

文章编号: 1002-1566(2019)04-0704-15
DOI: 10.13860/j.cnki.sltj.20190529-002

教育缓解了收入分配不平衡吗

许永洪^{1,2} 萧珍丽¹ 朱建平^{2,3}

(1. 厦门大学经济学院, 福建 厦门 361005; 2. 厦门大学数据挖掘研究中心, 福建 厦门 361005;
3. 厦门大学管理学院, 福建 厦门 361005)

摘要: 本文从教育扩展与教育不均两方面, 探究教育对城乡收入差距的影响, 利用 1997-2016 年的省级面板数据, 采用固定效应的面板模型进行分析。结果表明, (1) 研究发现教育扩展与收入差距之间的关系不符合“倒 U”曲线关系, 而是“正 U”曲线关系。早期的普适性教育政策缓解了收入分配的不均, 但是近年来由于基础教育和高等教育质量的分化, 社会整体教育水平提高起到了加剧收入分配不平衡的效果; (2) 短期内, 中国教育不均与收入差距之间是负相关关系, 说明目前我国教育不均的改善 (恶化), 并没有起到减缓 (加剧) 收入差距的作用, 但是长期看, 要特别注意教育不均的社会问题, 避免为未来社会分配埋下隐患。本文还发现在大学教育从精英教育向大众化教育再向普及化教育转变的背景下, 现行的教育统计口径无法准确衡量受教育差距和人力资本差距。

关键词: 教育扩展; 教育不均; 收入分配不平衡

中图分类号: C81, O212

文献标识码: A

Did Education Relief Unfairness of Income Distribution

XU Yong-hong^{1,2} XIAO Zheng-li¹ ZHU Jian-ping^{2,3}

(1. School of Economics, Xiamen University, Fujian Xiamen 361005, China, 2. Data Mining Research Center, Xiamen University, Fujian Xiamen 361005, China, 3. School of Management, Xiamen University, Fujian Xiamen 361005, China)

Abstract: This paper explores the influence of education on urban-rural income's gap from perspectives of education's expansion and educational inequality, and adopts the fixed effects model based on provincial panel data from 1997 to 2016. The results show that: (1) This research finds that the relationship between education's expansion and urban-rural income's gap does not present the inverted "U" curve, but a positive "U" curve. The early universal education policy alleviated the inequality of income distribution, but in recent years, due to the differentiation of the quality of basic education and higher education, the improvement of education level of society as a whole has played a role in aggravating the imbalance of income distribution. (2) In the short term, there is a negative correlation between educational inequality and urban-rural income's gap in China, indicating that the improvement (deterioration) of educational inequality has not narrowed (aggravated) the income gap. However, in the long term, special attention should be paid to the social problems of educational inequality to avoid potential dangers for future social

收稿日期: 2018年7月31日

收修改稿日期: 2019年3月20日

基金项目: 本文受国家社科基金重大项目 (13&2D148), 国家社科基金青年项目 (17CTJ007) 资助。

distribution. It is also found that in the context of the transformation of college education from elite education to mass education and then to popular education, current education statistical coverage cannot accurately measure the educational gap or the human capital gap.

Key words: education's expansion, educational inequality, unfairness of income distribution

0 引言

收入分配不平衡已经成为我国经济社会发展的尖锐矛盾，也是国家扶贫重大战略政策的原因所在。自 1978 年改革开放以来，中国经济经历了近四十年高增长，居民平均生活水平明显提高，但是也产生了一些社会问题。以城乡居民人均收入为例，该指标从 1978 年的 343.6 元、133.6 元增长至 2016 年的 33616 元、12363 元，二者的比值从 2.57 上升到 2.72，个别年份城乡收入比最高曾达 3.3。尽管一些年份城乡收入差距出现短暂性缓和，但是差距的整体趋势呈扩大化，如果考虑到三十多年来财富的累积效应和城市化进程中土地和房价价值的重估差异，城乡居民财富的差距比收入差距更加悬殊。国家统计局的数据表明，我国 2016 年的基尼系数高达 0.465，尽管近年来，我国基尼系数略有下降，但均超过了国际警戒线 0.4，远高于发达国家 0.24-0.36 的数值表现。巨大的收入差距已成为中国经济与社会发展不容忽视的羁绊。

教育的不均成为日益严重的社会问题。近年来，随着经济的发展和社会的进步，政府和人民群众对教育的重视程度较之以往都有明显提高。政府层面九年义务教育制度的强制实施以及一系列加强措施推动了教育水平的全面提升。然而伴随着教育的扩展程度（教育的广度与深度）整体提高，城乡教育方面的差距也愈演愈烈。在我国，由于农村对教育的重视程度偏低、农村师资匮乏以及对农村的教育投入普遍低于城镇等一系列问题导致农村的平均受教育程度明显低于城镇，成为农村发展的重大制约因素。

教育与收入的密切关系已成共识，学术界对于二者之间的作用机制一直没有定论。一方面，教育是家庭对子女的投资，对子女的未来收入高低有直接影响。受教育程度的高低会直接决定个人能从事劳动的复杂程度和精密程度。受教育程度高的群体会更容易进入高收入群体，因为他们能够从事更为复杂和精密的工作，并且更具有不可替代性。相反，受教育程度低的人群往往只能从事一些相对比较没有门槛的简单工作，容易被替代。他们对工作的选择面也很窄，更容易面对失业等问题。另一方面，教育作为居民的支出内容之一，也受到个人收入高低的直接影响。收入较高的群体会有更多资金用于子女的教育方面的投入，而这样又很可能又会加大收入差距。教育是形成人力资本的主要要素，它直接影响了个人的收入水平，那么是否可以通过提高整个社会的教育水平和教育均衡程度来缓解甚至消除城乡收入差距？

基于上述思考，本文旨在从教育方面入手，基于 1997-2016 年的省级面板数据，建立回归模型，探讨教育扩展水平与教育不均两个主要影响因素对收入不平衡的影响。后续篇幅中：第 1 节是文献综述，主要阐述了前人有关于教育扩展、教育不均与收入分配不平衡间关系的研究成果；第 2 节是研究设计和变量设置，解释各变量的选取和参数的构建方法；第 3 节是数据来源和处理，由于一些基础数据的缺失和统计口径变化，本文进行了大量的统计数据口径调整和指标计算；第 4 节是实证，展示了基于省级面板数据进行的实证建模以及实证结果分析，并进行稳健性检验；第 5 节是结论与启示。

1 文献综述

最早对教育和收入关系进行研究的主要是国外的一些学者，他们大多数利用不同国家的宏观数据进行研究。然而他们关于教育扩展、教育不均和收入分配不平衡研究结论仍有很大

争议,一方面他们对教育扩展、教育不均和收入分配不平衡采用的衡量指标有很大差别,另一方面三者之间的关系也一直没有定论。Becker 和 Chiswick (1966)^[1] 发现美国地区间的收入差距与教育不均呈正相关关系,与平均受教育程度呈负相关关系。Chiswick (1971)^[2] 研究了 9 个国家的截面数据,发现教育的离散程度与收入差距呈负相关关系。Tinbergen (1972)^[3] 利用美国、加拿大和荷兰的数据,分析发现教育扩展和教育差距均会对收入分配有显著作用。教育水平下降,收入分配不平衡也减小;平均受教育程度增大,收入差距将会减小。Malrin 和 Psacharopoulos (1976)^[4] 以收入取对数后的方差衡量收入分配不平衡,通过对美国数据的研究,发现受教育年限每增加一年,收入分配不平衡度会下降 10%。但 Psacharopoulos (1977)^[5] 使用 49 个国家数据却发现教育不均变量对收入分配一直有负的影响,但是回归方程中是否纳入教育不均,平均受教育年限对收入分配差距的影响却完全不同。Winegarden (1979)^[6] 用受教育年限取对数后的均值和方差分别表示教育扩展和教育不均,对 32 个国家的数据进行研究,发现教育扩展水平越高,收入最低的 80% 人口收入份额占比越大;但教育差距越大,收入最低的 80% 人口收入份额占比越小。Tilak (1989)^[7] 对 50 个国家数据分析后发现,初等教育和中等教育的入学率与收入最低以及收入中等的 40% 人口收入份额占比正相关,而与收入最高的 20% 人口收入份额占比负相关。Bourguignon 和 Morrisson (1990)^[8] 使用与 Tilak (1989)^[7] 类似的指标,利用 35 个发展中国家数据进行了回归分析,并得到了相似的结论,即中等教育入学率与收入最低的 40% 人口收入所占份额正相关。Ram (1990)^[9] 使用 94 个国家的截面数据进行回归,首次发现平均受教育年限和教育不均程度之间存在“倒 U”型关系,且在平均受教育年限为 7 年左右达到峰值点。Park (1996)^[10] 考虑收入的“倒 U”型假说,利用 59 个国家的数据建模,发现平均受教育年限越高,收入分配会更均衡;同时教育不均程度越大,收入差距会拉大。赖德胜 (1997)^[11] 对 49 个国家数据建模也验证了教育扩展与收入分配差距存在库兹涅茨“倒 U”型关系。Gregorio 和 Lee (2002)^[12] 用 15 岁以上人口的平均受教育年限代表教育扩展,用平均受教育年限的标准差代表教育不均,以基尼系数代表收入分配不平衡,对 100 多个国家的时序截面数据进行 SUR (seemingly unrelated regression) 估计。估计结果表明教育不均都会加大收入分配不均,并且教育不均与滞后五年的教育扩展之间有“倒 U”型关系。

就中国而言,三者关系同样没有定论。教育的投入与收益并非简单线性关系(杜洁瑞,崔霞,2017^[13]),白雪梅(2004)^[14] 利用 Becker 和 Chiswick (1966)^[1] 的人力资本模型,利用全国的时间序列数据建模,发现教育扩展与收入差距之间存在“倒 U”型关系。王小鲁、樊纲(2005)^[15] 用中国大陆 30 个省、自治区和直辖市的 1996-2002 年的省级面板数据建模,发现人均受教育水平和城镇基尼系数有显著的正相关关系,但是对乡村基尼系数没有显著作用,对于城乡收入差距具有不显著的负向作用。陆铭、陈钊(2005)^[16] 用联立方程组和分布滞后模型对 1987-2001 年中国省级面板数据进行收入差距、投资、教育和经济增长的相互影响研究发现,收入差距对教育的作用较小,但是其累积作用一直为正,且两者之间并不是简单的线性关系。杨俊、黄潇和李晓羽(2008)^[17] 指出教育与收入分配会相互影响,他们利用 1996-2004 年的省级面板数据,在考虑了分配滞后模型的基础上,建立了联立方程组模型。研究表明收入分配不平衡会促进教育不均,但是教育差距缩小,收入分配差距并没有相应的减小,并且两者并不是直接的线性关系;另外教育扩展有利于教育和收入不平衡的改善。之后王艳真和李秀敏(2015)^[18] 又对其模型进行了改进,把教育扩展看成是内生变量,以此建立了包含教育不均、教育扩展与收入分配差距三个内生变量的联立方程组模型,并利用 1996-2012 年中国的时间序列数据进行分析。她们发现教育不均对收入差距的长期影响为正;教育不均和收入分配差距均会减少教育扩展程度;教育扩展和收入分配差距有“倒 U”型关系,教育扩展会缩小教育不均。

大部分文献使用基尼系数作为收入不均的衡量指标,但是由于统计口径的变化,和数据的缺失,通过繁琐计算后的收入基尼系数仍然存在较大误差(阮敬、丁琳和纪宏,2018^[19]),陈昌兵(2007)^[20]给出一些估算的方法,但是部分年份仍然无法填补,类似的其他指标也存在类似问题。财富的分化已经成为城乡、不同城市间和不同群体间撕裂的现状,那么教育水平的整体提高是否会改善现状,教育不均的存在是否又会成为未来中国城乡、不同城市和群体间不可跨越的鸿沟?教育扩展与收入分配差距存在的库兹涅茨“倒U”型关系是否依然存在?这些问题需要寻找更可靠的变量和更细致的证据去实证。

2 计量模型与变量设置

2.1 模型设定

为了研究教育扩展水平和教育不均对收入差距的影响,本文拓展了Becker和Chiswick(1966)^[1]提出的人力资本实证模型,该模型是研究教育、人力资本和收入不平等常用的基准模型,扩展后的模型为:

$$INEQ_{it} = \alpha_0 + \alpha_1 EDINEQ_{it} + \alpha_2 EDU_{it} + \alpha_3 EDU_{it}^2 + \alpha_4 LNY_{it} + \alpha_5 LNY_{it}^2 + \alpha_6 URBAN_{it} + \alpha_7 TRADE_{it} + \alpha_8 GR_{it} + \alpha_9 AGRI_{it} + \alpha_{10} EDIN_{it} + \varepsilon_t, \quad (1)$$

式(1)中, $INEQ$ 表示衡量收入分配不平衡程度的指标; $EDINEQ$ 是衡量教育不均程度的指标; EDU 是衡量教育扩展水平的指标,加入平方项考虑了教育拓展的“倒U形”效应(Gregorio和Lee,2002^[12];白雪梅,2004^[14]等); LNY 为人均GDP的自然对数; $URBAN$ 为城市化水平; $TRADE$ 为进出口总额占GDP的比例,用以衡量地区经济开放程度是否会对收入差距产生影响; GR 为经济增长率; $AGRI$ 为政府对农业的财政支出占一般财政支出总额的比例,用以衡量政府政策对农业的支持是否会对收入差距产生影响; $EDIN$ 为政府对教育的财政支出占总支出的比例(参见吴延科、田茂再,2016^[21];周华等,2018^[22])等研究。

2.2 变量说明

2.2.1 收入分配不平衡($INEQ$)的衡量

常见的用来度量城乡收入差距的指标有三种:(1)城镇居民可支配收入与农村居民纯收入的比值,该指标概括的表示城乡居民收入的总体差距水平,不能细致的反映整个社会收入分配的不均衡程度;(2)基尼系数,衡量收入分配差距的基本工具。但是基尼系数的变化不能反映个别阶层的收入分配变动情况,因此很难区分基尼系数的扩大是因为各阶层普遍收入增加还是因为收入低的变得更低、收入高的变得更高。另外目前对基尼系数也没存在一个普遍接受且口径统一的计算公式,因此在实际操作中很难得到一个统一的标准;最后由于部分省份数据欠缺很多,尽管前人如陈昌兵(2007)^[20]给出一些估算的方法,但是部分年份仍然无法填补,并且能估算的部分年份口径也有不同,计算繁琐的同时,误差也相对较大;(3)泰尔指数,即用各地区的收入份额与人口份额之比的对数为权重对收入份额的加权平均和来衡量国家地区间的收入差距。因为该指数不仅充分考虑到人口的作用,而且对阶层两端收入的变动比较敏感,因此能很好辨别泰尔指数变化是由低阶层收入变动引起,是由高阶层收入变动引起,还是由城乡人口比例变动引起(见Winegarden,1979^[6])。前人大部分使用基尼系数为衡量指标,相对较少使用泰尔指数衡量城乡居民收入差距,又基于本文比较分析,泰尔指数相比于基尼系数更直接反映城乡人口比重,更清晰说明城乡收入差距的变动具体是由什么因素变动引起,

能更敏感的反映城乡两端收入的变动,因而本文选用更具针对性的泰尔指数作为衡量城乡收入差距的指标,更能反映城乡两者收入变化及差别。

$$TL_{it} = \sum_{j=1}^2 \frac{\frac{In_{ijt}}{In_{it}}}{\ln\left(\frac{In_{ijt}}{In_{it}} / \frac{P_{ijt}}{P_{it}}\right)} = \frac{\frac{In_{i1t}}{In_{it}}}{\ln\left(\frac{In_{i1t}}{In_{it}} / \frac{P_{i1t}}{P_{it}}\right)} + \frac{\frac{In_{i2t}}{In_{it}}}{\ln\left(\frac{In_{i2t}}{In_{it}} / \frac{P_{i2t}}{P_{it}}\right)} = \frac{I_{i1t}}{I_{it}} \frac{\frac{P_{i1t}}{P_{it}}}{\ln\frac{I_{i1t}}{I_{it}}} + \frac{I_{i2t}}{I_{it}} \frac{\frac{P_{i2t}}{P_{it}}}{\ln\frac{I_{i2t}}{I_{it}}}. \quad (2)$$

其中, TL_{it} 表示第 i 个地区 t 时期的泰尔指数; $j = 1, 2$, 分别表示城镇和农村地区, P_{ijt} 表示第 i 个地区 t 时期城镇 ($j = 1$) 或农村 ($j = 2$) 的人口数量, P_{it} 表示 t 时期 i 地区的总人口, 为了和城镇化率的计算口径保持一致, 这里总人口表示常住人口; In_{ijt} 表示 t 时期 i 地区城镇 ($j = 1$) 或农村 ($j = 2$) 的总收入 (用相应的人口和人均收入的乘积来表示), In_{it} 表示 t 时期 i 地区的总收入; I_{ijt} 表示 t 时期 i 地区城镇居民人均可支配收入 ($j = 1$) 或农村居民人均纯收入 ($j = 2$); I_{it} 表示 t 时期 i 地区的人均收入; 这里 $\frac{P_{i1t}}{P_{it}}$ 即下文的城镇化率, $\frac{P_{i2t}}{P_{it}}$ 即单位 1 减去城镇化率。推导过程中用到 $In_{it} = I_{it}P_{it}$, $In_{i1t} = I_{i1t}P_{i1t}$, $In_{i2t} = I_{i2t}P_{i2t}$ 。

2.2.2 教育扩展 (EDU) 与教育不均 (EDINEQ)

教育扩展能够增加受教育人数, 提高个体的受教育年限, 从而提高总体的平均受教育年限。本文用平均受教育年限衡量教育扩展水平。平均受教育年限详细的计算公式见式 (3), 采用以不同受教育程度的人口比重为权数对受教育程度加权平均得到。现有的文献中, 用来衡量教育不均的方法主要由两个: 一是用来代表教育的绝对离散程度的平均受教育年限的标准差或方差; 二是建立在衡量公平程度的基尼系数之上的教育基尼系数。因为相比于标准差, 教育基尼系数更能反应同一地区的不同受教育层次所占的比例, 且目前已经被大部分学者认同和采用, 本文用基尼系数测算方法作为教育不均程度的衡量, 见式 (4)。

$$EDU = \sum_{i=1}^n p_i y_i, \quad (3)$$

$$EDINEQ = \frac{1}{EDU} \sum_{i=2}^n \sum_{j=1}^{i-1} p_i |y_i - y_j| p_j, \quad (4)$$

其中 EDU 表示平均受教育年限, $EDINEQ$ 表示教育基尼系数, y_i 、 y_j 代表不同程度的受教育年限, p_i 、 p_j 代表相应受教育年限人口占总人口比重, n 为教育获得的分组数量。

2.2.3 控制变量

为了使模型的结果更稳健可信, 根据前人对影响收入差距的因素研究的结果, 本文选取的控制变量有: 人均 GDP 的自然对数 ($LN Y$); 城市化水平 ($URBAN$); 经济开放程度 ($TRADE$); 经济增长率 (GR); 政府对农业的支持 ($AGRI$); 政府教育投入水平 ($EDIN$)。

现有研究已经证明经济水平与收入差距之间也存在很强的相关关系, 并且由库兹涅茨提出的经济水平与收入差距之间存在“倒 U 型”关系也得到了许多学者的证实 (如白雪梅, 2004^[14])。本文采用人均 GDP 衡量经济水平, 并且对其取自然对数, 消除异常波动带来的影响, 用 $LN Y$ 表示。为验证“库兹涅茨”曲线是否在成立, 同样在方程 (1) 中加入人均 GDP 自然对数的平方项 ($LN Y^2$) 这一控制变量进行验证。

在经济发展影响因素这一类别中, 除了人均 GDP 这一项外, 也有学者验证了城市化水平 ($URBAN$) 与经济开放程度 ($TRADE$) 这两项对收入差距也会产生影响, 如杨俊等 (2008)^[17] 实证认为城市化水平 ($URBAN$) 的提高反而加剧了收入差距。因而本文在方程 (1) 中加入控制变

量城市化水平 (*URBAN*) 与经济开放程度 (*TRADE*) 进行探究, 其中经济开放程度 (*TRADE*) 采用一个地区的进出口总额来衡量。

3 数据来源与处理

由于核心变量均无法直接获得而需要自行计算, 因此变量的数据处理尤为重要, 实证所使用的基础数据来源于各年的《中国统计年鉴》与《中国教育统计年鉴》, 国家统计局官网及各地的地方统计年鉴, 并用 Stata12.0 软件进行实证。由于数据来源差异和不同年份年鉴间统计口径不同, 对于数据的特殊处理如下:

3.1 收入分配不平衡的泰尔指数

由于各省的城镇居民可支配收入与农村居民纯收入口径并不完全一致, 基于口径统一性的考虑, 数据来源以国家年鉴为准。1997–2013 年的城镇居民可支配收入和农村居民人均纯收入数据直接来源于 1998–2014 年的《中国统计年鉴》中“人民生活”章节下的“分地区城镇居民人均收入来源”和“分地区农村居民人均纯收入”。因为从 2012 年第四季度起, 国家统计局对城乡住户调查实施了一体化改革, 对城乡居民收入指标名称、分类和统计标准实行了统一的标准。改革后城镇居民人均可支配收入名称不变但其统计口径发生变化, 农村居民人均纯收入名称变为农村居民人均可支配收入且统计口径发生变化。从 2015 年《中国统计年鉴》能同时取得 2013 年和 2014 年的城乡居民收入数据为按照城乡一体化改革后新口径统计的数据, 因为 2013 年的城乡收入新旧口径数据都具备且完整, 所以本文统一以 2013 年为换算基点按比例把 2014 年的收入数据换算成旧口径, 换算方法为: 2014 年旧口径 = 2014 年新口径 / 2013 年新口径 × 2013 年旧口径。如天津市 2014 年的城镇居民人均可支配收入新口径数据是 31506.0 元, 天津市 2013 年的新旧口径的城镇居民人均可支配收入数据分别是 28979.8 元和 32293.57 元, 则 2014 年天津市以新口径计算的城镇居民人均可支配收入是 $31506.0 / 28979.8 \times 32293.6$, 即为 35108.63 元。按此方法, 则可以得到 2014 年所有地区用旧口径衡量的城乡居民人均收入数据, 同理得到 2015 年与 2016 年的数据。

西藏自治区 1997 和 1998 年的城镇居民人均可支配收入的数据缺失, 但是由于国家数据和西藏自治区的数据口径差距太大, 直接填补则导致数据不连贯, 所以本文将推算得出这两个数据。推算方法为: 把 1997 年数据当成 1996 和 1998 年的平均值, 1998 年的数据当成 1997 和 1999 年的平均值, 则可以同时推算出 1997 年和 1998 年西藏的城镇居民人均可支配收入。

由于《中国统计年鉴》的地区总体的人均可支配收入只有近两年的数据, 数据不全面。所以本文根据城镇居民人均可支配收入、农村居民人均纯收入城镇化率计算得到地区总体的人均收入。计算公式如下:

$$I_{it} = I_{i1t} \frac{P_{i1t}}{P_{it}} + I_{i2t} \frac{P_{i2t}}{P_{it}} = I_{i1t} \frac{P_{i1t}}{P_{it}} + I_{i2t} \left(1 - \frac{P_{i1t}}{P_{it}} \right). \quad (5)$$

在求出泰尔指数后, 由于数据太小, 为了防止回归时系数过大, 把泰尔指数统一乘以 100。

3.2 教育扩展数据

根据统计年鉴上人口受教育的程度的统计资料, 受教育程度分为 5 个层次: 识字很少或不识字 ($y_1 = 0$)、小学 ($y_2 = 6$)、初中 ($y_3 = 9$)、高中 (含中等职业教育) ($y_4 = 12$)、大学 (高职高专、研究生及以上学历) ($y_5 = 16$)。

除 2010 年外原始数据均来源于《中国统计年鉴》中“分地区的 6 岁以上人口的受教育程度人口数”条目。2010 年的数据来源于中国统计局的第六次人口普查。2001 年数据空缺,我们用 2000 年和 2002 年的平均值进行填补。为便于分析,防止系数太小,本文把教育不均数据乘以 100。

3.3 城镇化数据

城镇化率经常用来衡量一个国家或地区的经济发展水平和进步程度。目前主要有两种方法进行衡量。(1) 按照户籍制度,把总人口分为农业人口和非农业人口,用非农业人口占总人口的比重度量;(2) 按照常住人口划分标准,把总人口分成城镇人口和乡村人口,以城镇人口占总人口比重作为城镇化率。基于泰尔指数公式定义及数据的可取性,本文选用第二种方法度量城镇化率。城镇化率的所有原始数据来源于《中国统计年鉴》、各省(直辖市、自治区)的《统计年鉴》以及《新中国六十年统计资料汇编》。其中 2006-2014 年省份城镇化率直接来源于 2007-2015 年《中国统计年鉴》的“分地区年末城镇人口比重”,2005 年的省份城镇化率来源于 2006 年《中国统计年鉴》的“各地区人口城乡构成”。对于 1997-2004 年的数据,部分省份(直辖市、自治区)的统计年鉴数据比较齐全,可以直接得到按常住人口划分的城镇人口和乡村人口,然后运用公式即可求得城镇化率,这部分地区包括北京、内蒙古、吉林、黑龙江、安徽、江西、湖北、湖南、贵州、西藏、甘肃、青海。

另一部分省份(直辖市、自治区)的 1997-2004 年的统计年鉴数据比较齐全,但是并不是按照常住人口划分的,而是按照户籍人口划分的,且这些省份在《新中国六十年统计资料汇编》中不能直接找到城镇化率数据。对于这些地区,先按照非农业人口占总人口的比重计算出初步的城镇化率(如果直接有数据就直接摘抄),然后以 2005 年为基点(河北以 2003 年,因为可以在新中国统计资料汇编中可以找到追溯到 2003 年的城镇化率数据;上海以 2006 年为基点,因为 2005 年数据明显高于 2006 年,与实际不符;浙江以 2000 年为基点,因为在新中国统计资料汇编中可以找到追溯到 2000 年的城镇化率数据;福建以 2000 年为基点,并且由于 2005 年差距太大,明显不符合趋势,用福建省《统计年鉴》上的数据替换掉在《中国统计年鉴上》查找的数据;广西壮族自治区的数据分别以 1996 年和 2000 年为基点推算出 1997-1999 年的数据,然后再求平均值;云南、陕西和新疆虽然城镇人口数据一直存在,但是由于统计口径不一致,2000 年前后数据存在较大跳跃性,不可取,因此对 2000 年之前的数据以 2000 年为基点对非农业人口占常住人口的比例进行调整;宁夏以 2000 年为基点)对计算出的城镇化率数据进行比率调整,这里城镇化率的调整方法类比上文中的可支配收入调整方法。这部分地区包括天津、河北、山西、上海、浙江、福建、辽宁、山东、广东、广西、海南、四川、云南、陕西、宁夏、新疆。

重庆市的 1997-2004 年的统计年鉴数据不齐全,不能直接从相应的统计年鉴获得数据,作为替代,采用《新中国六十年统计资料汇编》的数据,且直接获得城镇化率的地区数据,不需要调整。

说明:①在进行 1997-2004 年的数据采集的过程中,如果相应省份(直辖市、和自治区)的《统计年鉴》和《新中国六十年统计资料汇编》上均有相应的数据,且根据比对,可以完全一一对应,则采纳其中更精确的数据。如黑龙江、江苏、河南虽然上面两个资料上都能找到相应数据,但是由于后者精确度更高,所以依然选择后者。②根据核对,贵州省 2005-2016 年的数据均是非农业人口占常住人口的比例,但是由于缺乏核准基点,贵州省的所有数据均由非农业人口占常住人口的比例计算得出。③由于城镇化率是不可逆的,通过用各省后一年数据减去前一年的数据求出连续两年之间的城镇化率变化值,如果出现负值,说明数据存在异常。通过检测发

现北京、黑龙江、上海、山东、湖北、青海和新疆均有年份出现异常点，但是其中北京、上海、青海和新疆的负值绝对值较小，数据误差仅在 0.1% 以内，予以忽略不计。对于负值绝对值超过 1% 的黑龙江 (1999 年到 2000 年，城镇化率变化值为 -2.25%)、山东 (1997 年到 1998 年，城镇化率变化值为 -0.45%) 和湖北 (1996 年到 1997 年，城镇化率变化值为 -2.5%)，此处按照采信后面数据的原则，对前面的数据进行如下调整：首先求出除此异常变化值之外的其他所有变化值的平均值，然后用此平均值代替异常的变化值，其他差值保持不变，以此逆向推出前面所有的城镇化率。以黑龙江省为例，求出除 -2.25% 之外的其他所有差值的平均值为 0.35%，则把 1999 年的数据调整为 2000 年的城镇化率 51.94% 减去平均差值 0.35%，即变成 51.59%，然后 1998 年的数据仍然是 1999 年的数据减去此两年间的原差值 0.20%，以此类推。^④最后为了便于分析，把城镇化率数据乘以 100。

3.4 教育投入

教育属于公共品，政府对教育的财政支出可以反映出政府对教育的支持力度和重视程度。本文用各省每年的总教育经费占各省 GDP 的比重乘以 100 后表示教育投入。理论上，教育投入会对教育的扩展水平和教育公平程度产生直接影响。教育投入可能会增加教育扩展水平，所以我们在式 (3) 中加入 $EDIN$ 进行验证。然而，教育投入对教育均衡程度的影响是事先难以判断的，因为教育投入可能由于具有城市偏向性而导致城乡教育不均程度拉大，所以这里我们在式 (2) 中加入教育投入来观察教育投入对教育不均程度的影响。原始数据来源于国家统计局与 1997-2016 年的《中国教育经费统计年鉴》。对于 2012 教育经费数据空缺，我们取 2011 年和 2013 年的平均值进行填补。对于 2014 年的数据空缺，我们把 2013 年看作是 2012 年和 2014 年的平均值进行预测推算，即用 2013 的教育经费的 2 倍减去 2012 年的教育经费计算得出。

3.5 其他变量

经济增长数据来源于各省 2017 年《统计年鉴上》，年鉴已经对数据进行了最新的统计调查口径调整。人均收入原始数据来源于国家统计局官网，本文将单位转换成万元。

4 实证分析

实证分析采用 Stata12.0，变量的基本描述性统计分析如表 1 所示。

表 1 变量的描述性统计分析

变量 (单位)	观测数	均值	标准差	最小值	最大值
$INEQ_{it}$ (%)	620	12.720	6.311	17.571	35.214
EDU_{it} (年)	620	8.091	1.331	2.920	12.300
$EDINEQ_{it}$ (%)	620	24.635	6.3569	17.000	59.900
LNY_{it} (万元)	620	0.563	0.897	-1.491	2.470
$URBAN_{it}$ (%)	620	46.492	16.293	14.000	90.000
$TRADE_{it}$ (%)	620	29.667	37.541	3.205	172.284
GR_{it} (%)	620	111.304	2.613	97.500	123.800
$AGRI_{it}$ (%)	620	8.855	3.729	0.910	18.970
$EDIN_{it}$ (%)	620	24.382	6.157	17.000	59.500

4.1 平稳性检验

在进行回归分析之前,先对面板数据的平稳性进行 Fisher-ADF 检验,以此确保估计结果的有效性,检验结果如表 2 所示。除了变量 *URBAN*,其他变量的 P 值都远远小于 0.01,且变量 *URBAN* 的 P 值也小于 0.05,在 5% 的显著性水平下均拒绝原假设,表明各变量均平稳,可进行下一步回归。

表 2 面板数据平稳性的 Fisher-ADF 检验

变量	统计量	P 值	变量	统计量	P 值	变量	统计量	P 值
<i>INEQ_{it}</i>	172.349	0.000	<i>LNY_{it}</i>	202.474	0.000	<i>GR_{it}</i>	170.387	0.000
<i>EDINEQ_{it}</i>	146.695	0.000	<i>URBAN_{it}</i>	81.541	0.049	<i>AGRI_{it}</i>	105.058	0.000
<i>EDU_{it}</i>	104.010	0.000	<i>TRADE_{it}</i>	164.873	0.000	<i>EDIN_{it}</i>	128.354	0.000

注: 0.000 代表数值小于 0.001。

4.2 F 检验与 Hausman 检验

面板数据模型有许多计量方法,常见的有混合模型、随机效应模型和固定效应模型。为选择合适的模型,本文先进行 F 检验,再进行 Hausman 检验,检验结果如下表 3 所示。F 检验和 Hausman 检验的统计量数值均较大,对应的 P 值均远远小于 0.01,说明在 1% 的显著性水平下拒绝原假设,适合选择固定效应模型进行实证分析。

表 3 模型设定的 F 检验与 Hausman 检验

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
F 值	51.060 (0.000)	59.670 (0.000)	42.880 (0.000)	45.190 (0.000)	48.640 (0.000)	48.840 (0.000)	49.890 (0.000)	49.780 (0.000)
H 值	54.680 (0.000)	54.680 (0.000)	55.240 (0.000)	126.390 (0.000)	139.580 (0.000)	130.720 (0.000)	132.720 (0.000)	122.040 (0.000)

注: 括号内为对应检验的 P 值; 0.000 代表数值小于 0.001; (1)-(8) 代表含有不同自变量的逐步回归模型。

4.3 回归分析

严格说,本文选取的控制变量并不是严格意义的控制变量,这些变量是否会影响因变量及其影响方向有一些文献证据,但是并没有统一公式,为验证参数的稳健性,使回归结果更有效,本文采取了逐步回归的估计方法。即先用只含主要解释变量 (*EDINEQ*、*EDU*) 的方程,再加入变量 *EDU*²,然后将控制变量一个一个加入方程中,分别得到模型 (1)-(8)。

表 4 结果显示,除了模型 (1) 的变量不显著外,其他 7 个方程的变量均在 5% 显著性水平下拒绝为零的原假设,且大多变量都在 1% 水平下显著。由模型 (2)-(8) 可以看出,教育不均、教育扩展与收入差距之间的关系是比较稳定的,在逐步加入控制变量后,各变量回归系数的符号与显著性均未发生变化,同时 8 个模型的拟合优度也达到了要求。经分析可知,实证的结果是稳健有效的。

在模型 (1) 中,只含教育不均 (*EDINEQ*) 和教育扩展 (*EDU*) 两个解释变量,回归结果均未达到显著性水平的要求。在模型 (2) 中加入教育扩展的平方项 (*EDU*²) 之后,回归拟合程度提高,系数符号发生了改变,且系数的统计性质显著,并在之后的模型中一直保持一。因此,在接下来的分析中,模型 (1) 的结果可以不参考。

表 4 模型 (1)–(8) 的 31 个省份逐步回归的估计结果

自变量	因变量: $INEQ_{it}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$EDINEQ_{it}$	0.046 (0.691)	-0.213*** (-3.052)	-0.232*** (-3.444)	-0.261*** (-3.939)	-0.224*** (-3.460)	-0.216*** (-3.344)	-0.209*** (-3.272)	-0.204*** (-3.170)
EDU_{it}	0.003 (0.576)	-0.077*** (-7.458)	-0.118*** (-10.099)	-0.121*** (-10.575)	-0.117*** (-10.491)	-0.118*** (-10.595)	-0.112*** (-10.009)	-0.113*** (-10.073)
EDU_{it}^2		0.005*** (8.738)	0.007*** (11.184)	0.007*** (11.446)	0.007*** (11.519)	0.007*** (11.647)	0.007*** (10.858)	0.007*** (10.915)
LNY_{it}			0.024*** (3.333)	0.016** (2.216)	0.016** (2.266)	0.022*** (2.942)	0.022*** (2.894)	0.021*** (2.870)
LNY_{it}^2			-0.009*** (-5.556)	-0.009*** (-5.451)	-0.009*** (-5.400)	-0.010** (-5.829)	-0.010*** (-6.163)	-0.010*** (-6.271)
$URBAN_{it}$				0.130*** (4.916)	0.150*** (5.759)	0.146*** (5.650)	0.138*** (5.342)	0.142*** (5.462)
$TRADE_{it}$					-0.042*** (-5.743)	-0.043*** (-5.894)	-0.043*** (-5.908)	-0.044*** (-6.027)
GR_{it}						-0.001** (-2.442)	-0.002*** (-2.994)	-0.002*** (-3.097)
$AGRI_{it}$							-0.137*** (-3.168)	-0.136*** (-3.144)
$EDIN_{it}$								-0.062 (-1.184)
常数项	0.069 (1.422)	0.473*** (7.267)	0.639*** (9.442)	0.618*** (9.304)	0.591*** (9.113)	0.739*** (8.350)	0.763*** (8.659)	0.787*** (8.702)
R^2	0.446	0.511	0.548	0.566	0.590	0.595	0.602	0.603

注：系数下方括号内的值为 t 值，*，**，*** 号分别表示 1%，5% 和 10% 的显著性水平下的假设检验。

根据模型 (2)–(8) 教育扩展及其二次方的回归系数符号可知，本文得出与前人不一致的结果，教育扩展与收入差距之间的关系不符合“倒 U”曲线关系，反而变成了“正 U”曲线关系，这也说明教育扩展与收入差距之间并不是简单的线性关系。经计算，教育扩展曲线的拐点在 7–8 年间，而根据我国教育数据，2016 年全国平均受教育年限为 9.10 年，这说明我国目前处于随着教育扩展水平的提高，收入差距正不断扩大的阶段。这一结果表明虽然未证实前人的“倒 U”曲线关系，但目前所处阶段与杜鹏 (2005)^[23] 的结论是一致的。

教育拓展与收入分配不均的“倒 U 形”关系认为，起初的教育水平的整体提升存在结构性差异，一些地域或者一些团体先行获得教育的改善，从而在劳动市场获得竞争优势，享受更高的报酬，从而早期社会平均受教育年限的增加将使得收入差距拉大，教育提升的后期是社会整体教育水平的提高，不同地域和不同团体间教育差距减少，收入差距开始降低。但是，这与我国的教育提升现状并不相符，改革开放后，我们推行了以基本普及九年义务制教育和基本扫除青壮年文盲的“两基”教育政策，这种教育政策是在全国普适性强制推广的，不存在“倒 U 形”假说中城市获得顺序的先后。相反，普适性的政策下，社会的整体教育水平有了很大的提高，作为人力资本的核心要素，教育整体水平的提高，和教育的正外部性影响，将缓解劳动者之间的出生地、年龄、性别等方面差异，减少劳动收入的分化，许多国家的经验都验证了整体教育水平提升对收入差距的缓解效应，如 Becker 和 Chiswick (1966)^[1] 的研究发现美国地区间

的收入差距与平均受教育年限负相关。那么整体教育水平提高缓解收入差距是教育“倒 U 形”假说中的拐点右侧部分还是“正 U 形”的左侧部分，还是两者关系是单调减函数形式？本文的证据在平均受教育年限 7-8 年发现了平均受教育年限对收入分配不均的拐点，因此不支持两者的单调函数形式，对于是否是“倒 U 形”的右侧部分，一方面我国的教育政策不支持“倒 U 形”的左侧一说，另一方面新的拐点出现，将使得“倒 U 形”假说演变成多个拐点的更高幂次函数假说，这些文章均没有做出解释。

关于教育拓展和收入分配差距“正 U 形”的解释，左侧部分可以认为是中国在整体教育水平低下的情况下推行的普适性强制“两基”教育提升了整个社会的教育水平，这种普适和强制在短期内缓解了教育人力资本的差异，缩小了收入差距，但是由于社会整体教育水平提高后，在大力发展经济的背景下，教育质量开始分化，平均受教育年限已经无法衡量人力资本的差异了，这可以从几个角度进行解释：(1) 义务教育质量的分化。随着家庭收入的提高和对教育的重视，近年来学区房兴起和火爆，说明义务教育质量的分化明显。(2) 高等教育质量分化明显。改革开放初 1978 年，中国的高等教育毛入学率只有 1.55%，1988 年达到 3.7%，1998 年升至 9.76%，1999 年开始大学扩招，高等教育毛入学率快速上升，2017 年教育部发布的《2016 年全国教育事业统计公报》显示，2016 年，我国高等教育毛入学率已达 42.7%。高等教育从精英教育向大众化教育转变，显然平均受教育年限无法衡量背后隐含的教育质量和人力资本的差异。基于以上两点原因，有理由认为平均受教育年限无法准确衡量教育质量的差异和人力资本的差异，因此随着社会平均受教育年限的提升，收入分配的差距将扩大。因此教育扩展和收入差距之间的“正 U 形”更符合中国教育发展的实情。

同时模型 (2)-(8) 中也显示教育不均与收入差距之间是负相关关系，说明目前我国教育不均的改善，并没有起到减缓收入差距的作用。这与杨俊等 (2008)^[17] 得出的结论相同，而与白雪梅 (2004)^[14] 的结论相反。一般认为，教育不均的缩小会改善收入分配不平衡，但是本文结论与之相反，其原因有两个：(1) 大学教育从精英教育向大众化教育转变，是否上过大学在“人口红利时期”对收入的影响不明显。尽管 2010 年开始，我国的整体抚养比触底后开始提升，但是年轻人仍是社会就业的主力军，在职业选择上，大多数受过大学教育的年轻人倾向于选择“体面”的工作环境，而技术工人和建筑工人等“不体面”工作需求量大却受到冷落，导致社会出现了大学生待遇和农民工倒挂的状况。因此，在教育大众化转变和经济高速发展对技术劳动力需求增大的背景下，是否受过大学教育和教育年限的差异并不会影响到收入分配的差距，甚至目前这个环境呈现了教育扩展和收入差距之间的负相关关系，但是长期看，随着产业转型升级，对劳动者技术要求的提高，是否受过良好教育将显著影响收入水平，届时可能会出现教育差距和收入差距的正相关关系。(2) 目前人口普查中教育的分类过粗，仅将人口教育程度分成“小学、初中、高中和大专及以上”无法准确衡量整个社会的受教育水平情况和受教育水平差距，大学教育从精英教育向大众化教育再向普及化教育转变，中国的大学毛入学率将超过百分之五十，大专、本科、硕士和博士的比重与改革开放初发生了重大变化，早期设立的分类规则已经无法反映教育水平差异现状，以至于大多数以教育年限测算人力资源来探讨社会经济的研究容易存在偏差。

模型 (3) 中加入了控制变量人均 GDP 的自然对数 ($LN Y$) 及其平方项 ($LN Y^2$)，一次项系数为正，二次项系数为负，验证了经济水平与收入差距之间的库兹涅茨“倒 U”曲线关系。即随着经济发展水平的提高，我国收入差距先会扩大到一定程度，随后再逐渐减少，这与主流研究结论是一致的。经计算，当 $LN Y$ 约为 1.0337 时，收入差距会达到最大值。而根据本文数据，2016 年全国 $LN Y$ 为 1.6529，这说明我国目前已经处于倒 U 曲线的右半段，即目前经济

发展水平的提高有利于缩小城乡收入差距。可能的原因是随着我国整体经济水平的发展，人们也越来越重视收入分配平等的问题。

模型 (4) 和模型 (5) 分别进一步加入了控制变量城市化水平 (*URBAN*) 与经济开放程度 (*TRADE*)。根据回归结果发现城市化水平的提高会加剧收入差距，可能的原因是城市化水平越高，政府实施的政策更容易偏向于城镇，导致城乡发展差异越大，而使收入差距越大。而经济开放程度的提高则会改善收入分配不平衡的问题。

模型 (6)、(7) 和 (8) 分别进一步加入了控制变量经济增长率 (*GR*)、政府对农业的支持 (*AGRI*) 与政府教育投入水平 (*EDIN*)。结果发现这三项控制变量都与收入差距是负相关关系，其中经济增长率和政府对农业的支持两项在 1% 水平下显著。这说明经济增长率越高，越有利于缩小收入差距，这也印证了 *LNY* 对收入差距的影响分析，即经济发展水平越高，越有利于收入差距问题的缓解。回归结果表明政府对农业的支持也会使收入差距趋于缩小，说明目前我国政策偏向因素确实会影响收入差距。虽然政府教育投入水平在本文中并没有得到显著影响的结果，但是系数也与预期是相同的，可能是由于目前教育投入力度还不够、教育投入分布不够合理导致。

4.4 稳健性检验

目前一些研究表明，教育不均会影响收入差距，同时收入差距也会影响教育不均。如杨俊等 (2008)^[17] 发现教育不均对收入差距有负相关影响的同时，收入差距的缩小也可以促进教育不均的缩小。这都说明教育不均可能是内生解释变量，教育不均与收入差距之间可能存在内生性问题。

为断定是否存在内生性问题，需要先对模型进行豪斯曼检验，以检验教育不均 (*EDINEQ*) 是否为内生解释变量。在进行豪斯曼内生性检验之前，必须先把可能的内生解释变量的工具变量寻找出来。有效的工具变量需要满足与解释变量的相关性和对被解释变量的外生性，在面板数据中，常常使用内生解释变量的滞后项作为工具变量。由于内生解释变量与其滞后项相关，且滞后项已经发生，为前定变量，理论上是符合相关性与外生性要求的，因此本文选择教育不均的滞后项作为工具变量。在恰好识别的情况下，无法检验工具变量是否与扰动项相关，于是选择教育不均 (*EDINEQ*) 的滞后一阶和滞后二阶作为工具变量，保证有效工具变量的个数大于内生解释变量个数，即此时为过度识别情况。寻找好工具变量后，对内生性进行豪斯曼检验，检验结果如下表 5。

表 5 解释变量 *EDINEQ* 的内生性 Hausman 检验

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
H 值	127.140 (0.000)	118.340 (0.000)	120.320 (0.000)	110.980 (0.000)	108.640 (0.000)	109.680 (0.000)	118.840 (0.000)	106.460 (0.000)

注：H 代表 Hausman 统计量，括号内为对应的 P 值；0.0000 代表数值小于 0.0001。

由表 5 结果可知，所有的统计量较大，对应的 P 值均远远小于 0.01，由此可以拒绝“所有解释变量均为外生变量”的原假设，即认为教育不均均为内生解释变量。

考虑到教育不均与收入差距之间的内生性问题，本文采用 IV-GMM 估计重新对面板模型进行逐步回归，以验证实证结果的稳健性，回归结果见表 6。

比较固定效应估计与 IV-GMM 估计的回归结果，可知各回归系数的符号中，人均 GDP 的自然对数 (*LNY*) 由正变为负，经济增长率 (*GR*) 的系数由负变为正，除此之外其他系数符

号均未发生改变。而尽管人均 GDP 的自然对数 ($LN Y$) 系数符号发生改变, 之前得出的经济水平与收入差距之间的库兹涅茨“倒 U”曲线关系仍未发生变化, 目前已经处于倒 U 曲线的右半段的结论也未变。控制变量经济增长率 (GR) 也不是本文的主要研究对象, 且变化也不大, 因此对整体回归结果的影响也不大。从显著性来看, 只有教育不均 ($EDINEQ$) 的显著性水平下降, 政府教育投入水平 ($EDIN$) 从不显著变为显著, 其他变量显著性水平基本保持原有水平。总体来说, 实证结果的稳健性基本得到证实。

表 6 模型 (1)–(8) 的 31 个省份逐步回归的 IV-GMM 估计结果

自变量	因变量: $INEQ_{it}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$EDINEQ_{it}$	-0.207 (-1.286)	-0.450** (-2.389)	-0.250 (-1.491)	-0.298* (-1.758)	-0.293* (-1.729)	-0.268 (-1.641)	-0.299* (-1.827)	-0.269 (-1.609)
EDU_{it}	-0.021*** (-4.730)	-0.095*** (-5.126)	-0.140*** (-8.284)	-0.144*** (-8.508)	-0.144*** (-8.514)	-0.132*** (-7.951)	-0.128*** (-7.800)	-0.127*** (-7.676)
EDU_{it}^2		0.004*** (4.610)	0.009*** (10.476)	0.009*** (10.631)	0.009*** (10.653)	0.008*** (9.883)	0.008*** (9.597)	0.008*** (9.458)
$LN Y_{it}$			-0.000 (-0.096)	-0.010** (-2.419)	-0.010** (-2.397)	-0.018*** (-4.213)	-0.013*** (-2.945)	-0.015*** (-3.319)
$LN Y_{it}^2$			-0.019*** (-10.394)	-0.018*** (-10.226)	-0.019*** (-10.079)	-0.012*** (-6.227)	-0.012*** (-6.393)	-0.013*** (-6.567)
$URBAN_{it}$				0.126*** (3.640)	0.129*** (3.715)	0.150*** (4.546)	0.143*** (4.362)	0.147*** (4.533)
$TRADE_{it}$					-0.007 (-0.807)	-0.017** (-2.161)	-0.018** (-2.267)	-0.019** (-2.431)
GR_{it}						0.003*** (6.610)	0.003*** (5.973)	0.002*** (5.637)
$AGRI_{it}$							-0.124*** (-2.875)	-0.137*** (-3.179)
$EDIN_{it}$								-0.086* (-1.729)
R^2	0.162	0.209	0.366	0.383	0.383	0.436	0.443	0.447

注: 系数下方括号内的值为 t 值, *, **, *** 号分别表示 1%, 5% 和 10% 的显著性水平下的假设检验。

为保证稳健性检验结果的有效性, 下面对工具变量的有效性进行检验。先对工具变量的第一个条件相关性进行检验。对工具变量进行弱工具变量检验, 检验结果如表 7。

表 7 工具变量的弱相关检验

回归	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
F 值	71.238	54.741	57.734	55.736	55.412	55.794	54.676	53.224

注: F 代表 Cragg-Donald Wald F 统计量; 10%, 15%, 20%, 25% maximal IV size 分别为 19.93, 11.59, 8.75, 7.25。

根据 Stata12.0 报告的 F 统计量的临界值标准 (19.93), 可知模型 (1)–(8) 的 F 统计量均远大于 19.93, 因此可以拒绝“存在弱工具变量”的原假设, 即工具变量满足相关性。

接下来进一步考察工具变量有效性的第二个条件, 即外生性是否成立。对模型进行过度识别检验, 检验结果如表 8。

表 8 工具变量的过度识别检验

回归	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
S 值	1.855 (0.173)	1.113 (0.292)	0.852 (0.356)	1.340 (0.247)	1.391 (0.238)	0.957 (0.328)	1.133 (0.287)	0.477 (0.490)

注：S 代表 Sargan 统计量，括号内为对应的 P 值。

由表 8 结果可知，Sargan 统计量的值均较小，且 P 值均大于 0.1，因此我们不能拒绝“所有工具变量均外生”的原假设，即工具变量满足外生性。

综上所述，我们可知工具变量满足有效性，因此稳健性检验的结果有效性也得到了保证，实证结果的稳定性也得到了证实。

5 结论

本文利用 1997-2016 年 31 个省份的省级面板数据，采用固定效应模型，并用 IV-GMM 估计进行稳健性检验，研究了我国教育对收入差距的影响。在考虑了一些控制变量的前提下，模型得出了和前人并不完全一致的结果。根据实证结果与对我国现状的分析，形成的主要结论与建议如下。

研究发现教育扩展与收入差距之间的关系不符合“倒 U”曲线关系，而是“正 U”曲线关系，我国目前处于随着教育扩展水平的提高，收入差距正不断扩大的阶段。改革开放后强制性的普适性以“基本普及九年义务教育、基本扫除青壮年文盲”为宗旨的“两基”政策和其他一些教育政策，较为平均的提升了整个社会的教育水平和文化水平，对缩小收入差距起到了较好作用。但是，近年来，基础教育在不同城市 and 不同学校间分化明显，高等教育转向普及化后，教育质量在学校间也分化明显，教育质量的分化使得在整个社会教育总水平提高的同时，带来了收入分化的社会问题。因此，中国的教育应该防止教育水平差异化，特别要保障基础教育公平，防止基础教育质量分化而造成潜在的社会不公。

现阶段，中国教育不均与收入差距之间是负相关关系，说明目前我国教育不均的改善，并没有起到减缓收入差距的作用。短期内，在教育大众化转变和经济高速发展对技术劳动力需求增大的背景下，技术工人的紧缺和大学毕业生对工作环境的选择，一定程度上出现了“体力劳动者”待遇高过大学生的状况，所以短期内是否受过大学教育和教育年限的差异并不会影响到收入分配的差距，甚至目前呈现了教育扩展和收入差距之间的负相关关系，但是长期看，随着产业转型升级，对劳动者技术要求的提高，是否受过良好教育将显著影响收入水平，届时可能会出现教育差距和收入差距的正相关关系。尽管短期教育不均对收入分配不平衡没有造成影响，但是长期看，仍要特别注意教育不均的社会问题，避免为未来社会分配埋下隐患。

本文还发现我国现行的人口普查中教育的分类过粗，在大学教育从精英教育向大众化教育再向普及化教育转变的背景下，仅将人口教育程度分成“小学、初中、高中和大专及以上”无法准确衡量整个社会的受教育水平情况和受教育水平差距，以至于大多数以教育年限测算人力资源来探讨社会经济的研究容易存在偏差。

[参考文献]

- [1] Becker G S, Chiswick B R. Education and the distribution of earnings [J]. American Economic Review, 1966, 56(1): 358-369.
- [2] Chiswick. Earnings inequality and economic development [J]. The Quarterly Journal of Economics,

- 1971, 85: 21-39.
- [3] Tinbergen J. The impact of education on income distribution [J]. *Review of Income and Wealth*, 1972, 16(2): 221-234.
- [4] Marin A, Psacharopoulos G. Schooling and income distribution [J]. *Review of Economics and Statistics*, 1976, 58(3): 332-338.
- [5] Psacharopoulos G. Unequal access to education and income distribution [J]. *De Economist*, 1977, 125(3): 383-392.
- [6] Winegarden C R. Schooling and income distribution: Evidence from international data [J]. *Economica*, 1979, 46(181): 83-87.
- [7] Tilak J B G. Rates of return to education and income distribution [J]. *De Economist*, 1989, 137: 454-465.
- [8] Bourguignon F, Morrisson C. Income distribution, development and foreign trade: A cross-sectional analysis [J]. *European Economic review*, 1990, 34: 1113-1132.
- [9] Ram R. Education expansion and schooling inequality: International evidence and some implications [J]. *Review of Economics and Statistics*, 1990, 72(2): 266-273.
- [10] Park K H. Educational expansion and educational inequality on income distribution [J]. *Economics of Education Review*, 1996, 15(1): 51-58.
- [11] 德胜. 教育扩展与收入不平等 [J]. *经济研究*, 1997, (10): 46-53.
- [12] Gregorio J D, Lee J W. Education and income distribution: New evidence from cross-country data [J]. *Review of Income and Wealth*, 2002, 48(3): 395-416.
- [13] 杜洁瑞, 崔霞. 数据缺失条件下我国教育收益率估计和分析 [J]. *数理统计与管理*, 2018, 37(2): 198-204.
- [14] 白雪梅. 教育与收入不平等: 中国的经验研究 [J]. *管理世界*, 2004, (6): 53-58.
- [15] 王小鲁, 樊纲. 中国收入差距的走势和影响因素分析 [J]. *经济研究*, 2005, (10): 24-36.
- [16] 陆铭, 陈钊. 因患寡, 而患不均 — 中国的收入差距、投资、教育和增长的相互影响 [J]. *经济研究*, 2005, (12): 90-101.
- [17] 杨俊, 黄潇, 李晓羽. 教育不平等与收入分配差距: 中国的实证分析 [J]. *管理世界*, 2008, (1): 38-47.
- [18] 王艳真, 李秀敏. 中国教育扩展、教育不平等与收入分配差距间的相互影响 [J]. *税务与经济*, 2015, (6): 29-34.
- [19] 阮敬, 丁琳, 纪宏. 收入分布视角下的收入分配研究 [J]. *数理统计与管理*, 2018, 37(1): 104-121.
- [20] 陈昌兵. 各地区居民收入基尼系数计算及其非参数计量模型分析 [J]. *数量经济技术经济研究*, 2007, 24(1): 133-142.
- [21] 吴延科, 田茂再. 社会经济地位与教育回报率的无条件分位回归分析 — 基于 CGSS2010 年调查数据 [J]. *数理统计与管理*, 2016, 35(4): 692-699.
- [22] 周华, 闫琴, 滕娇, 张赫. 公共、私人教育投入分配对益贫式增长的作用机制 — 基于分层次教育体系的研究 [J]. *数理统计与管理*, 2018, 37(3): 422-437.
- [23] 杜鹏. 我国教育发展对收入差距影响的实证分析 [J]. *南开经济评论*, 2005, (4): 47-52.