



北京工商大学经济学博士文库

# 分位数回归模型 及其应用

罗玉波◎著

F  
ENWEISHU HUIGUI MOXING  
JIQI YINGYONG

知识产权出版社

# 分位数回归模型 及其应用

罗玉波◎著

C812

L970

知识产权出版社

## 内容提要

本书围绕分位数估计,以分位数回归模型为主线,研究了变换可加分位数回归模型的估计方法、分位数回归模型在空间数据中的应用以及如何利用鞍点逼近技术求解分位数。变换可加分位数回归模型既能够体现出非参数模型的优点,又能克服一般非参数方法的缺点。将分位数回归思想用于空间数据分析是今年研究的一个热点,我们对这方面的问题也进行了一些探索。

责任编辑:兰 涛

## 图书在版编目(CIP)数据

分位数回归模型及其应用/罗玉波著. —北京:知识产权出版社, 2009. 9

ISBN 978-7-80247-530-4

I. 分… II. 罗… III. 统计分析—研究 IV. C812

中国版本图书馆 CIP 数据核字 (2009) 第 159821 号

## 分位数回归模型及其应用

罗玉波 著

---

出版发行: 知识产权出版社

社 址: 北京市海淀区马甸南村 1 号

邮 编: 100088

网 址: <http://www.ipph.cn>

邮 箱: [bjb@cnipr.com](mailto:bjb@cnipr.com)

发行电话: 010-82000893 82000860 转 8101 传 真: 010-82000860-8325

责任编辑: 010-82000860-8325

责任编辑: [lantao@cnipr.com](mailto:lantao@cnipr.com)

印 刷: 知识产权出版社电子印制中心

经 销: 新华书店及相关销售网点

开 本: 880mm×1230mm 1/32

印 张: 5.25

版 次: 2009 年 9 月第 1 版

印 次: 2010 年 9 月第一次印刷

字 数: 112 千字

定 价: 16.00 元

ISBN 978-7-80247-530-4/C·091 (10280)

---

版权所有 侵权必究

如有印装质量问题, 本社负责调换。

# 丛书编委会

(按拼音顺序)

主任委员：杨德勇

|      |     |     |     |     |
|------|-----|-----|-----|-----|
| 委 员： | 葛红玲 | 耿莉萍 | 郭馨梅 | 洪 涛 |
|      | 胡俞越 | 季 铸 | 姜 竹 | 李宝仁 |
|      | 李书友 | 李友元 | 栗书茵 | 廖运凤 |
|      | 刘成璧 | 刘 毅 | 马若微 | 马韬光 |
|      | 王绪瑾 | 王再文 |     |     |

# 总 序

北京工商大学是北京市重点建设的多科性大学，经济学院作为北京工商大学规模最大的学院之一，拥有一支结构合理、力量雄厚的教学科研队伍，教师中享受国务院特殊津贴的专家四人，近五年来学院教师出版专著 90 余部，发表学术论文 1100 余篇，承担国家级省部级课题 60 余项，为经济理论的创新和发展、为国家的经济建设作出了较大贡献。

在当今世界处于经济危机的大环境下，经济社会的发展必须依靠青年科技人才。作为培养青年人才的高等学府，高校更是需要注重青年教师的培养，只有赢得青年教师，才能赢得高校的未来和发展。北京工商大学始终以培养青年教师为基本发展理念。近年来，经济学院引进了一大批拥有较高专业素养和研究能力的博士、博士后，使学院师资队伍职称结构、学历结构、年龄结构和学缘结构发生了巨大变化，作为学院教学、科研工作的新生力量和学术骨干，这些中青年教师正在各自的学科领域崭露头角，其中多位中青年教师步入北京市优秀教师、北京市中青年骨干教师、北京市跨世纪理论人才“百人工程”、“北京市跨世纪优秀人才工程”、北京市学科带头人的行列，在各自的专业领域内脱颖而出。

这些中青年教师都接受过系统的经济学教育，掌握了规范

的经济学研究方法，具有较新的研究理念，部分教师在其研究领域还取得了创新性的成果。为了充分展示他们的学术成果、打造高端学术平台、提升整体科研水平、推动学院学科建设可持续发展，北京工商大学经济学院特组织院内部分专家、教授组成了“北京工商大学经济学博士文库”编委会，从这些中青年教师的研究成果中精选、汇集了一批在经济研究领域具有代表性的研究成果，并从学科建设经费中拨出专项资金予以资助出版。本套丛书是在学院中青年教师博士论文、博士后出站报告的基础上修改而成的，是学院中青年教师学术研究成果的一次充分展示，因此，具有较高的学术内涵和实践价值。

“北京工商大学经济学博士文库”的出版，一方面旨在保存和交流学术研究成果，鼓励学术研究，提高学院教师的科研积极性，通过本套丛书向社会推出一批学术精品，为解决现实经济问题献计献策，发挥现代高校应有的贡献；另一方面，通过资助在学术方面具有独到见解和创新思想的科研成果，培养优秀的高水平科研人才，提高学院学术队伍的整体实力，促进北京工商大学经济类学科建设的发展。

丛书编委会

2009年8月

# 前 言

在许多的数据分析和建模过程中，回归模型可以说是应用最为广泛的统计方法之一。回归模型有着许多的运用，可以用来分析变量间的关系，或者用来进行预测，或者用来控制等。比如，假定因变量能够表示为自变量的线性函数和一个随机误差项之和，并且随机误差项的期望值为零，就可以得到最常用的简单线性回归模型。简单线性回归模型等价于假定因变量的条件期望是自变量的线性函数。多数传统的回归模型都可以归为这一类条件均值模型，能够揭示因变量条件期望与自变量的关系。我们知道，一般情况下，均值只是描述随机变量的指标之一，要充分描述一个随机变量只有了解随机变量的分布。同样，刻画因变量与自变量的关系，只有条件期望可能是不够的，一些时候我们需要了解因变量的条件分布。

正是基于以上的分析，Koenker 和 Bassett 于 20 世纪 70 年代末期提出了具有重要影响的“回归分位数”概念，其核心正是对因变量的条件分位数进行建模。我们知道随机变量分位数与分布之间存在着对应关系，了解随机变量的分位数，也就了解了随机变量的分布，通过建立自变量与因变量分位数的关系，能够全面了解自变量与因变量之间的关系。从分位数回归模型提出到现在，它已经在各个方面都取得了巨大的发展，特

别是随着近十几年计算机的迅速普及,分位数回归模型的研究和应用更是成为统计和计量领域的热点之一。

本书围绕分位数估计,以分位数回归模型为主线,研究了变换可加分位数回归模型的估计方法和应用、分位数回归模型在分析空间数据中的应用,以及如何利用鞍点逼近技术求解分位数。变换可加分位数回归模型既能够体现出非参数模型的优点,又能克服一般非参数方法的缺点。将分位数回归思想用于空间数据分析是近年研究的一个热点,我们对这方面的问题也进行了一些探索。

本书的完成,得益于多方面的支持和帮助。作者感谢吴喜之教授长期给予的悉心指导和鼓励,感谢北京工商大学经济学院学术出版计划的支持,使得本书得以出版。特别感谢李宝仁教授,方燕副教授对本书出版给予的关心和帮助,感谢马若薇副教授为本书出版所做的相关工作。

由于水平所限,书中一定有不少错误和缺点,恳请批评指正。

# 目 录

|                      |    |
|----------------------|----|
| 第 1 章 概述             | 1  |
| 第 2 章 变换可加分位数回归      | 8  |
| 2.1 分位数回归简介          | 8  |
| 2.2 研究现状             | 13 |
| 2.3 模型               | 21 |
| 2.4 模型估计方法和性质        | 24 |
| 2.4.1 模型估计方法         | 24 |
| 2.4.2 估计量的相合性        | 30 |
| 2.5 性质证明             | 34 |
| 2.5.1 MH 估计性质证明      | 35 |
| 2.5.2 TSE 估计性质证明     | 56 |
| 2.6 模拟和经验运用          | 59 |
| 2.6.1 Monte Carlo 模拟 | 59 |
| 2.6.2 经验数据分析         | 65 |
| 2.7 讨论               | 69 |
| 第 3 章 分位数空间自回归模型     | 71 |
| 3.1 介绍               | 72 |
| 3.1.1 空间数据           | 72 |
| 3.1.2 空间自回归模型        | 74 |
| 3.2 模型和 IVQR 估计      | 77 |

|                                |            |
|--------------------------------|------------|
| 3.3 Bootstrap 方法构造置信区间 .....   | 82         |
| 3.3.1 空间数据的再抽样 .....           | 84         |
| 3.3.2 模拟研究 .....               | 88         |
| 3.4 经验数据应用 .....               | 90         |
| 3.5 半参数分位数空间自回归模型 .....        | 95         |
| 3.5.1 模型性质讨论 .....             | 99         |
| 3.6 模拟以及经验应用 .....             | 99         |
| 3.6.1 Monte Carlo 模拟 .....     | 99         |
| 3.6.2 经验数据分析 .....             | 101        |
| 3.7 讨论 .....                   | 107        |
| <b>第4章 鞍点逼近方法用于分位数估计 .....</b> | <b>115</b> |
| 4.1 介绍 .....                   | 115        |
| 4.2 鞍点逼近的原理和方法 .....           | 117        |
| 4.2.1 普通鞍点逼近 .....             | 119        |
| 4.2.2 经验鞍点逼近 .....             | 123        |
| 4.2.3 广义鞍点逼近 .....             | 123        |
| 4.3 鞍点逼近方法的运用 .....            | 125        |
| 4.3.1 VaR 值估计 .....            | 125        |
| 4.3.2 广义卡方型混合分布 .....          | 130        |
| 4.3.3 独立逆抽样下优势比 .....          | 138        |
| 4.4 讨论 .....                   | 148        |
| <b>参考文献 .....</b>              | <b>149</b> |

本书围绕分位数的三个相关主题展开研究，第一部分内容研究参数变换可加分位数回归模型的估计以及运用，第二部分内容研究分位数回归模型用于分析空间数据，第三部分内容研究鞍点逼近方法及运用。下面给出对每一个主题的简要介绍以及本书结构，具体内容参见本书各章。

第一部分研究变换可加分位数回归模型。Koenker and Bassett (1978) 引入的分位数回归方法被视为统计学中的一次重要突破。自该方法提出以来，从理论和运用两个方面都得到了很多的研究成果，包括各种新的分位数模型和估计推断方法，以及广泛的实际运用。分位数回归的直观原理非常简单明确。例如，线性分位数回归模型假设响应变量的条件分位数是协变量的线性函数，就像通常的线性回归模型假设响应变量的条件均值是协变量的线性函数一样。分位数回归方法扩展了通常的分位数概念，形成了一类新的模型。人们所熟知的线性最小一乘回归 (LAD) 就是线性分位数回归的一个特殊形式，相当于假设响应变量的条件中位数具有线性形式。我们知道，中

位数具有所谓的稳健性，在一些情况下比均值更适合作为随机变量的位置参数，同样，相对于条件均值模型，条件分位数回归模型也有稳健性的优势。更重要的是，分位数回归还能通过考察不同分位点的回归曲线，更全面的揭示响应变量和协变量之间的关系，而均值回归模型只能够描述对均值的影响。关于分位数回归的简要综述和最新进展，可以参看本书的介绍部分，以及综述性文章 Yu et al. (2004)，著作 Koenker (2005) 等。

对于一些实际问题，线性模型的假设可能过于苛刻，而完全非参数的方法会受到所谓“维数灾难”的影响。因此一些研究者考虑了更为灵活的模型。例如，Powell (1991)，Buchinsky (1995)，Mu (2005)，Mu and He (2007) 等将 Box-Cox 变换用于分位数回归，建立了变换线性分位数回归模型，这一模型能够包含普通的线性分位数回归模型作为特例。De Gooijer and Zerom (2003)，Yu and Lu (2004)，Horowitz and Lee (2005) 等研究了可加分位数回归模型。可加模型能够克服所谓的“维数灾难”问题，并且模型解释上比一般的非参数模型更加容易。

综合上面变换与可加模型两种思想，本书研究一类特殊的分位数回归模型，参数变换可加分位数回归模型。简单地讲，我们假设响应变量进行参数变换之后的条件分位数取可加模型。这一模型综合了变换模型和可加模型两者的优点，因此能够适应更多实际的数据分析。本书提出两个模型的估计方法，并在一定假设条件下，证明了估计量的相合性质。最后对估计方法进行 Monte Carlo 模拟分析，并将模型用于经验的数据研究。

第二部分研究分位数回归方法在分析空间数据中的运用。本书具体讨论分位数约束下的一类空间自回归模型，分位数空间自回归模型及运用。虽然分位数回归在许多领域都有了应用，但是将其用于分析空间数据，近几年才逐渐引起研究者的注意。

在实际数据分析和研究中，人们常常遇到一类与空间位置相关的数据。例如，区域经济数据，地质勘探数据，航空和卫星遥感数据等。这类数据的主要特点就是包含空间位置信息。空间关系的信息往往可以加以利用，有时候空间关系更是研究的重点。空间数据的一个基本特点是往往表现出空间相关性，与时间序列数据的自相关性类似。从经验上讲，在空间中距离较近的点，它们表现出来的关系也往往更密切。这种空间相关性能对模型的估计和预测产生影响，而有些时候我们的兴趣也在于发现这种空间相关性。其次，由于空间数据通常来自二维平面或者三维空间上，没有自然的先后顺序，因此空间数据还能表现出方向性（如山脉的走向等），比时间序列的相依性更复杂，这也是空间数据和时间序列数据重要的不同之处。由于空间现象中存在诸多不同方向，不同距离成分间的相互作用，致使空间数据的分析变得复杂，更多关于空间数据分析的复杂性问题的可以参考 Cressie (1993)。

空间模型在区域经济以及地理分析中已经有比较长的历史，近年在统计学和经济学等领域也逐渐得到了更多的重视。Cliff and Ord (1973, 1981) 引入的空间自回归模型是目前广泛应用的空间模型之一，与时间序列分析中的自回归模型类似，它能够比较简洁地描述空间交互作用。空间自回归模型有

着许多不同的设定,参见 Aselin (1988), Kelejian and Prucha (2005) 关于空间自回归模型的一般形式。在已有的模型估计方法中,通常假定扰动项具有零均值和有限方差,也就是在均值约束下进行建模。因此将分位数的方法与空间自回归相结合是比较自然的,得到的新模型能从多个方面拓宽空间自回归模型的运用范围。首先,相比已有的模型,新模型对异方差、异常点等有着更稳健的表现,其次,新模型可以通过估计不同分位数下的参数,得到更丰富的模型结果,从而揭示更多的空间信息。

近来一些研究考虑了将普通空间自回归模型扩展为分位数模型,其中 Su and Yang (2007) 考虑了一个估计方法,并给出了估计量的渐近正态分布,并且通过模拟表明新的模型具有良好的性质。我们注意到, Su and Yang 文章中所给出的渐进分布方差形式相当复杂,因此本文考虑利用 bootstrap 方法构造参数的置信区间,从而便于统计推断。在利用 Bootstrap 方法的时候,需要考虑到空间数据所具有的特殊结构,因此不同于独立样本下的 Bootstrap 方法。作为实际数据分析的例子,我们将新模型用于分析 Anselin (1988) 中一组具有空间结构的经验数据,得到了不同于普通空间自回归模型的一组结果,表明新模型能够发现其他模型难以发现的有趣现象。

另外,为使模型更加灵活,我们考虑解释变量以非参数的形式进入模型,提出分位数约束下的半参数分位数空间自回归模型,并考虑了模型经验的估计方法,最后给出了模拟和经验数据分析的例子。

第三部分研究鞍点逼近技术及其运用。

Daniles (1954) 将鞍点逼近技术引入到统计学中, 用于近似随机变量的密度函数、分布函数、条件密度和条件分布函数。鞍点逼近突出的特点是能够达到很高精度的逼近水平, 要优于常用的正态近似和 Edgeworth 展开, 在小样本情况下, 也能够有很好的近似效果。特别是对于分布函数的尾部密度逼近, 鞍点逼近方法能够达到很小的相对误差, 这一性质使得鞍点逼近方法能够用于小样本和小概率问题的研究。虽然鞍点逼近技术的引入已经有很长历史, 但只是近十几年鞍点方法在统计中的运用才得到更多的关注, 到目前为止, 已经有很多关于鞍点逼近方法理论和运用的研究。例如, 将鞍点逼近方法用于构造极大似然估计的分布函数, Jensen (1995) 将鞍点逼近技术用于极值理论等。由于鞍点逼近技术推导中涉及复变函数的积分, 所以略显得晦涩难懂, 因此其运用还不广为人知。但鞍点逼近的近似公式相对简单, 就运用来讲比较直接。本书将研究鞍点逼近方法的几个具体运用。

**风险值 (VaR, Value at Risk)** 是当今最流行的风险量化与评估方法, 已经成为金融资产风险管理和控制的重要工具之一。Jorion (1997) 定义 VaR 是“一段特定时期之内, 正常市场条件下, 给定置信水平下投资组合预期的最大可能损失。”换句话说, VaR 是描述给定投资组合可能遭受损失大小的一种直观方法。简单来讲, VaR 就是投资收益率分布的一个分位数, 通常是取小概率 (如 0.05, 0.01) 上的分位数。现在已经有 VaR 建模和计算的方法, 如 Risk Metrics 方法, 双曲分布模型, 极值方法, 分位数回归方法等。由于鞍点逼近方法能够对小概率近似达到很高的精度, 因此本书考虑将鞍点逼近

方法用于近似收益率的分布,进而计算 VaR 值。通过模拟和经验数据分析表明,鞍点逼近技术能够有效的运用于 VaR 值的计算。

在非参数模型检验中,许多检验统计量都属于一类所谓的广义卡方型混合随机变量。计算这类随机变量的分布或者密度,通常采用正态近似、卡方型近似,或者利用 Bootstrap 等计算密集型方法,参见 Zhang (2005)。本书注意到广义卡方型混合随机变量的密度函数虽然没有显式的表达式,但是其矩生成具有简单的形式,因此能够利用鞍点逼近方法进行近似。通过模拟计算表明,鞍点逼近方法要优于其他近似方法。

**优势比** (Odds Ratio) 是流行病学和医学研究中用来度量风险因子和疾病之间关系最常用的指标之一。优势比指标的不变性,使得其能够用于许多不同场合,参见 Lui (2004)。另外,逆抽样方法(也称为负二项抽样)是流行病学和医学中常用的抽样设计,特别是当潜在病例很稀少的时候,为了得到足够的观测病例数,多采用逆抽样方法进行抽样。在独立逆抽样设计下,优势比的置信区间估计存在三个常用的估计,所有这些估计都基于大样本近似。本书用鞍点逼近方法讨论构造优势比的置信区间,并进行了模拟和实际数据的分析,结果表明鞍点近似方法有着较好的效果,能够作为已有方法的有用补充。

全书安排结构如下:

第 2 章讨论变换可加分位数回归模型,包括模型的估计,估计的渐进性质和模拟以及经验数据的运用;第 3 章讨论分位数回归方法用于分析空间数据,建立分位数空间自回归模型,包括用 Bootstrap 方法构造参数的置信区间,以及一系列模拟

和经验运用结果；第4章研究鞍点逼近方法及其运用。从研究方法和研究内容上讲，主要创新和贡献有以下几点。

- 本书建立了变换可加分位数回归模型，并给出模型的两个估计方法，在一定假设条件下，证明了估计量的相合性质。通过模拟和经验数据分析表明，新模型能够有效的用于实际数据分析。
- 本书研究了分位数约束下的一类空间自回归模型，并利用 Bootstrap 方法构造了参数的置信区间，给出了模拟和经验运用的例子，表明分位数回归方法与空间模型相结合，能够更有效的揭示空间数据中所包含的信息，因此拓宽了分位数方法的运用范围，为空间数据分析提供了一个新的工具和思路。
- 本书研究了鞍点逼近及其运用。将鞍点逼近方法用于三类重要的问题之中，通过模拟和经验运用表明，鞍点逼近方法相比于一些已有近似方法有着更高的精度，是计算密度函数，分布函数，分位数近似值的有效工具。