Statistics and Application 统计学与应用, 2018, 7(5), 489-494
Published Online October 2018 in Hans. http://www.hanspub.org/journal/sa
https://doi.org/10.12677/sa.2018.75057

Hans汉斯

The Analysis of the Factors Influencing China's Gross Agricultural Output from the Perspective of Quantile Regression

Zhefan Hou

College of Sciences, The University of Mining and Technology, Beijing
Email: 18811138401@163.com

Received: Sep. 30th, 2018; accepted: Oct. 12th, 2018; published: Oct. 19th, 2018

Abstract

This paper briefly introduces quantile regression model, and compares it with least squares model on their theories and results, quantile regression model is more robust and more informative. Using the model with the *R* project package, this paper analyses the relationships between the gross agricultural output and five agricultural input factors on the latest data about our 31 provinces and cities from China Statistical Yearbook.

Keywords

Quantile Regression, Least Squares, Gross Agricultural Output

基于分位数回归的我国农业总产值影响因素分析

侯哲凡

中国矿业大学(北京)理学院，北京
Email: 18811138401@163.com

收稿日期：2018年9月30日；录用日期：2018年10月12日；发布日期：2018年10月19日

摘　要

本文对分位数回归模型的理论基础进行了简要介绍，并将其与线性回归的最小二乘法进行了算法和结果方面的比较，分位数回归模型更加稳健而且可以提供更完整的信息。在*R*语言软件包的辅助下，基于《2017

文章引用: 侯哲凡. 基于分位数回归的我国农业总产值影响因素分析[J]. 统计学与应用, 2018, 7(5): 489-494.
DOI: 10.12677/sa.2018.75057

年中国统计年鉴》的我国31个省市最新相关数据，运用分位数回归模型对影响我国农业总产值的五项投入值的关系进行分位数回归分析。

关键词

分位数回归，最小二乘法，农业总产值

1. 分位数回归的合理性

1.1. 线性回归与最小二乘法

假设因变量 Y 和自变量 X 之间存在相关关系，对 X 的各个确定值 x，随机变量 Y 都有对应的分布 $F(y/x)$。最小二乘回归方法用于确定 $E(y/x)$，即自变量 X 取值 x 时，与之相应的 Y 取值均值。假定 Y 的一组随机样本取值为 $y_1, y_2, \cdots, y_n$，最小二乘原理即寻找 α 使误差平方和最小，即满足：

$$\min_{\alpha \in R} \sum_{i=1}^{n} (y_i - \alpha)^2$$

该问题的解就是：$\alpha = \frac{1}{n} \sum_{i=1}^{n} y_i$。

1.2. 最小二乘法缺陷

最小二乘法广泛应用于经典线性回归的数据分析中，其原因在于最小二乘法的解释与人们的直观想象一致，同时该方法易于计算，手工即可完成，其优越性在前计算机时代是不言而喻的。而且当误差项服从正态分布时，其解具有无偏性、有效性等优良性质。但在实际问题中，应用最小二乘回归时需要满足较为严格的条件，如误差项同方差、自变量两两不相关等条件。当需要进行回归系数的显著性推断时，还需要假设误差项服从正态分布。当误差项的分布是重尾或有离群点情况时，其结果的稳健性较差。但在实际问题中完全满足这些基本假设的情况并不多见，一旦违背了某一项基本假设，就难以得到无偏的、有效的参数估计量。而且大量的数据仅仅只能得到包含有限信息的一条回归曲线，比较浪费。

为此，作为最小二乘法的拓展，分位数回归方法应运而生。相较于最小二乘法模型，分位数回归模型具有以下几个优势：首先，它对模型中的误差项不需做任何分布的假定，表现出较强的稳健性；其次，对条件分布的刻画更加细致，尤其能有效地分析数据分布中极端值的影响；还有，分位数回归通过使加权误差绝对值之和最小得到参数的估计量具有大样本理论下的渐近优良性。

1.3. 分位数回归原理[1]

分位数回归在自变量 X 取值 x 时，与之对应 Y 取值的任意分位数 $Q(y,\tau)$ 作为 $F(y/x)$ 的近似，与之对应的估计方法原理是寻找 α 使不对称加权绝对值之和最小，即满足：

$$\min_{\alpha \in R} \left\{ \tau \sum_{i:y_i \geq \alpha} |y_i - \alpha| + (1-\tau) \sum_{i:y_i \leq \alpha} |y_i - \alpha| \right\}$$

上式又可以表达为：

$$\min_{\alpha\in R}\sum_{i=1}^{n}\rho_{\tau}\left(y_i-\alpha\right)$$

其中 $\rho_{\tau}(\mu)=\mu\left(\tau\text{-}I(\mu<0)\right)$，$I(\cdot)$为简单示性函数。

现有样本 $y_1,y_2,\cdots,y_n$，其次序统计量为 $y_{(1)},y_{(2)},\cdots,y_{(n)}$。当 $n\tau>m$ 时，$\sum_{i=1}^{n}|y_i-\alpha|$ 是 α 的严格减函数；当 $n\tau<m$ 时，$\sum_{i=1}^{n}|y_i-\alpha|$ 是 α 的严格增函数，即当 $y_{(m)}<\alpha<y_{(m+1)}$ 时，最小不对称加权绝对值之和函数可表示为

$$\sum_{i=m+1}^{n}\tau y_{(i)}-\sum_{i=1}^{m}(1-\tau)y_{(i)}-(n\tau-m)$$

其图像当 $n\tau<m$ 时递减，当 $n\tau>m$ 时递增，从而，样本 $y_1,y_2,\cdots,y_n$ 的分位数 $Q(y,\tau)=y_{([n\tau])}$ 为满足要求的 α。

由以上分析可知，最小二乘回归确定的是自变量 X 取值 x 时，与之相应 Y 取值的均值 $E(y/x)$，而分位数回归确定的是给出自变量 X 取值 x 时，相应 Y 取值的各种分位点的估计，从而可以给出更多的信息，比如自变量 X 取值 x 时，相应 Y 取值的较大值(高分位点)或较小值(低分位点)与自变量 X 的相关关系的确定，模型能够更加全面详细地刻画出被解释变量条件分布的全部特征。该方法分位数回归适用范围更广，对模型没有苛刻要求，给出的统计信息更丰富，对于变量间的相关关系能够提供更加全面的分析结果，而且结果更加稳健，是普通回归分析的有益补充。

2. 实例分析我国 31 省市农业总产值与各项投入之间关系的分析[2]

2.1. 研究方法[3] [4]

重视农业的发展，确保农业经济的稳步发展对于中国经济发展至关重要。所以，“农村，农业，农民”问题近年来广受关注。而现发展阶段破解“三农问题”并且加快深化新农村建设步伐的关键在于提高地区农业生产的效益和效率。农业生产的增长主要依靠两种力量推动，第一种是依靠各种物质性投入的密集使用；第二种是依靠农业生产技术效率的改进，这种增长方式提高了技术在农业生产中的贡献份额。

为了研究影响我国农业总产值的主要因素，分析我国农业总产值对各项影响因素的依赖程度，本文采取典型相关分析法，选取农业总产出作为体现农业产出水平的主要指标，以农业劳动力投入，农业机械化程度，农业生产资料投入，农业土地投入，国家财政支出等作为影响农业生产水平的投入组指标，构建农业投入产出的典型相关模型。基于此选择建立回归模型，运用 *R* 软件对我国农业总产值及其相关影响因素的经济数据进行计量经济分析，得出不同水平农业总产值与其影响因素等各项数据的回归关系。分别建立线性回归模型和分位数回归模型，对其关系进行分析。

2.2. 回归模型分析[5]

一般地，农业产出水平主要体现为农业总产出(亿元)，用农业总产值表示，记为 y。影响农业生产水平的投入要素较多，结合各指标的实际经济意义，并参考国内的一些研究成果，通过筛选确定以下指标作为自变量组，称为“投入组”：1) 农业劳动力投入(万人)，用乡村人口数量表示，记为 X_1；2) 农业机械化程度(万千瓦)，用农业机械总动力表示，记 X_2；3) 农业生产资料投入(万吨)，用农用化肥施用折纯量表示，记为 X_3；(4)农业土地投入(千公顷)，用农作物总播种面积表示，记为 X_4；5) 国家财政支出(亿元)，用地方财政农林水事务支出表示，记为 X_5。本文选择目前使用最为广泛的描述投入产出关系的柯

布—道格拉斯生产函数作为研究的基本模型。

首先建立柯布－道格拉斯型农业投入与产出模型：

$$Y = AX_1^{\beta_1} X_2^{\beta_2} \cdots X_n^{\beta_n} \tag{1}$$

为便于应用最小二乘法进行多元线性回归，将式(1)两边取对数，转化为双对数的多元线性回归模型：

$$\ln Y = \ln A + \sum_{i=1}^{n} \beta_i \ln X_i \tag{2}$$

其中 Y 为产出量，$X_1, X_2, \cdots, X_n$ 为 n 个投入要素量，A 为综合技术水平，$\beta_1, \beta_2, \cdots, \beta_n$ 为各种生产要素的产出弹性。

最小二乘回归使我们在期望的意义上关于各投入量对各省农业总产值的影响有了一个基本的判断。条件分位数回归的结果则能帮助我们分析在各省不同农业发展水平上各因素的影响之间的差异和变化规律。

在构建的生产函数的基础之上，分别使用分位数回归和传统的最小二乘回归方法进行计量分析，并对两种估计方法的结果进行比较，以期全面了解各变量对农业生产总值的影响及其条件分位数的分布特征。

2.3. 数据来源及统计分析[3] [4]

为进行我国各省农业生产总值的影响因素得研究，本文选取了来自《2017 年中国统计年鉴》的数据，截取了包括 2016 年全国 31 个省、自治区、直辖市(不包括台湾省、香港特别行政区和澳门特别行政区)的农业生产总值(亿元)，乡村人口(万人)，农业机械总动力(万千瓦)，农用化肥施用折纯量(万吨)，农作物总播种面积(千公顷)，国家财政支出(亿元)这 6 个主要农业指标的数据。通过建立柯布—道格拉斯生产函数，并取其对数形式进行多元线性回归，对相应地区的农业生产总值进行解释。

用最小二乘法拟合得到的参数拟合结果见表 1。

通过 R 用分位数回归得到的拟合结果中，选取分位点 τ 为 0.1，0.2，0.3，0.3，0.5，0.6，0.7，0.8，0.9，共 9 个拟合结果，将各参数的估计值罗列见表 2。

各个系数 $\hat{\beta}_i(\tau)$ 与 τ 的关系见图 1 所示。

Table 1. Results of least squares regression fitting
表 1. 最小二乘回归拟合结果

截距	X_1	X_2	X_3	X_4	X_5
1.2484	0.5012	−0.2886	0.6031	0.0367	0.1637

Table 2. Results of quantile regression fitting
表 2. 分位数回归拟合结果

分位点	截距	X_1	X_2	X_3	X_4	X_5
0.1	−1.23	0.6245	−0.3789	0.6875	0.0858	0.012
0.2	1.73	0.4291	−0.3865	0.739	−0.0929	0.0993
0.3	1.61	0.4994	−0.3552	0.7759	−0.0154	0.0808
0.4	0.41	0.6106	−0.0283	0.4252	−0.1015	0.1545
0.5	0.49	0.5588	−0.1059	0.4199	−0.0171	0.1923
0.6	0.65	0.4676	−0.1252	0.4391	0.0898	0.1721
0.7	0.8374	0.3773	−0.0789	0.4693	0.1377	0.1318
0.8	0.9432	0.5812	−0.4652	0.4652	0.2818	0.2029
0.9	2.19	0.3228	−0.0461	0.5635	−0.0114	0.0799

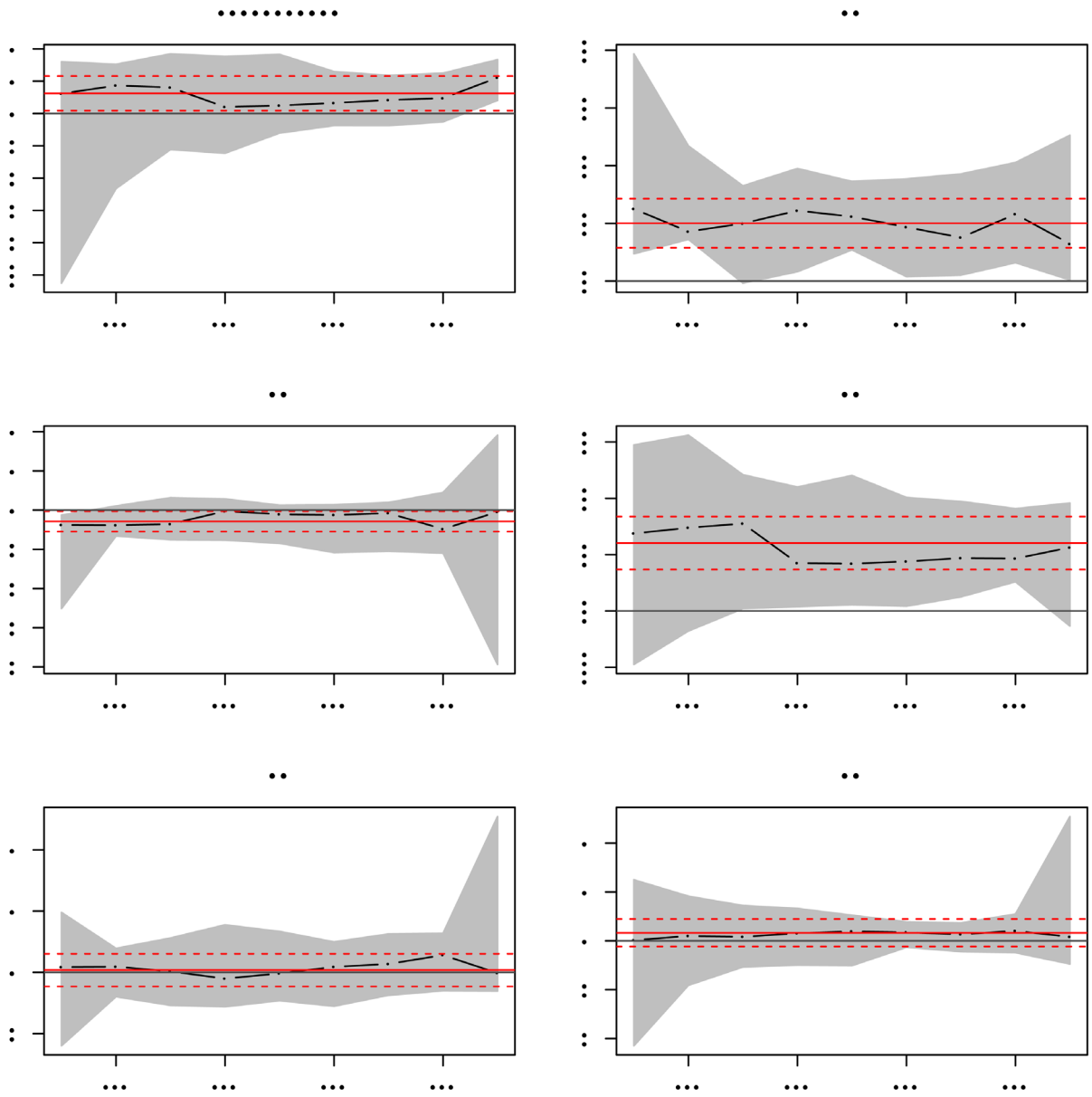

Figure 1. Curve: Coefficients of quantile regression fitting
图 1. 分位数回归拟合系数图

图 1 中点、线、坐标等元素含义与回归模型分析部分中解释一样。每个子图是其上面标注的变量(X_i对应$\ln X_i$)对应的系数的估计量与τ的关系。

由拟合结果和图 1 中的信息可知：$\ln A$ 解释为当各个投入要素都为 1 时的农业总产值，对于一个省来讲，投入 1 相当于零投入，对应的农业总产值不妨认为是无其他投入量时，各省当年自然环境状况自行可产生的农业生产总值的能力。最小二乘法估计的结果是任何时候，尽管零投入，仍有自然农业生产总值$e^{1.238173}$，这一结果符合常理，它表示自然环境正向影响力越大，农业总产值会而越高。分位数回归给出的结果可以得出，自然环境的生产能力对各省农业生产的影响是有很大差别的。对于农业比较落后的省份，自然环境对农业总产值不起积极作用，抑制其农业生产，符合常理。

对于变量乡村人口，其系数最小二乘估计值为正数结果符合常理。由分位数回归估计结果得出更详

细的结果。从拟合结果来看，各分位数下，系数均为正数，且有逐渐增大的趋势。这可以解释为：乡村人口增加相当于乡村劳动力增加，且在当今国家环境下，劳动力增加对各省的农业总产值有积极作用。

对变量农业机械总动力，其系数最小二乘估计为正值结果符合常理，农业机械总动力越多，其替代人工劳动力越多，促进农业生产效率提高，从而促进农业生产总值。详细情况由分位数回归分析表明,对任何农业发展的省增加农业机械量对该省的农业生产情况产生消极影响。这说明，2016 年全国农业生产水平已达到较饱和状态，再加大该方面投入只会造成机械过剩，造成浪费，反而降低农业总产值。并且对农业发展状况为中等情况的省消极影响较小，而对农业发展状况较差或较好的省的消极作用较为明显。

对变量农用化肥施用量，其系数的最小二乘法表明加大化肥投用量对农业总产值有积极作用，但这是一个较高的投入量，与化肥投用量过大会污染环境而要求减少化肥投用量的政策主旨相反；通过分位数回归分析，各分位点处农用化肥施用量系数虽均接近零值，但其图像的变动趋势仍呈现随分位数数值增加而减少的趋势，且在低尾处，农用化肥施用量对应的系数大于均值，在高尾处，低于均值。这说明，目前农业发展程度较低的省市自治区仍需要大量使用化肥来增加粮食产量并且化肥使用量增加仍会带来农业总产值增加量增大，而农业较发达的省份则已经有了较充足的化肥施用量或相关农业技术带来了较高的农业生产总值，所以加大化肥投入带来的农业总产值增幅有限。

对变量农作物总播种面积，其对应的系数的估计量均为正数，表示播种面积越大，种植粮食越多，从而收获越多，这是符合常理的。而且从图像观察，其图像基本围绕其的最小二乘估计值 0.12483 波动，没有明显的上升或下降趋势。这说明农作物总播种面积的变化对农业发达或不发达的省份的影响是一致的。

对变量国家财政支出，其对应的系数的最小二乘估计值为正值但比较小，这表明其对农业总产值的影响虽是正面的但影响度较小。通过分位数回归估计结果可以得出更多结果,目前国家财政拨款给农业不发达的地区带来的效益正向的，而对农业发达地区带来的效益已为负向。根据有关资料，产生负向影响的原因可能为：其一，财政支持对象不合理。其二，农业内部支持结构不合理，在中国农业投入中，用于流通环节的补贴太多。

3. 总结

分位数回归方法提出之初，由于其算法相较于经典回归分析比较麻烦，并没有被立即广泛使用。但随着计算机技术的发展其得到了迅猛的发展及应用，在理论和实践方法上都越来越成熟。本文给出了分位数回归原理的说明，并基于此对我国 31 个省市的农业总产值与其各项投入项之间的关系进行了两类回归分析，得出了分位数回归比只做最小二乘回归结果更加丰富的结论，这表明分位数回归可以基于已有数据得出更多合理的信息分析结果。

基金项目

中国矿业大学(北京)数据的回归分析的应用研究小组。

参考文献

[1] 乔舰, 李再兴. 分位数回归的理论再说明及实例分析[J]. 统计与决策, 2012(19): 104-107.

[2] 裴耀. 分位数回归及其应用[D]: [硕士学位论文]. 武汉: 华中师范大学, 2014.

[3] 常春华. 中国农业产值影响因素分析[J]. 农业经济与科技, 2006(8): 51-52.

[4] 冯启磊, 王红瑞, 白颖, 刘琼. 农业产出水平的影响因素分析[J]. 安徽师范大学学报(自然科学版), 2010(3): 276-280.

[5] 袁磊. 我国农业总产值影响因素研究[J]. 山东农业大学学报(社会科学版), 2013(3): 29-33.

知网检索的两种方式：

1. 打开知网页面 http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD
   下拉列表框选择：[ISSN]，输入期刊 ISSN：2325-2251，即可查询
2. 打开知网首页 http://cnki.net/
   左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：http://www.hanspub.org/Submission.aspx
期刊邮箱：sa@hanspub.org