,罗幼喜.非参数面板数据模型的贝叶斯分位回归方法研究[J].*统计与决策*,2020,36(19):9-14.
- [39] 张玉臣,吕宪鹏.高新技术企业创新绩效影响因素研究[J].*科研管理*,2013,34(12):58-65.
- [40] 周梦雨.基于 Guass-Seidel 型迭代算法的分层模型统计推断[D].兰州财经大学,2021.

后记

日逢今昔，岁在壬寅，三载悠悠，文末搁笔，所有相遇，一生感激。

东西南北春常在。在那个秋夏交界的乙未年，我走进了兰财的门，踏进了统计学院，这一入，已过七年。这七年，是兰财统院的栽培；这七年，经历了自我塑造到自我怀疑，再趋于平稳的过程。岁月虽清浅，时光亦潋潋。

先生之风，山高水长。入田门，为兰州关门弟子，感恩我的导师田茂再老师，田老师专业精通，治学严谨，腹笥渊博，做事认真，风趣幽默。不仅在我的学习和论文写作过程中给予非常细致地指导，在日常生活中也常常答疑解惑，也时常带领我们游览祖国的大好河山。先生于我，影响甚远，为人处世，虽不能至，也向往之。同时感谢统计学院所有老师对我的帮助。幸遇良师，谨记教诲。

春晖寸草，山高海深。祖父母疼爱，兄姊宠爱，亲人友爱。祖父母拼尽全力撑起了我的人生，他们对生活的坚强让我坚定了面对困难的勇气和底气，他们也从未入学堂却深知读书的重要，而我的步伐，随着读书的深入一步步从小村庄到乡镇县城省会乃至全国各地。往后，惟愿身体健康，平安顺遂。

萍水相逢，三生有幸。圈子很小，但有一隅，极为重要，与吴军胜先生、马素平女士、陈玉国先生、白明军先生等长辈因一份善意相识于 2015 年，自此一直被关爱照顾，对我而言，他们的关怀启迪和对晚辈的善意使我受益良多，不是亲人胜似亲人。打开一扇窗，静听花开声。

山河足重遇知己。天下快意之事莫挚友，李艳艳、颜小凤是我的张怀民，是我这几年的情绪出口，感谢在我失落时毫不顾忌的致电并听我的唠叨；感谢李肖凤对我学习之余生活的调节；感谢各位师兄师姐以及所有同学朋友们的引导和鼓励，尤其感谢梁永玉和周梦雨师姐在我学习特别是写论文过程中极大的帮助。感谢我的室友们一起学习熬夜吃火锅。人生漫长，奔赴前程。

瘟疫散尽，国泰民安。整个研究生生涯都被笼罩在疫情下，封校、线上活动都已成了常态，但何其有幸生于华夏，才使我们的生活学习相对能够正常运转，这世界那么多人，在此致敬所有医护人员和志愿者。山河锦绣，万民康健。

愿此去经年，心怀赤诚，行远自迩，踵事增华，一生向阳。兰州段家滩路 496 号，再会！

2022 年 5 月