

# 分位数回归发展、原理及其计算机实现<sup>①</sup>

史金凤<sup>1</sup> 刘维奇<sup>2</sup>

(1. 山西大学经济与管理学院, 太原 030006; 2. 山西大学管理与决策研究所, 太原 030006)

**摘要:** 分位数回归技术由于其假设条件宽松、提供更完整的统计关系而得到广泛的关注, 形成丰富研究成果。本文在梳理分位数回归理论发展脉络中归纳分位数回归的应用范围, 在介绍分位数回归基本原理和估计方法基础上, 详细讲述 Eviews、Stata 等软件窗口化操作步骤以及 R、SAS 等语言实现分位数回归的具体程序, 并就分位数回归的未来提出展望, 为分位数回归理论的拓展及其应用提供指引。

**关键词:** 分位数回归; Eviews; R

**中图分类号:** O212.4; F224.0 **文献标识码:** A **文章编号:** (2015) 01-0092-104

## 0 引言

分位数回归 (quantile regression) 是应用于统计、经济、金融等领域的一种回归分析方法, 是一种系统研究位置、规模、分布的方法, 是传统线性回归的扩展与补充, 挖掘的信息更加丰富。传统的线性回归模型描述的是因变量的条件均值分布受自变量的影响过程, 分位数回归不仅能够获得因变量均值水平受自变量的影响过程, 还能进一步提供因变量任一分位点受自变量的影响。分位数回归依据因变量的条件分位数对自变量进行回归, 得到了所有分位数下的回归模型, 相比普通的最小二乘回归, 能够更加精确地描述自变量对因变量的变化范围及条件分布形状的影响, 给出包含平均水平的相关关系在内一组不同分位点的相关关系结构。在现实生活中我们关心的不仅仅是平均水平上的关系, 而是更多分位点上的相关关系的结构, 此时分位数回归成为更好的选择。

相对于传统的线性回归模型系数估计的基本方法最小二乘法 (OLS), 分位数回归的假设条件更少。传统的线性回归模型的最小二乘估计在数据分布满足一定条件时具有无偏性、有效性等优良性质, 估计结果可靠。如果模型的随机误差项来自均值为零、方差相同的分布, 那么回归系数的最小二乘估计为最佳线性无偏估计 (BLUE); 如果随机误差项是正态分布, 那么回归系数的最小二乘估计与极大似然估计一致, 均为最小方差无偏估计 (MVUL)。然而, 在实际的经济生活中, 这种假设通常不能够满足, 特别地当数据中存在严重的异方差, 或厚尾、尖峰情况时, 最小二乘法的估计将不再具有上述优良性质, 致使最小二乘估计结果的正确性、准确性有待于进一步磋商。为了弥补普通最小二乘法在回归分析中的缺陷, 1818年 Laplace 提出了中位数回归 (最小绝对偏差估计), 1978年 Koenker 和 Bassett<sup>[1]</sup> 把中位数回归推广到了一般情形, 最早提出分位数回归的思想。不同于普通最小二乘法利用因变量的条件均值来建模并通过使残差平方和最小获得回归参数的估计, 分位数回归利用因变量的条件分位数来建模, 通过最小化加权的残差绝对值之和来估计回归参数。分位数回归的求解可以转化为线性规划问题, 使用单纯形法、内点算法等技术得到分位数回归的点估计, 通过直接估计法、秩得分法、重复抽样法得到区间估计, 使其对数据的要求放宽, 适用于具有异方差、尖峰后尾等特征的数据, 在经济金融等领域得到了广泛的发展。应用发展也推动了分位数回归理论模型的拓展, 分位数回归的时间序列模型<sup>[2]</sup>、面板模型<sup>[3]</sup>、非参数模型<sup>[4]</sup>、混合模型<sup>[5]</sup>应运而生。

<sup>①</sup> 基金项目: 教育部人文社科基金青年项目 (14YJC630163)、山西大学校科研基金 (011551901006)。

作者简介: 史金凤 (1982—), 女, 山西榆次人, 博士, 研究方向: 金融工程与风险管理等, Email: 65564355@qq.com; 刘维奇 (1963—), 男, 山西忻州人, 博士, 山西大学管理与决策研究所教授, 博士生导师, 研究方向: 金融工程与风险管理、时间序列分析, Email: liuwq@sxu.edu.cn。

分位数回归的思想起源于1760年,然而,这一回归方法计算的复杂性却是一大挑战。尽管计算机功能的快速发展和统计软件的广泛应用才使得拟合分位数回归模型变得方便,应用分位数回归解决现实问题却仍不容易。本文拟将从分位数回归发展脉络中探析分位数回归技术的应用范围,介绍基本原理的基础上讲述Eviews、Stata等软件、MatLab、R等语言实现分位数回归的具体操作,并以实际数据案例进行实证分析,以期初学者提供快速指引,为社会科学研究者更加熟悉分位数回归贡献力量。

## 1 分位数回归的发展

分位数回归技术由于其假设条件宽松得到广泛的关注,形成了丰富的研究成果。1978年第一篇分位数回归论文才刚刚问世,经过30余年的发展相关论文数量就已破万,分位数回归理论成长之快速不言而喻,特别是近年来发展速度更为迅猛。英文文献方面,在Elsevier SD数据库中以“quantile regression”作为关键词,选择All field选项搜索可以得到10805条结果,选择Title、Abstract、Keywords选项搜索则可得819条结果<sup>①</sup>,每年发表相关论文情况如图1所示,发文量前20的期刊如表1所示。从发表论文数量来看,分位数回归相关论文数量加速增长。1995年前的近20年时间里,分位数回归相关的论文总共仅有754篇,以分位数回归为题目摘要或关键词的论文仅有29篇,2000年前每年的发文量也仅仅百余篇,之后分位数回归的研究增长飞跃,特别是2005年之后,每年的发文数量比上年净增长100篇左右,仅2014年1到10月份的分位数相关论文发文量已是1633篇,是1996—2000年5年间发文总量的2.6倍。表1则展示出分位数回归涉及的领域也很广泛。从发表论文所在期刊来看,从深度上进一步拓展、改进和优化分位数回归模型的相关论文发表在统计学期刊*Computational Statistics & Data Analysis*、*Journal of Statistical Planning and Inference*、*Statistics & Probability Letters*、*Journal of Multivariate Analysis*等以及计量学期刊*Journal of Econometrics*、*Economic Modelling*等,这类论文不断深入探索参数估计、模型识别、变量选择、方差估计等统计技术,提出半参数模型、多分位数混合模型、两步模型等分位数回归模型,并将分位数回归模型的适用范围拓展到受限数据、面板数据、大规模数据等。从广度上进一步拓展应用的相关论文则发表于经济学期刊(如*Economics Letters*、*Labour Economics*等)、金融学期刊(如*Journal of Banking & Finance*)、运筹学期刊(如*European Journal of Operational Research*)、水文学期刊(如*Journal of Hydrology*)、环境学期刊(如*Atmospheric Environment*)等,这类论文应用分位数回归研究全球汇率的时间依赖结构、公职与收入不平等的关系、多元化对企业绩效的影响、系统性风险的度量、母亲沮丧对孩子认知能力的影响、货币环境对股票收益的影响、道路交通对地面臭氧的作用等问题,分析不同分位点处因变量与自变量之间的依赖关系变化,提供更多的相关性分析结论。

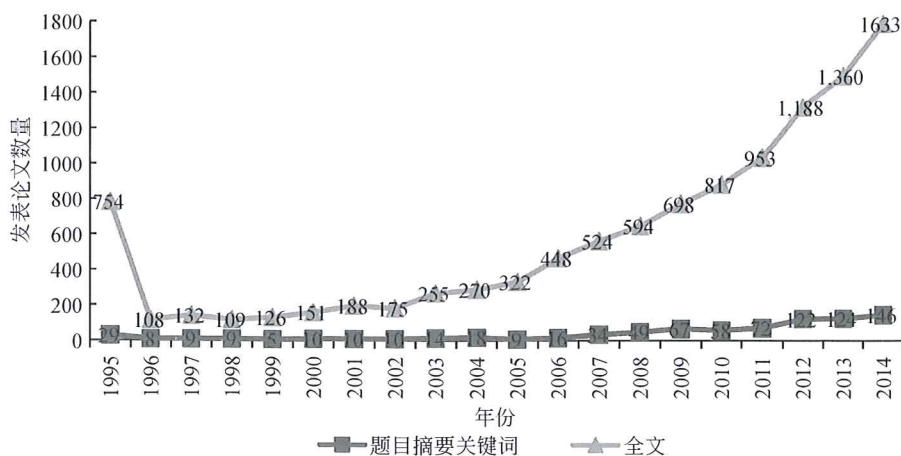


图1 Quantile regression 相关英文论文发表数量

Fig. 1 The number of published papers in English about quantile regression

图中1995年对应的发表论文数量是1995年以前年份的发表相关主题论文的总量。

<sup>①</sup> 文中所有数据库中检索数据的检索时间为2014年10月8日。



中文文献方面,在中国知网以“分位数回归”为关键词在中国学术期刊网络出版总库(CAJD)精确检索得到911条信息,文献的学科、作者、基金分布情况如表2所示。分位数回归的中文论文主要聚焦于分位数回归在经济、金融、管理中的应用,涉及宏观经济管理与可持续发展、金融、经济体制改革、投资、证券等学科。作者与单位角度,中国矿业大学的王新宇、宋学锋团队、西南交通大学的魏宇探索分位数回归在股票市场风险度量的研究,合肥工业大学的许启发、蒋翠侠团队则主要探索分位数回归工资收入、消费行为等经济问题,宁波大学的俞立平着眼于分位数回归在科技创新、期刊评价两个方面的应用,吉林通化师范学院的赵海英、张明旭运用分位数回归分析粮食产量影响因素相关问题,厦门大学的陈建宝南开大学的周兴、王芳等基于分位数回归探讨收入差别影响因素及其后果,厦门大学的方匡南探讨社会保障对城乡家庭消费的影响、股指期货价格发现功能等问题研究,同济大学的简泽探析竞争的增强对企业层面全要素生产率增长的影响,山西大学的刘维奇致力于基于分位数回归多的金融市场稳定性度量,等等。从基金资助层面来看,受到国家自然科学基金资助的有165篇,受到国家社会科学基金资助的有148篇,合计达到所有相关论文的34.36%,可见国家自然科学基金委员会对分位数回归相关问题研究的支持力度不小,此外,分位数回归相关研究还得到跨世纪优秀人才、博士后基金、国家科技支撑计划、国家高技术研究发展计划、湖南省教委科研基金、国家软科学研究计划、国家留学基金、浙江省自然科学基金、长江学者奖励计划以及上海、广东、山东、湖南、浙江等地区的自然科学基金资助。

表1 Quantile regression 相关英文论文发表期刊

Table 1 The journal of published papers in English about quantile regression

排名	在题目摘要关键词中搜索		在全文中搜索	
	期刊名	数量	期刊名	数量
1	Computational Statistics & Data Analysis	64	Computational Statistics & Data Analysis	620
2	Journal of Econometrics	54	Journal of Statistical Planning and Inference	540
3	Journal of Statistical Planning and Inference	40	Journal of Econometrics	537
4	Statistics & Probability Letters	34	Journal of Hydrology	330
5	Journal of Multivariate Analysis	31	Journal of Multivariate Analysis	288
6	Economics Letters	29	Statistics & Probability Letters	261
7	Economics & Human Biology	18	Journal of Banking & Finance	147
8	Journal of Hydrology	17	Handbook of Statistics	138
9	Economic Modelling	15	Economics Letters	126
10	Labour Economics	13	European Journal of Operational Research	117
11	Economics of Education Review	12	Insurance; Mathematics and Economics	109
12	Journal of the Korean Statistical Society	12	Forest Ecology and Management	105
13	Journal of Banking & Finance	10	International Journal of Forecasting	100
14	European Journal of Operational Research	8	The American Journal of Human Genetics	99
15	Atmospheric Environment	7	Remote Sensing of Environment	91
16	Ecological Economics	7	Economic Modelling	82
17	Insurance; Mathematics and Economics	7	Reliability Engineering & System Safety	75
18	International Journal of Forecasting	7	Economics of Education Review	74
19	Journal of Development Economics	7	Labour Economics	73
20	World Development	7	Journal of Empirical Finance	64

表 2 分位数回归相关的中文论文分布情况

Table 2 The outline of papers in Chinese about quantile regression

学科	数量	作者	作者单位	数量	基金支持	数量
宏观经济管理与可持续发展	508	王新宇	中国矿业大学	11	国家自然科学基金	165
数学	209	许启发	合肥工业大学	10	国家社会科学基金	148
金融	197	俞立平	宁波大学	7	跨世纪优秀人才培养计划	28
经济体制改革	156	赵海英	吉林通化师范学院	7	中国博士后科学基金	12
投资	142	张明旭	吉林通化师范学院	7	高等学校博士学科点专项科研基金	10
证券	101	蔡超	山东工商学院	6	上海市重点学科建设基金	8
人才学与劳动科学	86	陈建宝	厦门大学	5	广东省自然科学基金	8
企业经济	84	蒋翠侠	合肥工业大学	5	山东省自然科学基金	8
市场研究与信息	60	周兴	南开大学	5	国家科技支撑计划	8
农业经济	58	魏宇	西南交通大学	4	国家高技术研究发展计划	7
工业经济	51	王艳梅	温州大学	4	湖南省教委科研基金	6
贸易经济	45	简泽	同济大学	4	国家软科学研究计划	5
财政与税收	40	董乃铭	温州大学	4	国家留学基金	5
教育理论与教育管理	21	王杰	山东大学	4	浙江省自然科学基金	4
社会学及统计学	19	张世伟	吉林大学	4	长江学者奖励计划	4
保险	19	Simon	诺丁汉大学	4	江苏省教育厅人文社会科学研究基金	4
高等教育	18	宋学锋	中国矿业大学	4	吉林省自然科学基金	3
环境科学与资源利用	13	刘启浩	北京工业大学	4	安徽省自然科学基金	3
经济理论及经济思想史	13	张文君	江西财经大学	4	湖南省自然科学基金	3
水产和渔业	13	夏庆杰	北京大学	4	湖南省社会科学基金	3
人口学与计划生育	10	宋丽娜	诺丁汉大学	4	江苏省青蓝工程基金	3
医药卫生方针政策与法律法规研究	9	宋利明	上海海洋大学	3	北京市教委科技发展基金	2
服务业经济	8	武夷山	中国科学技术信息研究所	3	广西壮族自治区软科学研究计划	2
行政学及国家行政管理	8	阮素梅	安徽财经大学	3	教育部科学技术研究项目	2
科学研究管理	6	王芳	南开大学	3	中国科学院知识创新工程基金	2
电力工业	6	郭念国	河南工业大学	3	安徽省软科学研究计划	2
动力工程	5	张海燕	长春工业大学	3	上海市高等学校科学技术发展基金	2
图书情报与数字图书馆	5	李小波	河北工程大学	3	陕西省软科学研究计划	2
经济统计	5	张文君	江西行政学院	3	山西省软科学研究计划	2
政党及群众组织	5	刘维奇	山西大学	3	山东省软科学研究计划	2
石油天然气工业	5	张良桥	顺德职业技术学院	3	重庆市教委科研基金	2
旅游	4	姚先国	浙江大学	3	北京市优秀人才基金	2
管理学	3	李红梅	北方工业大学	3	美国中华医学基金	2
出版	3	锁志海	西安交通大学	3	全国教育科学规划	1
信息经济与邮政经济	3	任燕燕	山东大学	3	农业部“948”项目	1
审计	3	钱文荣	浙江大学	3	湖北省软科学研究计划	1
中国政治与国际政治	3	方匡南	厦门大学	3	广东省软科学研究计划	1
会计	3	陈新军	上海海洋大学	3	霍英东教育基金	1
水利水电工程	3	王维国	东北财经大学	3	福建省软科学研究计划	1
地质学	3	吴拥政	湖南师范大学	3		

综上，分位数回归在学者的关注与政府的支持下快速发展，形成丰富的成果，相关研究主要集中于两个视角，一是统计模型本身的深化与拓展，二是模型在经济、金融、管理中的应用。前者关注于参数估计、渐进理论等理论框架的探索，后者则更多的是实证研究，通过分位数回归探讨不同情况下的相关关系



结构, 提供了更加全面的分析。

分位数回归发展迅速, 论文数量颇丰, 哪些论文受到关注更多呢? 以金融为例, 分位数回归引用频率最高的中文论文如表3所示, 呈现出一些特点。第一, 《经济研究》发表的相关论文引用更多, 2009年苏冬蔚与曾海舰发表于《经济研究》的“宏观经济因素与公司资本结构变动”一文引用最多, 被引用149次, 下载5286次, 其次是2008年张车伟与薛欣欣发表于《经济研究》的“国有部门与非国有部门工资差异及人力资本贡献”, 被引用135次, 下载3171次, 且被引前两名均发表于《经济研究》。第二, 综述性的文章均有较高的引用, 发表于《统计与信息论坛》的三篇综述引用率相对较高。金融学科方面的分位数回归相关论文发表期刊分布见图2, 可见分位数回归在《统计与决策》、《系统工程理论与实践》、《中国管理科学》等期刊更受欢迎, 是应用分位数回归技术解决金融问题相关研究的投稿首选。

表3 分位数回归引用频率最高的中文论文

Table 3 The highly cited papers in Chinese about quantile regression

	论文名	作者	刊名	年/期	被引次数	下载次数
1	宏观经济因素与公司资本结构变动	苏冬蔚, 曾海舰	经济研究	2009/12	149	5286
2	国有部门与非国有部门工资差异及人力资本贡献	张车伟, 薛欣欣	经济研究	2008/04	135	3171
3	分位数回归技术综述	陈建宝, 丁军军	统计与信息论坛	2008/03	128	3295
4	技术引进与中国企业的自主创新: 基于分位数回归模型的经验研究	孙文杰, 沈坤荣	世界经济	2007/11	103	2672
5	分位数回归及应用简介	李育安	统计与信息论坛	2006/03	94	2864
6	中美股票市场的联动性研究	张兵, 范致镇, 李心丹	经济研究	2010/11	87	3421
7	中国垄断行业的高收入及其原因: 基于整个收入分布的经验研究	傅娟	世界经济	2008/07	84	1667
8	教育和经验对中国居民收入的影响——基于分位数回归和审查分位数回归的实证研究	刘生龙	数量经济技术经济研究	2008/04	65	2383
9	贸易开放、人力资本与中国全要素生产率——基于分位数回归方法的经验研究	魏下海	数量经济技术经济研究	2009/07	61	2326
10	随机森林方法研究综述	方匡南, 吴见彬, 朱建平, 谢邦昌	统计与信息论坛	2011/03	60	3237

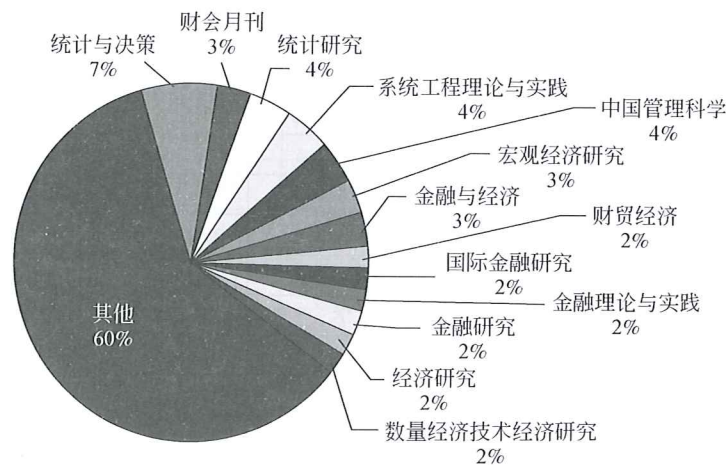


图2 Quantile regression 相关金融学科中文论文发表期刊分布

Fig. 2 The outline of published papers in English about quantile regression and finance

注: 图中1995年对应的发表论文数量是1995年以前年份的发表相关主题论文的总量。

## 2 分位数回归的原理<sup>[6]</sup>

分位数回归的研究对象是变量不同分位数,可以描述一些因素如何影响研究对象的中位数、1/4分位数、3/4分位数等,这些不同分位数代表了处于不同水平的研究对象。而不同分位数下的参数估计量往往也不同,这就代表同样的影响因素对处在不同水平的研究对象的作用大小不同。特别是在研究对象的分布呈现异质性,如不对称、厚尾、截断性等特征时,这一方法往往对数据中出现的异常点具有耐抗性;其次,从需要满足的假设条件来看,一般的线性回归需要满足一系列较强的假设条件,而经济变量却往往做不到这一点。相比之下,分位数回归的假设条件要弱很多,与现实情况更为接近,如它对模型中的随机扰动项不需做任何分布假定,这样整个回归模型就具有很强的稳健性,因此其参数估计量无论在有效性还是其他方面都优于一般的线性回归。特别是当经济问题中存在一些离群值(指某一个样本点的取值突然明显高于或低于其他样本)时,中位数回归方法的优势更加明显。所有这些都使得运用分位数回归得到的经济变量之间的关系更加真实、准确。另外,分位数回归还具有比较好的弹性性质,对因变量具有单调变幻性,其估计出来的参数具有在大样本理论下的渐进优良性。

从概率与数理统计角度来说,对一个连续随机变量 $y$ ,如果 $y$ 小于等于 $q_\tau$ 的概率是 $\tau$ ,则我们说 $y$ 的 $\tau$ 分位值是 $q_\tau$ ,或者说 $q_\tau$ 就称作 $y$ 的第 $\tau$ 分位数。类似地,如果我们将被解释变量 $y$ 表示为一系列解释变量 $X$ 的线性表达式(又称为拟合值),并使得该表达式满足小于等于 $q_\tau$ 的概率是 $\tau$ ,就称为分位数回归。如果用数学语言描述其估计原理,实际上是使一个关于 $y$ 与其拟和值之差(又称作残差)的绝对值的表达式最小。

分位数回归是对以古典条件均值模型为基础的最小二乘法的延伸,用多个分位函数来估计整体模型。中位数回归是分位数回归的特殊情况,用对称权重解决残差最小化问题,而其他的条件分位数回归则用非对称权重解决残差最小化。从而反映出不同分位点处变量间的相关关系。

### 2.1 分位数回归的参数估计

一般线性回归模型可写成如下形式

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k + u \quad (1)$$

其中, $u$ 为随机扰动项, $\alpha_0, \alpha_1, \alpha_2, \cdots, \alpha_k$ 为待估参数。在满足高斯——马尔可夫假设前提下,上述模型的均值回归(OLS)描述的是自变量与因变量平均水平的关系,可表示为

$$E(y | x) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k \quad (2)$$

基于均值回归模型的思想,分位数回归模型可以定义为

$$Q_\tau(y | x) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k + Q_\tau(u) \quad (3)$$

其中, $Q_\tau(y) = \inf\{y: F(y) \geq \tau\}$ , $0 < \tau < 1$ 为 $y$ 的 $\tau$ 分位数, $F(y)$ 为 $y$ 的分布函数。

类似于均值回归模型参数估计的最小化残差平方和的思路,分位数回归模型参数通过最小化加权绝对偏差来求解,即

$$\min E \rho_\tau(y - \alpha_0 - \alpha_1 x_1 - \alpha_2 x_2 - \cdots - \alpha_k x_k) \quad (4)$$

其中 $E \rho_\tau(X - x) = (\tau - 1) \int_{-\infty}^{\hat{x}} (x - \hat{x}) dF(x) + \tau \int_{\hat{x}}^{\infty} (x - \hat{x}) dF(x)$ , $\rho_\tau(u) = u(\tau - I(u < 0))$ , $0 < \tau < 1$ 为检查函数,见图3; $I$ 为示性函数,自变量大于0时为1,反之为0。从图3容易看出, $\rho_\tau(u)$ 相当于对 $u$

然而,一般线性回归模型的原理是使得被解释变量 $y$ 与其拟合值之差(称作残差)的平方和最小,而分位数回归是使得这个残差的绝对值的一个表达式最小,这个表达式不可微,因此传统的求导方法不再适用,而是采用线性规划方法。采取线性规划法(LP)最小化加权绝对偏差来求解解释变量的回归系数,可以求解得到参数的估计值 $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2, \cdots, \hat{\alpha}_k$ ,相应的预测值为

$$Q_\tau(y | x) = \hat{\alpha}_0 + \hat{\alpha}_1 x_1 + \hat{\alpha}_2 x_2 + \cdots + \hat{\alpha}_k x_k \stackrel{\Delta}{=} Z' \hat{\alpha}(\tau) \quad (5)$$

其中, $Z = (1, X)'$ , $\hat{\alpha}(\tau) = [\hat{\alpha}_0(\tau), \hat{\alpha}_{1k}(\tau)]'$ 。

上述线性规划求解时,常用的计算方法有单纯性算法、内点算法、平滑算法、自适应算法等。



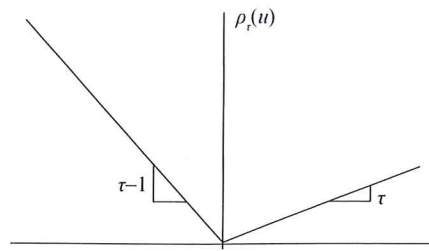


图3 检查函数  $\rho_\tau(u)$

Fig. 3 The picture of checking function  $\rho_\tau(u)$

### 2.2 分位数回归的假设检验

分位数回归估计的假设检验主要包括拟合优度检验、拟似然比检验、Wald 检验、斜率相等检验、斜率对称检验。

#### 2.2.1 拟合优度检验

类似于普通最小二乘估计，拟合优度检验考察解释变量对被解释变量第  $\tau$  分位数回归拟合的好坏，检验统计量为

$$R_{(\tau)}^* = 1 - \frac{\hat{Q}_{(\tau)}}{\tilde{Q}_{(\tau)}} \tag{6}$$

其中代表无约束的和有约束目标方程的极小值分别为  $\hat{Q}_{(\tau)}$  和  $\tilde{Q}_{(\tau)}$ ，

$$\hat{Q}_{(\tau)} = \min \left[ - \sum_{i: y_i < X' \hat{\beta}_{(\tau)}} (1 - \tau)(y_i - \hat{\alpha}_{0(\tau)} - Z' \hat{\alpha}_{1k(\tau)}) + \sum_{i: y_i \geq X' \hat{\beta}_{(\tau)}} \tau(y_i - \hat{\alpha}_{0(\tau)} - Z' \hat{\alpha}_{1k(\tau)}) \right] \tag{7}$$

$$\tilde{Q}_{(\tau)} = \min \left[ - \sum_{i: y_i < X' \hat{\alpha}_{(\tau)}} (1 - \tau)(y_i - \hat{\alpha}_{0(\tau)}) + \sum_{i: y_i \geq X' \hat{\alpha}_{(\tau)}} \tau(y_i - \hat{\alpha}_{0(\tau)}) \right] \tag{8}$$

因为  $\hat{Q}_{(\tau)} \leq \tilde{Q}_{(\tau)}$ ，所以  $R_{(\tau)}^*$  的值在 0 和 1 之间，解释变量的作用越强， $\hat{Q}_{(\tau)}$  越远远小于  $\tilde{Q}_{(\tau)}$ ， $R_{(\tau)}^*$  越接近于 1，反之，越接近于 0。

#### 2.2.2 拟似然比检验

Koenker 和 Machado (1999) 根据目标函数在施加约束条件前后得到的两个极小值构造了两个拟似然比检验统计量 (QLR)，两统计量的表达式如下

$$L_T(\tau) = \frac{2(\tilde{Q}_{(\tau)} - \hat{Q}_{(\tau)})}{\tau(1 - \tau)s(\tau)} \tag{9}$$

$$A_T(\tau) = \frac{2\hat{Q}_{(\tau)}}{\tau(1 - \tau)s(\tau)} \log\left(\frac{\tilde{Q}_{(\tau)}}{\hat{Q}_{(\tau)}}\right) \tag{10}$$

两个统计量都渐近服从自由度为  $q$  的卡方分布，其中  $q$  是原假设目标函数中约束条件的个数， $s(\tau)$  是分位数密度函数。如果约束是无效的，有约束的最大似然函数值当然不会超过无约束的最大似然函数值，但如果约束条件“有效”，有约束的最大值应当“接近”无约束的最大值，这正是似然比检验的基本思路。因此，选择原假设为  $H_0: g(\alpha) = C$ 。

#### 2.2.3 Wald 检验

给定分位数回归参数估计量的渐近方差协方差矩阵，我们就可以构造 Wald 形式的统计量进行各种约束形式的参数检验。Wald 统计量的一种表达形式

$$w = \frac{n(e'_* e_* - e'e)}{e'e} \tag{11}$$

其中， $e'_* e_*$  是有约束模型残差平方和， $e'e$  是有约束模型残差平方和， $w$  服从自由度为  $q$  的卡方

分布。

#### 2.2.4 斜率相等检验

斜率相等检验，即检验对于不同的分位点，估计得到的结构参数（在线性模型中即为斜率）是否相等。原假设被设定为  $H_0: \hat{\alpha}_i(\tau_1) = \hat{\alpha}_i(\tau_2) = \dots = \hat{\alpha}_i(\tau_m)$ ,  $i = 1, \dots, k$ 。  $\hat{\alpha}_i$  指常数项以外的解释变量所对应的  $(k-1)$  维参数列向量。因此，零假设共含有  $(k-1)(m-1)$  个约束条件，构造 Wald 形式的统计量检验零假设是否成立，它渐近服从自由度为  $(k-1)(m-1)$  的卡方分布。如果接受该假设，说明每个斜率对于不同分位点具有不变性，此时，应该采用普通最小二乘估计；如果拒绝该假设，说明模型应该采用分位数回归估计，以反映每个斜率在不同分位点的不同值。

#### 2.2.5 斜率对称检验

斜率对称性检验，即检验对于给定的  $X, Y$  的分布是否是对称的。假设我们要检验的分位数回归模型有  $m$  个， $m$  是奇数，且中间值  $\tau_{(m+1)/2}$  是 0.5，其他  $\tau$  都关于 0.5 对称，即  $\tau_j = 1 - \tau_{m-j+1}$ ,  $j = 1, \dots, (m-1)/2$ 。参数估计量按照  $\tau_k$  的大小排序，则对称性检验的零假设为  $H_0: [\alpha(\tau_j) + \alpha(\tau_{m-j+1})]/2 = \alpha(0.5)$ 。可以构造 Wald 形式的统计量检验上述  $k(m-1)/2$  个约束条件是否成立。该统计量服从自由度为  $k(m-1)/2$  的卡方分布。如果接受斜率相等性假设，就不必进行斜率对称性检验。如果拒绝斜率相等性假设，则可以进一步进行斜率对称性检验，若接受原假设，则认为斜率具有对称性，否则，则认为斜率不具有对称性。

### 3 分位数回归的实现

分位数回归技术经过 30 余年的发展，已成为解决经济、金融、工程等领域问题的一种纯熟技术，在软件方面也得到了有力的发展。随着计算机技术的不断突破，上述算法可以很方便地由各种软件实现。现在主流统计、计量与科学计算软件 EViews、STATA、R、SAS、MATLAB 等中都可以加载软件包。目前，在 Eviews、Stata 等软件中有内化的可视化选项执行分位数回归，SAS、MatLab、R 等语言的分位数回归程序包也随着分位数回归技术的发展不断丰富，为分位数回归技术的应用提供了广阔的空间。下面将介绍分位数回归技术的计算机实现方法，为分位数回归技术的使用提供指引。

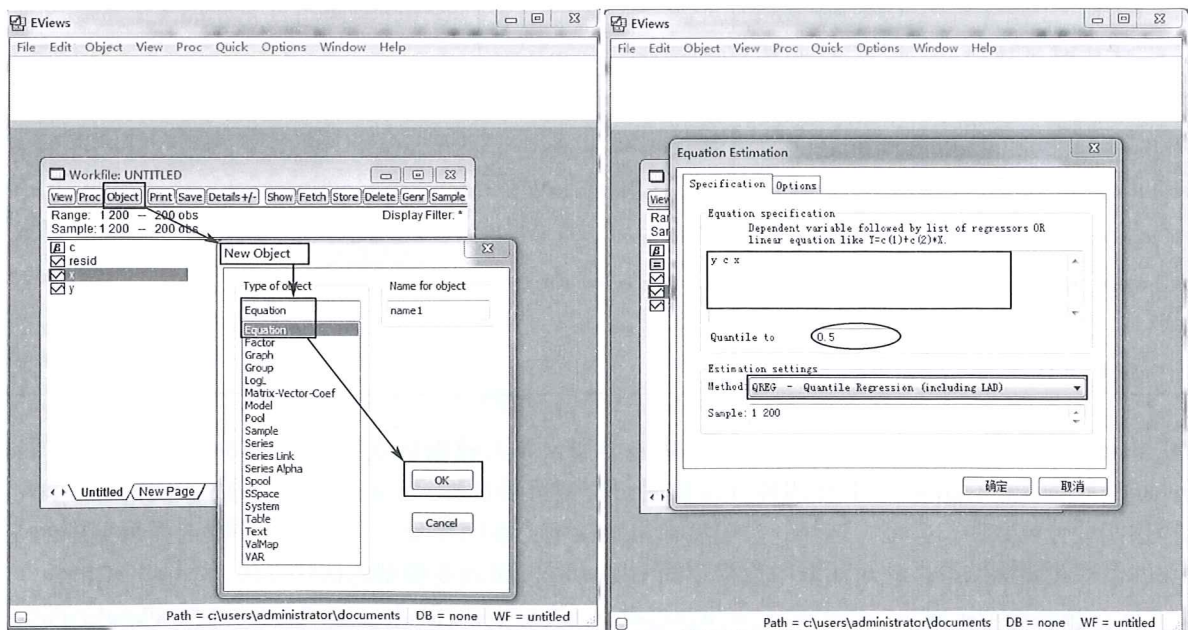


图 4 Eviews 软件分位数回归操作窗口图

Fig. 4 The windows of quantile regression in eviews



### 3.1 窗口化软件实现步骤

在 Eviews 软件中，建立工作簿导入数据，在图 4 左图窗口点菜单栏 object→new object→Equation→OK，弹出图 4 右图窗口中 Equation specification 中输入回归方程，如上回归方程则输入回归模型“y C X”（因变量 常数项 自变量），在 Quantile to 后边的框内输入分位数  $\tau$  的值（大于 0 小于 1 的数字），在 method 选项中选择 QREG-Quantile Regression (including LAD) 选项，点击确定弹出输出结果。如表 5 所示。上方的信息依次是因变量名称；回归模型（选择  $\tau$  的值为 0.5，此行显示 Median，若选择  $\tau$  的值为 0.6，此行显示 tau=0.6）；回归时间与回归耗时；样本；模型设定 Huber Sandwich 方法估计系数协方差，Siddiqui (mean fitted) 方法得到稀疏度，用 Hall-Sheather 计算带宽；最后一行说明成功得到唯一解。Variable、Coefficient、Std. Error、t-Statistic、Prob. 分别对应变量名，估计系数、估计标准差、t 统计量及 P 值。Pseudo R-squared、Adjusted R-squared 为 R<sup>2</sup>，Objective 为目标函数最小值，Objective (const. only) 仅含常数的目标函数最小值，Quasi-LR statistic、Prob (Quasi-LR stat) 为似然比及其 P 值。该软件还提供了过程系数 (Process Coefficients)、斜率相等 (slope equality test)、对称检验 (symmetric quantile test) 三个检验。

表 4 Eviews 软件分位数回归输出结果  
Table 4 The results of quantile regression in eviews

Dependent Variable: AS				
Method: Quantile Regression (Median)				
Date: 12/30/14 Time:				
Sample (adjusted):				
Included observations: 39 after adjustments				
Huber Sandwich Standard Errors & Covariance				
Sparsity method: Kernel (Epanechnikov) using residuals				
Bandwidth method: Hall-Sheather, bw =				
Estimation successfully identifies unique optimal solution				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C				
AS (-1)				
Pseudo R-squared	Mean dependent var			
Adjusted R-squared	S. D. dependent var			
S. E. of regression	Objective			
Quantile dependent var	Objective (const. only)			
Sparsity	Quasi-LR statistic			
Prob (Quasi-LR stat)				

在 stata 软件中的实现方法与 Eviews 软件基本类似，导入数据后点菜单栏下 Statistics→linear models and related→Quantile regression 下有四种分位数回归模型可供选择。Statistics→Nonparametric analysis 下也可选择这四种分位数回归模型。选任一种模型弹出对话框，在因变量、自变量选项中选择模型中的变量，分位点的值即可输出结果，此对话框中选择不同 Option 选项可改变模型的估计方法。

### 3.2 程序化软件实现步骤

#### R 语言 quantreg 包

R 语言分位数回归实现需要加载程序包，在联网情况下在 R 的命令窗口中输 install.packages ("quantreg") 即可安装 quantreg 包。值得注意的是，每次使用分位数回归前需要输入 library ("quantreg") 加载 quantreg 包，此外还可以通过如下方式进入帮组页面：

```
help.start()          #进入 R 帮助首页
help(rq)              #获取 rq 函数的帮助,也可以写成:? rq
example(rq)           #显示分位数回归函数 rq() 的一个简单示例代码
```

MATLAB、SAS 中同样也可以下载 quantreg 包后实现分位数回归,SAS 帮助中如后所示有具体程序。  
分位数回归模型具体实现程序如下:

R 语言实现程序<sup>①</sup>:

##参数估计及基本检验

```
data(engel)          #加载 quantreg 包自带的数据集,使用时导入要分析的数据集。
fit1=rq(foodexp ~ income,tau=0.6,data=engel) #进行60%分位点的分位数回归
fit1                 #直接显示分位数回归的模型和系数
summary(fit1)        #得到更加详细的显示结果
r1=resid(fit1)       #得到残差序列,并赋值为变量 r1
c1=coef(fit1)        #得到模型的系数,并赋值给变量 c1
summary(object,se=NULL,covariance=FALSE,hs=TRUE,...) #通过设置参数 se 可以得到系数假设
检验
```

其中主要参数有:

# object: 分位数回归对象,根据 rq() 函数等得到的结果如 fit1。

# se: 用于计算参数估计值标准差的方法,可以选取的值包括:

-rank: 根据 Koenker(1994)的秩检验得到标准差的估计值。默认情况下假定残差是服从独立同分布。如果补充另一个参数 iid=FALSE,则采用 Machado(1999)的方法计算标准差(参数的写法:se="rank",iid=FALSE)。

-iid: (这个与上面提到的 iid=FALSE 不同,这里是参数 se 的一个取值,而上面的 iid 是一个逻辑参数)假定残差服从独立同分布,并按照 KB(1978)的方法计算残差。

-nid: 用 sparsity 算法计算的参数估计值标准差。

-ker: 用 Powell(1990)的核密度估计方法得到标准差。

#boot: 采用 bootstrap 自助抽样的方法计算标准差。

-默认情况下,se=NULL 且 covariance=FALSE,标准差的默认算法是 se="rank";其他情况下,se 默认值为"nid"。

##不同分位点下的系数估计值的比较

```
fit1=summary(rq(foodexp ~ income,tau=2:98/100))
fit2=summary(rq(foodexp ~ income,tau=c(0.05,0.25,0.5,0.75,0.95)))
plot(fit1)
plot(fit2)
# #散点图
attach(engel)        #打开 engel 数据集,直接运行其中的列名,就可以调用相应列
plot(income,foodexp,cex=0.25,type="n", #画图,说明①
      xlab="Household Income",ylab="Food Expenditure")
points(income,foodexp,cex=0.5,col="blue") #添加点,点的大小为 0.5
abline(rq(foodexp ~ income,tau=0.5),col="blue") #画中位数回归的拟合直线,颜色蓝
abline(lm(foodexp ~ income),lty=2,col="red") #画普通最小二乘法拟合直线,颜色红
taus=c(0.05,0.1,0.25,0.75,0.9,0.95)
for(i in 1:length(taus)) { #绘制不同分位点下的拟合直线,颜色为灰色
```

① 北京师范大学经济管理学院詹鹏,http://www.ciidbnu.org/news/201212/20121218191555706.html。



```
abline(rq(foodexp ~ income, tau = taus[i]), col = "gray")
```

```
detach(engel)
```

```
##比较不同分位点下,相关性是否相同
```

```
fit1=rq(foodexp ~ income, tau=0.25) fit2=rq(foodexp ~ income, tau=0.5)
```

```
fit3=rq(foodexp ~ income, tau=0.75) anova(fit1, fit2, fit3)
```

```
SAS 实现程序
```

```
proc quantreg data=a;
```

```
model y=x1 x2;
```

```
run;
```

```
proc quantreg algorithm=interior(tolerance=1e-6)
```

```
ci=none data=a;
```

```
model y=x1 x2/itprint nosummary;
```

```
run;
```

```
matlab 实现程序
```

```
function b=rq_fnm(X,y,p)
```

```
% Construct the dual problem of quantile regression
```

```
% Solve it with lp_fnm
```

```
[m n]=size(X); u=ones(m,1); a=(1-p). * u; b=-lp_fnm(X',-y',X' * a,u,a)';
```

```
function y=lp_fnm(A,c,b,u,x) % Solve a linear program by the interior point method;
```

```
% min(c * u), s. t. A * x=b and 0<x<u % An initial feasible solution has to be provided as x
```

```
% Function lp_fnm of Daniel Morillo & Roger Koenker Translated from Ox to Matlab by Paul Eilers 1999
```

```
% Modified by Roger Koenker 2000-- More changes by Paul Eilers 2004
```

```
% Set some constants
```

```
beta=0.9995; small=1e-5; max_it=50; [m n]=size(A);
```

```
% Generate initial feasible point
```

```
s=u-x; y=(A'\ c)'; r=c-y * A; r=r+0.001 * (r==0); % PE 2004
```

```
z=r. * (r>0); w=z-r; gap=c * x-y * b+w * u;
```

```
it=0; % Start iterations
```

```
while (gap)>small & it<max_it
```

```
it=it+1; % Compute affine step
```

```
q=1./(z'./x+w'./s); r=z-w; Q=spdiags(sqrt(q),0,n,n);
```

```
AQ=A * Q; % PE 2004
```

```
rhs=Q * r'; % "
```

```
dy=(AQ' \ rhs)'; % "
```

```
dx=q. * (dy * A-r)'; ds=-dx; dz=-z. * (1+dx./x)'; dw=-w. * (1+ds./s)';
```

```
% Compute maximum allowable step lengths
```

```
fx=bound(x,dx); fs=bound(s,ds); fw=bound(w,dw); fz=bound(z,dz);
```

```
fp=min(fx,fs); fd=min(fw,fz); fp=min(min(beta * fp),1); fd=min(min(beta * fd),1);
```

```
% If full step is feasible, take it. Otherwise modify it
```

```
if min(fp,fd)<1
```

```
mu=z * x+w * s; % Update mu
```

```
g=(z+fd * dz) * (x+fp * dx)+(w+fd * dw) * (s+fp * ds);
```

```
mu=mu * (g/mu)^(3/(2 * n));
```

```
% Compute modified step
```

```
dxdz=dx. * dz'; dsdw=ds. * dw'; xinv=1./x; sinv=1./s;
```

```

xi=mu * (xinv-sinv); rhs=rhs+Q * (dxdz-dsdw-xi);
dy=(AQ'\rhs)'; dx=q * (A' * dy'+xi-r'-dxdz+dsw); ds=-dx;
dz=mu * xinv'-z-xinv' * z * dx'-dxdz'; dw=mu * sinv'-w-sinv' * w * ds'-dsw';
% Compute maximum allowable step lengths
fx=bound(x,dx); fs=bound(s,ds); fw=bound(w,dw); fz=bound(z,dz);
fp=min(fx,fs); fd=min(fw,fz); fp=min(min(beta * fp),1); fd=min(min(beta * fd),1);
end
% Take the step
x=x+fp * dx; s=s+fp * ds; y=y+fd * dy;
w=w+fd * dw; z=z+fd * dz; gap=c * x-y * b+w * u; % disp(gap);
end
function b=bound(x,dx) % Fill vector with allowed step lengths,Support function for lp_fnm
b=1e20+0 * x; f=find(dx<0); b(f)=-x(f)./dx(f);

```

#### 4 分位数回归的展望

分位数回归在理论方法与实证分析两个方面都形成了丰富的成果，现有成果的梳理展示出分位数回归是经济、金融、管理领域分析现实问题的有力工具，对工资收入、汇率变动、风险度量、环境评估、粮食产量等的影响因素分析不仅给出平均水平上的相关关系，还能进一步分析不同分布条件下相关关系的变化，能更好地满足现实需求中关注重点是尾部特征、极端特征等现象的探讨。基于回归分析研究的问题改用分位数回归则能够分析得出信息含量更多的结论，可见分位数回归的应用前景广阔，但值得注意的是分位数回归并不适用于任何场合，特别是小样本情形经常出现回归模型不显著等问题。同时，在大数据冲击而来的现在，分位数回归求解的技术也不再适用，大规模数据的应用而生<sup>[7]</sup>，如何适应现实需要不断深化分位数回归理论，拓展适用性更强的模型必将是未来的需要。

参考文献：

- [1] Koenker, R., G. Bassett. Regression quantiles [J]. *Econometrica*, 1978, 46 (1): 33-50.
- [2] Xiao, Zhijie. Time Series Quantile Regressions [M]. *Handbook of Statistics*, Elsevier, 2012.
- [3] Galvao, A. F., L. Wang. Efficient minimum distance estimator for quantile regression fixed effects panel data [J]. *Journal of Multivariate Analysis*, 2015, 133: 1-26.
- [4] Qu, Z., J. Yoon. Nonparametric estimation and inference on conditional quantile processes [J]. *Journal of Econometrics*, 2015, 185: 1-19.
- [5] Geraci, M., M. Bottai. Linear quantile mixed models [J]. *Stat. Comput.*, 2014, 24: 461-479.
- [6] Koenker, R. Quantile Regression [M]. *Econometric Society Monograph Series*, Cambridge University Press, 2005.
- [7] Fu, L., Y. Wang. Quantile regression for longitudinal data with a working correlation model [J]. *Comput. Statist. Data Anal.*, 2012, 56: 2526-2538.

### Quantile Regression: Development, Principle and Computer Implementation

Shi Jinfeng<sup>1</sup>, Liu Weiqi<sup>2</sup>

1. School of Economics and Management, Shanxi University, Taiyuan 030006, China;
2. Institute of Management and Decision, Shanxi University, Taiyuan 030006, China

**Abstract:** Quantile regression has been paid widely attention due to its loose assumptions and could show a more complete statistical relationship, forming rich research results. Based on quantile regression theory development



context, the application scope of quantile regression is analyzed, after introducing the basic principle and estimation method of quantile regression, detailed Windows operating steps in software such as Eviews, Stata and program in R, SAS, Matlab are supplied, and the future of quantile regression is proposed. It will provide guidance for the future development on theory and application of quantile regression.

**Key words:** Quantile regression; Eviews; R