

基于多项式回归模型的中国城乡收入差距变化特征与地区差异分析

王晓洁¹, 周文涛¹, 姜源源¹, 杨炼², 陈国华^{2*}

¹湖南人文科技学院研究生教育教学部, 湖南 娄底

²湖南人文科技学院数学与金融学院, 湖南 娄底

收稿日期: 2023年1月17日; 录用日期: 2023年2月7日; 发布日期: 2023年2月23日

摘要

为更好的探究中国城乡收入差距的变化特征与地区差异, 本文建立多项式回归模型, 将影响城乡收入差距的因素划分为内部因素和外部因素分别作为中介变量和控制变量, 将城乡收入差距作为被解释变量, 利用主成分分析法对数据进行处理降维得到代表数字经济的指数作为解释变量, 采用泰尔指数测算城乡收入差距, 以定量的方式分析中国城乡收入差距的变化特征以及地区之间的差异。研究发现: 内部因素和外部因素均与城乡收入差距存在显著相关性, 我国城乡收入差距在波动中呈逐步扩大的趋势并体现出阶段性特征, 城乡收入差距的地区差异较为明显, 地区偏远、环境恶劣、经济发展落后、乡村振兴战略实施难度较大的西部地区以及农业在当地所占的经济比重较高、经济欠发达的中部地区城乡收入差距相对大, 而经济较发达、平原广阔、物产丰富、交通便利、农业人口所占比重较低的东部地区城乡收入差距相对较小, 并且, 西部地区城乡收入差距扩大的速度较快, 中部地区的城乡收入差距在不断缩小。

关键词

城乡收入差距, 地区差异, 多项式回归模型

The Variation Characteristics and Regional Differences of Urban-Rural Income Gap in China Based on Polynomial Regression Model

Xiaojie Wang¹, Wentao Zhou¹, Yuanyuan Jiang¹, Lian Yang², Guohua Chen^{2*}

¹Department of Graduate Studies, Hunan University of Humanities, Science and Technology, Loudi Hunan

²School of Mathematics and Finance, Hunan University of Humanities, Science and Technology, Loudi Hunan

*通讯作者。

文章引用: 王晓洁, 周文涛, 姜源源, 杨炼, 陈国华. 基于多项式回归模型的中国城乡收入差距变化特征与地区差异分析[J]. 统计学与应用, 2023, 12(1): 121-127. DOI: 10.12677/sa.2023.121014

Abstract

In order to better explore the changing characteristics and regional differences of China's urban-rural income gap, this paper establishes a polynomial regression model. The factors affecting the urban-rural income gap are divided into internal factors and external factors as mediating variables and control variables respectively. The urban-rural income gap is used as the explained variable, and the principal component analysis method is used to process the data to reduce the dimension to obtain the index representing the digital economy as the explanatory variable. The Theil index is used to measure the urban-rural income gap, and the changing characteristics of China's urban-rural income gap and the differences between regions are analyzed quantitatively. The study found that there is a significant correlation between internal factors and external factors and the urban-rural income gap. The urban-rural income gap in China is gradually expanding in the fluctuation and reflects the stage characteristics. The regional differences in the urban-rural income gap are more obvious. The western regions with remote areas, harsh environment, backward economic development, and difficulty in implementing the rural revitalization strategy, as well as the central regions with high proportion of agriculture in the local economy and underdeveloped economy, have a relatively large urban-rural income gap, while the eastern regions with more developed economy, vast plains, rich products, convenient transportation, and low proportion of agricultural population have a relatively small urban-rural income gap. The income gap between urban and rural areas in the western region is expanding faster, and the income gap between urban and rural areas in the central region is shrinking.

Keywords

Urban-Rural Income Gap, Regional Difference, Polynomial Regression Model

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

自改革开放 40 年来,中国经济发展迅速,GDP 总量已经稳居世界第二,即使在全球经济增长乏力、复苏缓慢的后疫情时代,中国经济增速仍然保持在 5%以上,成为推动世界经济增长的重要动力。一直以来,国家积极推进城乡发展,党的十九大报告提出实施乡村振兴发展战略,标志着中国的乡村治理进入新阶段。党的二十大报告中强调全面推进乡村振兴,坚持农业农村优先发展。这一系列政策为加速破解传统城乡二元结构,缩短城乡收入差距,推动城乡一体化发展提供了新方向。据国家统计局数据显示,截至 2021 年,中国城镇化水平已经达到 64.7%,城镇居民人均可支配收入达到 47,412 元,同比增长率为 7.1%。目前,我国通过加速建设数字经济、结合乡村振兴等的一系列统筹城乡发展举措,缩小城乡收入差距,但城乡收入差距问题仍然比较突出,城乡收入差距水平仍然高于西方发达国家和绝大多数发展中国家,为维护我国社会和谐稳定,推动乡村振兴,缩小城乡差距问题刻不容缓。

本文选择中国城乡收入差距的典型实例,从数字经济与乡村振兴两个方面进行描述性统计分析,建立多项式回归模型,分析影响中国城乡收入差异的主要因素,并区分这些影响因素中哪些对中国城乡收

入差距产生积极(或消极)作用;由于我国明显的城乡二元经济结构,采用泰尔指数测算城乡收入差距,以定量的方式分析中国城乡收入差距的变化特征以及地区之间的差异。

通过探究中国城乡收入差距变化特征与地区差异,增进对城乡收入差距影响因素的认知,有助于进一步缩小城乡收入差距,对推动城乡均衡发展具有重要意义。

2. 数据预处理

2.1. 指标的选取

指标选取主要考虑政策依据、理论依据和现实依据三个方面,政策依据指的是党和政府对城乡收入差距的相关文件指导;理论依据指的是学者们对城乡收入差距影响因素已有的相关研究;现实依据主要考虑指标的可获得性和可操作性。本文主要借鉴靳贞来[1],李红[2],杨茜[3]等学者的指标选择,通过因素归类将影响城乡差距的因素归纳为内部影响和外部影响两个方面,并结合实际情况,并选取了9个统计指标作为影响城乡收入差距的因素(见表1)。

Table 1. Influencing factors of urban-rural gap

表 1. 城乡差距影响因素

影响因素	指标参考
内部因素	人力资本
	创新活力
	城镇化
外部因素	交通发达程度
	投资水平
	开放程度
	财政农业支出占比
	教育科研投入占比
	产业结构偏离度

2.2. 变量的选取

1) 被解释变量

城乡收入差距,由于我国明显的城乡二元经济结构,采用泰尔指数测算城乡收入差距[4]:

$$Theril = \sum_{j=1}^2 \left\{ \frac{I_{ij,t}}{I_{i,t}} \ln \left[\frac{I_{ij,t}}{I_{i,t}} / \frac{P_{ij,t}}{P_{i,t}} \right] \right\} \quad (1)$$

其中,表示城镇总收入,表示农村总收入(由人均可支配收入与人口数的乘积计算得到),泰尔指数越接近0,说明城乡收入差距越小,反之越大。

2) 解释变量

数字经济,查阅相关资料选择使用互联网普及率(百人中互联网宽带接入用户数量)、互联网行业从业人数(计算机和软件从业人员与总单位从业人员数的比值)、互联网相关产出(人均电信业务总和)和移动电话普及率(百人中使用移动电话人数)4个指标,运用主成分分析法标准化处理数据后降维得到代表数字经济的指数。

3) 中介变量

人力资本(人均受教育年限)、创新活力(专利授权总数的对数)、城镇化(城镇常住人口与总人口的比值)。

4) 控制变量

- a) 交通发达程度
- b) 投资水平
- c) 开放程度
- d) 财政农业支出占比
- e) 教育科研投入占比
- f) 产业结构偏离度

注：产业结构偏离度计算公式[5]：

$$\lambda = \lg \left\{ \left| \frac{\text{各产业产值比重}}{\text{各产业就业比重}} - 1 \right| + 1 \right\}$$

2.3. 数据来源

城乡居民可支配收入于 2013 年开始统计，故选取我国 30 个省份及自治区直辖市(除港澳台及西藏地区外) 2013~2021 年的省级面板数据，数据来自《中国人口与就业统计年鉴》、国家统计局、各省的统计局官网以及中国教育部网站等。下表 2 为各个变量的描述性统计结果。

Table 2. Descriptive statistics of variables
表 2. 变量描述性统计

	变量	样本量	均值	标准差	最大值	最小值
被解释变量	泰尔指数	346	0.097	0.102	0.64	0.001
解释变量	数字经济指数(替换)	346	37.26	14.45	105.48	23.51
	数字经济指数(基准)	346	30.12	13.80	89.21	14.32
中介变量	人力资本	346	12.87	0.43	20.07	2.39
	创新活力	346	4.31	2.31	9.54	1.86
	城镇化	346	1.22	1.58	5.72	0.65
控制变量	交通发达程度	346	0.98	2.32	3.98	0.29
	投资水平	346	1.74	1.28	9.61	0.22
	开放程度	346	2.79	0.99	4.10	0.08
	财政农业支出占比	346	12.33	2.47	22.07	5.83
	教育科研投入占比	346	17.21	3.73	27.30	6.92
	产业结构偏离度	346	0.32	0.51	0.94	0.06

3. 模型建立

3.1. 多元回归模型

多元回归模型是基于最小二乘法的基础上得到的自变量与因变量的相关关系，并以误差平方和最小为准则来估计[6] [7]。得到不同次数的多项式表达式，再参考标准差选择最合适多项式表达式来描述一定作用下影响因素与城乡收入差距的关系。其数学表达式为：

$$y(x, a) = \sum_{i=0}^h a_i x^i \tag{2}$$

其展开式为：

$$y(x, a) = a_0 + a_1 x + a_2 x^2 + \dots + a_h x^h \tag{3}$$

其中 a 为所求参数， h 为多项式次数。

回归模型的显著性检验：分析得到的各个自变量与因变量之间的相关关系，得到回归方程的相关系数，再分别对误差进行残差检验，综合可得到我们回归模型的显著性，通过对显著性的比较，可以说明模型对该题的适用性。

3.2. 基准回归模型

从乡村振兴和数字经济维度讨论城乡收入差距的影响建立基准回归模型[5]，具体公式如下：

$$Theril_{it} = \alpha_0 + \alpha_1 S_{it} + \theta X_{it} + \varepsilon_{it} \tag{4}$$

建立如下非线性模型，检验数字经济与城乡收入差距间是否为非线性关系：

$$Theril_{it} = \alpha_0 + \alpha_1 S_{it} + \alpha_2 S_{it}^2 + \theta X_{it} + \varepsilon_{it} \tag{5}$$

其中， $Theril_{it}$ 表示被解释变量， i 省份 t 时期城乡收入差距； S_{it} 表示核心解释变量， i 省份 t 时期数字经济的发展水平， X_{it} 为 i 省份 t 时期的一组控制变量； ε_{it} 表示随机扰动项。

3.3. 面板门限模型

将人力资本、创新活力、城镇化设为门槛变量，建立面板门限回归模型：

$$Theril_{it} = \alpha_1 S_{it} I(M_{it} \leq r_1) + \alpha_2 S_{it} I(r_1 < M_{it} \leq r_2) + \dots + \alpha_n S_{it} I(r_{n-1} < M_{it} \leq r_n) + \alpha_{n+1} S_{it} I(M_{it} > r_n) + \theta X_{it} + \mu_i + \varepsilon_{it} \tag{6}$$

其中， M_{it} 表示人力资本、创新活力、城镇化在 i 省份 t 时期的门槛变量； $\alpha_1, \alpha_2, \dots, \alpha_n$ 表示回归系数； I 为示性函数，若括号内条件得到满足，则 $I=1$ ，反之，则 $I=0$ 。

4. 模型求解

4.1. 城乡收入差距影响因素分析

由主成分分析，可从这些影响因素中提取出两个主成分，这两个主成分作为自变量，城乡收入差距作为因变量，将数据进行标准化处理，得到多元回归方程：

$$W = 4.21 \times 10^{-22} + 0.329 F_1 + 0.907 F_2 \tag{7}$$

模型的总体回归显著， F_1 、 F_2 与 W 正相关，回归系数显著，即外部因素和内部因素对城乡收入差距存在明显的扩张作用，将得到的表达式带入回归方程：

$$W = 4.21 \times 10^{-22} - 0.125 X_1 - 0.018 X_2 - 0.374 X_3 + 0.121 X_4 + 0.187 X_5 - 0.045 X_6 - 0.570 X_7 - 0.106 X_8 + 0.358 X_9 \tag{8}$$

结果表明，乡村振兴战略的大部分举措对缩小城乡收入差距具有正效应关系，属于积极的影响因素。通过研究发现，内部因素对缩小中国城乡收入差距具有积极作用；外部因素中的交通发达程度、投资水平、开放程度、财政农业投入、教育科研投入等因素对缩小中国城乡收入差距具有积极作用，但加快产业结构偏离则会扩大城乡收入差距；当投资水平在偏向城市时会扩大城乡收入差距，提高城市居民收入

水平，结果反而不利于缩小城乡收入差距；教育是立国之本，教育的发展很大程度上体现了一个地区经济的繁荣或衰落，乡村振兴战略加大了对农村教育的投入和财政农业支出的比重，对缩小城乡收入差距有积极的作用；“想要富，先修路”，说明交通发展是一个地区经济繁荣的前提，同时便利的交通能加快城乡经济的发展，在同步发展的前提下，城乡收入差距也在缓慢扩大。

4.2. 实证分析

为反映我国城乡收入差距的地区差异，根据区域划分，北京市、天津市、河北省、辽宁省、山东省、江苏省、上海市、浙江省、福建省、广东省和海南省为东部区域；黑龙江省、吉林省、山西省、河南省、湖北省、湖南省、江西省和安徽省为中部区域；云南省、广西省、贵州省、青海省、宁夏省、陕西省、甘肃省、内蒙古自治区和新疆省为西部区域。根据泰尔指数计算公式得出各区域不同时期的泰尔指数(见表 3)。

Table 3. Theil Index for the Eastern, Central and western regions 2017~2021

表 3. 东、中、西部区域 2017~2021 年的泰尔指数

		2017	2018	2019	2020	2021
东部	均值	0.091	0.093	0.074	0.063	0.082
	标准差	0.083	0.142	0.241	0.852	0.291
中部	均值	0.125	0.107	0.096	0.099	0.128
	标准差	0.181	0.522	0.324	0.631	0.329
西部	均值	0.210	0.098	0.121	0.172	0.084
	标准差	0.024	0.473	0.489	0.378	0.283

由上表分析，我国城乡收入的差距并不呈现倒“U”型，在经历最下方的拐点后，目前城乡收入差距在波动中呈现扩大的趋势；我国城乡收入差距具有阶段性变化的特征，利用基准分析模型和门限面板回归模型结合相关发展经济学的理论及库兹涅兹提出的倒“U”型分析得出，城乡收入差距随着国家的发展反而扩大，这是各发展中国家遇到的共同问题。在收入水平较低时，城乡收入差距表现得并不是非常明显；随着收入的升高，城市居民收入增长趋势比农村居民收入增长趋势强，这是由于农村经济结构比较单一，难以跟上城市互联网经济的进程。当所有居民均处于高收入水平时，城乡收入的差距才会逐渐缩小。

反观不同地区的泰尔指数，我国城乡收入差距的地区差异较为明显，以 2017 年为例，东部地区平均泰尔指数为 0.091，中部地区平均泰尔指数为 0.125，西部地区平均泰尔指数为 0.210。这说明西部地区的城乡收入差异远高于东、中部地区，这是由于西部地区范围广阔，地区偏远，环境恶劣，经济发展落后，乡村振兴战略实施难度较大；中部地区的城乡收入差异高于东部，东部属于沿海地区，平原广阔，物产丰富，交通便利，也是外商投资企业的首选，东部地区城市的发展也带动了农村的发展，所以东部地区的城乡收入差距扩大速度缓慢；中部地区大多都是山区地带，农业在当地所占的经济比重较高，但随着近几年来数字经济和乡村振兴举措的实施，中部地区的城乡收入差距在不断缩小。

5. 结论

本文基于多项式回归模型来分析中国城乡收入差距变化特征与地区差异。研究发现，内部因素和外部因素均与城乡收入差距存在显著相关性，其中，内部因素对缩小中国城乡收入差距具有积极作用；外部因素中的交通发达程度、投资水平、开放程度、财政农业投入、教育科研投入等因素对缩小中国城乡

收入差距具有积极作用，而产业结构偏离则会扩大城乡收入差距。

基于我国东部、中部以及西部地区 2017~2021 年的泰尔指数，利用基准分析模型和门限面板回归模型结合相关发展经济学的理论及库兹涅兹提出的倒“U”型分析发现，我国城乡收入差距在波动中呈逐步扩大的趋势并体现出阶段性特征。

我国城乡收入差距的地区差异较为明显，地区范围广阔、地区偏远、环境恶劣、经济发展落后、乡村振兴战略实施难度较大的西部地区以及农业在当地所占的经济比重较高、经济欠发达的中部地区城乡收入差距相对大，而经济较发达、平原广阔、物产丰富、交通便利、农业人口所占比重较低的东部地区城乡收入差距相对较小，并且，西部区域城乡收入差距扩大的速度较快，中部地区的城乡收入差距在不断缩小。由此说明，在科学发展观的指导下，政府采取提高农民收入、推动乡村振兴、缩小城乡收入差距的一系列措施具有较强的科学性与针对性。

致 谢

感谢湖南人文科技学院数学与金融学院对本文的支持。

基金项目

湖南省学位与研究生教育改革研究项目“专业学位硕士研究生创新人才培养的数学建模‘两轮驱动’模式研究与实践”(湘教通[2019] 293 号 No. 2019YGYB67)。

参考文献

- [1] 靳贞来. 城乡居民收入差距变动及其影响因素的实证研究[D]: [博士学位论文]. 南京: 南京农业大学, 2006.
- [2] 李红, 梁炳礼, 龙雨. 对外开放对城乡收入差距的空间溢出效应研究[J]. 华东经济管理, 2019(11): 86-93.
- [3] 杨茜, 石大千. 交通基础设施、要素流动与城乡收入差距[J]. 南方经济, 2019(9): 35-50.
- [4] 程名望, 张家平. 互联网普及与城乡收入差距: 理论与实证[J]. 中国农村经济, 2019(2): 19-41.
- [5] 李豫新, 李枝轩. 乡村振兴背景下数字经济发展与城乡收入差距[J]. 金融与经济, 2022(6): 60-67+96.
- [6] 孙毅, 刘仁云, 王松, 冷晓冰, 臧雪柏. 基于多元线性回归模型的考试成绩评价与预测[J]. 吉林大学学报(信息科学版), 2013, 31(4): 404-408.
- [7] 司守奎, 孙玺菁. 数学建模算法与应用[M]. 北京: 国防工业出版社, 2016.