

计划生育对子女教育水平的影响

——来自中国的微观证据

秦雪征 庄 晨 杨汝岱*

摘 要 随着“二孩”政策的逐步放开,新政策对中国经济发展的长期影响引发了新一轮的关注。本文基于 2005 年全国 1% 人口抽样调查数据,利用多种计量方法对中国人口数量与质量之间的因果关系进行识别。结果显示,人口数量的减少显著提高了子女的教育水平,但该影响只在收入较低、信贷发展较落后的地区得以显现。随着中国经济的发展,全面放宽家庭对子女人力资本投资的约束将可能减轻计划生育全面放宽的负面影响,从而有利于经济的可持续发展。

关键词 计划生育,教育水平,断点回归设计

DOI: 10.13821/j.cnki.ceq.2018.04.02

一、引 言

面对人口老龄化的压力,党中央在十八届三中全会通过的《关于全面深化改革若干重大问题的决定》中提出启动“单独二孩”的政策,标志着中国自 20 世纪 70 年代以来的人口政策首次在全国得到放松。随后,十八届五中全会提出了全面放开“二孩”政策。新政策实施以来,人口管控的放松对经济发展的影响备受关注。国内部分学者认为,如果放松当前的人口控制政策,中国经济增长仍将承受较大的“人口负担”,并面临着发展可持续问题的挑战(刘永平和陆铭,2008;黄少安和孙涛,2013)。当然,国内也有学者认为,继续执行严格的计划生育政策将加重人口老龄化负担,并对中国的经济增长带来负面影响(周俊山和尹银,2011;马骁骁,2010)。从理论层面来看,以上两种观点争论的焦点是,在中国当前的经济发展水平下,人口数量与质量之间是否存在权衡取舍的替代关系:如果人口数量的大幅度增长将导致人口质量的下降,那么全面放松人口政策将可能对中国“以人为本”的创新型经

* 秦雪征、杨汝岱,北京大学经济学院;庄晨,华盛顿大学经济学系。通信作者及地址:庄晨,美国华盛顿州西雅图华盛顿大学经济学系 305 办公室,98195;电话:1-1206-48706429;E-mail:zogcee@gmail.com, cczhuang@uw.edu。本文得到国家自然科学基金常规面上项目(71573003)、北京市社会科学基金研究基地项目(16JDLJB001)以及北京大学经济学院科研种子基金的资助。

济增长模式带来不利影响;反之,如果人口数量与质量之间不存在这样的负相关关系,甚至人口数量的提升还会导致人口质量的改善,那么当前人口政策的放松将很可能会促进中国经济的长期增长。

人口数量与质量的替代(quantity-quality trade-off)是经济学和人口学领域长期关注的重要理论命题。这一命题最早由“人力资本学说”的创始人 Gary Becker 在 Becker and Lewis(1973)中正式提出[Willis(1973)在同一时期也对这一命题有所阐述]。根据经典理论,在父母对子女一视同仁、家庭面临预算约束(budget constraint)和信贷市场失灵(credit market failure)这三条假定下,由于家庭资源的稀缺性,子女数量的增多必然意味着父母只能对每个子女进行更少的人力资本投资,从而家庭中子女的数量与质量(人力资本)之间便会存在负向关系。为检验这一理论,文献中往往使用子女的教育水平或健康水平作为人力资本的代表,用实证研究方法来考察家庭中子女的平均数量与平均质量的关系。大部分文献发现,两者之间确实存在负向关系(Glick *et al.*, 2005; Li *et al.*, 2008; Rosenzweig and Zhang, 2009; Hatton and Martin, 2010; Liu, 2014),但也有一些文献发现两者之间没有显著关系(Black *et al.*, 2005; Park and Chung, 2012)或它们存在正向关系(Qian, 2009; Lordan and Frijters, 2013)。

经验证据与理论预测的背离往往是因为现实条件不符合理论假设。例如,Black *et al.* (2005)基于挪威全国人口数据发现,在考虑子女的出生顺序之后,家庭中子女的数量对其教育水平的影响不再显著,并且子女的出生次序与其教育水平存在负相关关系;这表明父母对其子女可能不是“一视同仁”的,父母有可能对部分子女投入了更多的时间与资源,或者说某些子女在家庭中受到了特殊的偏爱(例如父母可能对长子寄予更高期望或对女孩存在性别歧视)。Park and Chung(2012)通过对孟加拉国 MHSS 数据的实证研究发现,家庭规模对教育水平的负向影响只在前两个子女中显著,而对于第三个及以后的子女不再显著;其原因可能是,当家庭达到一定规模时,预算约束或信贷约束可能不再成为家庭的限制因素,多生育的决策恰恰是父母收入或能力提高的结果,从而那些来自大规模家庭的子女在人力资本积累方面未必低于小规模家庭的子女(事实上,这也说明家庭内部生育决策存在内生性)。Qian (2009)利用 CHNS 数据对中国农村地区进行的实证研究同样发现了这一现象;其研究表明,家庭中子女数量增加将显著提高长子或长女的入学率,并且在所有子女的性别相同时这一效应更大。Lordan and Frijters(2013)则利用秘鲁数据发现,只有在家庭意外生育的情况下,子女的数量与质量(身高)之间才存在负向关系,而在家庭主动生育的情况下,子女数量与质量的关系变为正向。

家庭内部生育决策的内生性是相关实证研究所面临的重要挑战。一方面,子女的数量与质量往往同时受到不可观测因素的影响,例如父母的社会地位

或对子女特殊的偏好有可能同时影响子女的数量和质量；另一方面，子女的数量与质量之间还可能存在反向因果关系，即子女质量的提高可能导致父母偏好生育更多的子女。为了控制内生性的影响，文献中往往采用工具变量(IV)方法，而常用IV包括“是否生育双胞胎”(Glick *et al.*, 2005; Li *et al.*, 2008; Angrist *et al.*, 2010; Hatton and Martin, 2010)或“是否实施人口政策”(Qian, 2009; Park and Chung, 2012; Cameron *et al.*, 2013)，因为这些变量一般外生于家庭生育决策，并且会直接影响家庭规模。然而，由于人口政策只在部分国家得到实施，并且实施力度、时间和范围不同，因此并非所有政策都适合用于IV方法，这使得基于人口政策的研究受到限制。用“是否生育双胞胎”作为家庭规模的IV也可能存在问题：一方面，尽管双胞胎的出生会“意外”增大家庭规模，然而基因等不可观测的因素也可能同时对人口质量和双胞胎生育带来影响，从而削弱工具变量的有效性；另一方面，父母后续生育决策和对子女的态度也可能受到双胞胎意外出生的影响而发生改变，比如在中国双胞胎家庭无法享受独生子女家庭奖励或优待，并且同时增加多个子女往往会在短期和长期给家庭带来更大的经济压力，这可能使父母放弃后续的生育计划或对双胞胎子女采取不同的态度。

断点回归设计(regression discontinuity design, RDD)作为近年来兴起的计量经济学方法还很少被应用于该问题的研究中。RDD的基本思想是，在制度的外生变化环境中寻找合适的驱动变量，使其能够导致人口数量发生“跳跃”，但不直接影响人口质量，由此便可估算出人口质量的变化在多大程度上是由人口数量的外生变化所引起的，即验证了两者的因果关系。相对于传统的IV方法而言，RDD的优势是能够充分利用政策冲击所带来的“跳跃”效应，对变量本身的外生性要求较为温和，从而有利于解决家庭生育决策的内生性问题，进而得到对两者因果关系的无偏估计。

本文是RDD方法在相关实证研究中的一次尝试。由于中国是一个社会主义人口大国，历史因素和制度特点导致中国的人口控制政策在实施力度上显著大于其他国家，这就为RDD的实现提供了可能性。计划生育政策在全国范围内的实施为研究人口数量与质量的关系提供了很好的自然实验(natural experiment)机会，而2005年全国1%人口抽样调查数据则满足了RDD对大样本量的需求。基于此，本文利用中国20世纪80年代开始推行的“一胎化”强制计划生育政策，对“Q-Q替代”这一理论命题进行RDD实证检验。为了验证RDD实证结果的可信性，我们同时提供一个双重差分(DID)的回归分析。

本文的贡献主要集中于以下几个方面：(1) 本文尝试使用了断点回归设计(RDD)等多种计量方法解决家庭内部生育决策的内生性问题，为相关领域的研究提供了一个示范性的参考；(2) 本文对Becker and Lewis(1973)和Willis(1973)提出的人口质量与数量替代的相关前提假设进行了检验，为理解该经典理论提供了更为全面的视角；(3) 我们的实证模型基于具有全国代表性的大

样本数据,为重新评估中国计划生育政策的实施效果提供了最新的估计结果;(4)我们得到的经验结果还有助于深刻认识中国人力资本的投资与决定机制,为中国当前和未来人口政策的制定提供微观理论基础和政策启示。

本文其余部分安排如下:第二部分简单介绍中国计划生育政策和“一胎化”政策的背景与制度安排;第三部分介绍本文的研究方法和模型设定;第四部分介绍数据选择及样本描述统计;第五部分汇报实证结果,并对结果进行稳健性检验和安慰剂检验;第六部分通过分样本回归对经典理论的前提条件进行系统的验证;第七部分总结全文并提出政策建议。

二、计划生育与“一胎化”政策

计划生育是中国的一项基本国策,也是一个长期以来不断发展的政策。该人口管控政策自中华人民共和国成立以来大体经历了六个阶段。表 1 是对这六个阶段基本情况的总结。

表 1 中华人民共和国成立以来中国计划生育政策的主要历史沿革

阶段	年份	内容	文件
第一阶段	1949—1961	计划生育概念提出,引发了关注与争论	《避孕及人工流产办法》(1953 年 8 月 11 日)、《中共中央对卫生部党组关于节制生育问题的报告的批示》(1955 年 3 月 1 日)、《关于发展国民经济的第二个五年计划的建议》(1956 年 9 月 16 日)
第二阶段	1962—1978	计划生育政策提出,计划生育机构成立	《关于认真提倡计划生育的指示》(1962 年 12 月 18 日)、《关于第四个五年国民经济计划纲要》(1970 年 5 月 20 日)、《关于 1975 年国民经济计划的报告》(1974 年 12 月 29 日)、《关于全国计划生育工作汇报会的报告》(1978 年 2 月 24 日)
第三阶段	1979—1991	计划生育成为国策,“一胎”数量抽紧	《中华人民共和国宪法》(1978 年 3 月 5 日,1982 年 12 月 4 日)、《关于控制我国人口增长问题致全体共产党员、共青团员的公开信》(1980 年 9 月 25 日)、《中共中央、国务院关于进一步做好计划生育工作的指示》(1982 年 2 月 9 日)
第四阶段	1992—2000	计生体制逐步规范,组织机构更加健全	《中国计划生育工作纲要(1995—2000)》(1995 年 1 月 14 日)、《关于加强人口与计划生育工作稳定低生育水平的决定》(2000 年 3 月 2 日)
第五阶段	2001—2012	低生育水平的稳定,人口问题统筹解决	《中华人民共和国人口和计划生育法》(2001 年 12 月 29 日)、《中共中央国务院关于全面加强人口和计划生育工作统筹解决人口问题的决定》(2006 年 12 月 17 日)
第六阶段	2013 年至今	计划生育政策放松,“二胎”限制放宽	《中共中央关于全面深化改革若干重大问题的决定》(2013 年 11 月 15 日)、《中共中央关于制定国民经济和社会发展第十三个五年规划的建议》(2015 年 10 月 26 日)、《人口与计划生育法修正案(草案)》(2015 年 12 月 21 日)

注:本表中的部分资料转引自杨发祥(2003)和徐剑(2010),阶段划分稍有区别,主要根据“文件”的内容和性质进行划分。

可以看到,中国的人口控制经历了“先松、再紧、又松”的曲折发展过程:第一、二阶段是计划生育的预备和起步阶段,第三、四阶段是计划生育的加强和深化阶段,而第五、六阶段则是计划生育的成熟与放宽阶段。人口控制政策的变化将引起独生子女率在政策转折点附近发生“跳跃”:当政策从“先松”转变为“再紧”时,独生子女率有可能发生向上的“跳跃”;而当政策从“再紧”转变为“又松”时,独生子女率可能发生向下的“跳跃”。变量在政策断点附近发生“跳跃”是使用断点回归设计的基础条件,有效的RDD需要独生子女率在断点处发生“足够”的跳跃;而如果政策实施力度不够大¹,那么RDD方法将很难得到显著的估计结果。由于大多数国家的人口政策实施力度有限,其政策引发的变化往往不足以被识别,这或许也是目前几乎没有文献使用RDD方法进行相关实证研究的原因之一。

从20世纪70年代末开始,计划生育政策在内容和性质上发生重大转变:从之前的“晚、稀、少”等鼓励性政策全面转向“一胎化”的强制性生育管制(Zhang and Spencer, 1992; Li, 1995; 杨发祥, 2003; 徐剑, 2010; Cameron *et al.*, 2013)。该政策在半数以上的省份(郭志刚等, 2003)²通过法规严格规定中国的城镇居民和国家干部、职工除特殊情况外只能生育一胎,而不少夫妇也自发响应政府号召,在生育一个子女后便领取“独生子女光荣证”并不再继续生育。该政策的实施采取了“胡萝卜加大棒”的方式:一方面,“独生子女光荣证”为遵从“一胎化”政策的父母及其子女提供诸多福利(如住房分配和子女入学的优先权、现金和医疗补贴等);另一方面,对不服从政策的“超生”父母,收取相当于地方年人均可支配收入的3—10倍的罚金(Banister, 1987; Cooney *et al.*, 1991; Li and Cooney, 1993)。³

上述“一胎化”政策代表了中国计划生育史上的一次重大制度变革:区别于此前相对缓和并以自愿为主的人口政策,它在实施上具有强制性和较高的贯彻力度。考虑到该政策在短时间内对城镇⁴父母生育决策具有较强的冲击性,我们认为该政策推出后中国城镇人口的独生子女率会发生向上的“跳跃”。由于“独生子女光荣证”在1979年年底开始大范围发行(根据中国卫生和计划生育委员统计,仅1979年全国就颁发了610万张光荣证),而“超生罚款”也在1980年年初迅速在各省推行(梁中堂, 2014)。以此政策执行时点为基础,考虑到“怀胎十月”的孕期,我们预测在1980年10月前后中国新

¹ 这将导致只有少部分人成为顺从者(complier),即实施“一胎化”政策后只有少部分人放弃生育多胎而只生育一胎,那么断点前后的独生子女率就不会发生实质性的变化。

² 实行或基本实行“一胎化”的省份大致包括上海、江苏、北京、天津、四川、重庆、辽宁、黑龙江、广东、吉林、山东、江西、湖北、浙江、湖南、安徽、福建和山西等。

³ 1979年年底至1980年年初,各地均陆续开始出现“超生罚款”的相关规定。根据各地经济水平的不同,“超生罚款”也有所差异;然而,对于大多数父母而言,“超生罚款”力度不轻(Ebenstein, 2010)。

⁴ 城市地区基本做到了“一胎化”,而农村地区在贯彻该政策时受到较大阻力,农民往往以非正常办法(如隐瞒户口)抵制该政策的执行。

出生人口中可能会出现一个独生子女率的“断点”。尽管各地区实施政策的严格程度和执行政策的时点有区别,但我们相信在预期的“断点”处,政策影响的力度和密度都将达到最大值。为了证实此预测,图 1 基于 2005 年全国 1%人口抽样调查数据画出了在 1980 年 10 月前后出生的人口独生子女率的变化情况。从图中可以发现,独生子女率在该政策“断点”处的跳跃是比较明显的。

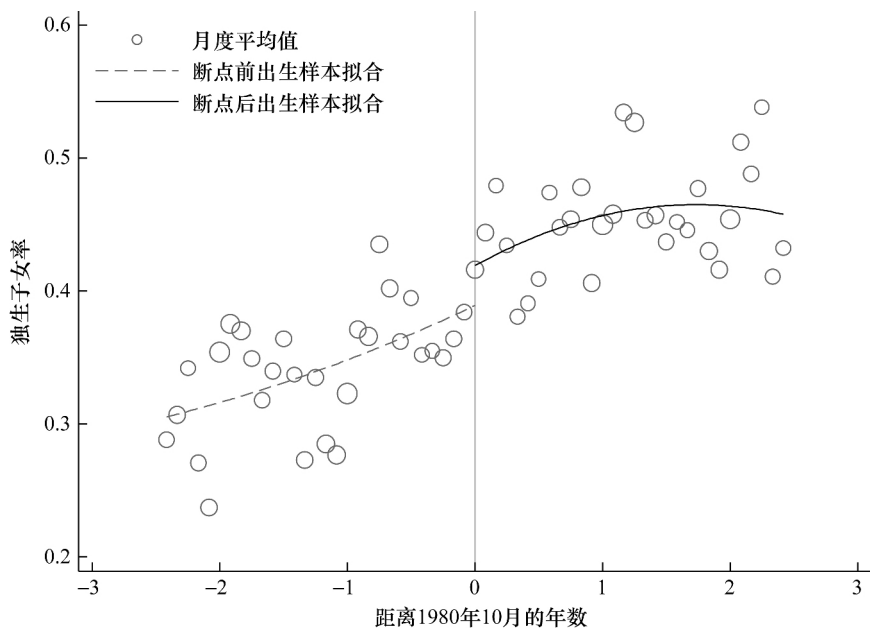


图 1 独生子女率在断点前后的跳跃

注:图中散点表示每个月的独生子女率,通过计算各月出生的样本中独生子女所占比例而得到;图中的两条曲线分别代表对断点前、断点后两个时期的散点进行非线性回归所得到的拟合线。

三、估计方法与实证策略

本文将利用断点回归设计(RDD)和双重差分(DID)等计量方法进行实证检验。下面简要阐述前两种方法在本文中的设定。

RDD 的基本思路是,在外生冲击下有且只有两个变量发生跳跃,从而获得两者之间的因果关系。本文使用的是模糊型的断点回归设计,因为个体在临界值两边接受处理效应(treatment effect)的概率(即成为独生子女的概率)在 0 到 1 之间发生变化,这也是诸多相关文献使用的类型(Angrist and Lavy, 1999; Van der Klaauw, 2002; 雷晓燕等, 2010; 余静文和王春超, 2011; Chen *et al.*, 2013)。

本文的指派变量(assignment variable) X 定义为样本个体的出生月份距离断点(定义为 1980 年 10 月)的时间⁵, 结果变量(outcome variable) Y 是样本个体的人力资本水平(用受教育年限衡量), 处理变量(treatment variable) D 定义为样本个体是否为独生子女(是则取 $D=1$, 否则取 $D=0$)。我们基于式(1)对独生子女与非独生子女之间受教育水平的差异(处理效应)进行估计:

$$\tau = \frac{\lim_{x \downarrow c} E[Y | X] - \lim_{x \uparrow c} E[Y | X]}{\lim_{x \downarrow c} E[D | X] - \lim_{x \uparrow c} E[D | X]}. \quad (1)$$

该 Wald 估计量的分母是处理变量的期望值在断点 c 处发生的“跳跃”, 分子则是结果变量在断点 c 处发生的“跳跃”。我们选取断点 c 附近(距离 $\pm h$)的样本进行估计, 相当于做一个局部试验。可以证明, 在异质性假设下, 式(1)是对顺从者(complier, 指按“一胎化”政策进行生育的个体, 即在“一胎化”政策实施之前打算生育两个或多个子女, 而政策实施后转而只生育一个子女的个体⁶)的局部平均处理效应(local average treatment effect, LATE)的估计。对于式(1)所涉及的四个期望值, 我们可以利用工具变量(IV)的原理, 通过参数估计方法(Imbens and Lemieux, 2008; Lee and Lemieux, 2010)获得。基于 RDD 中的指派变量构造分组指代变量(记为 T , 如果指派变量将个体分配到断点右边的处理组则 $T=1$, 否则 $T=0$), 然后将 T 作为 D 的 IV, 利用传统的两阶段最小二乘法(2SLS)来估计 D 与 Y 之间的因果关系。该参数估计可以用以下回归等式刻画⁷:

$$Y = \theta + \tau D + f(X - c) + \epsilon, \quad (2)$$

$$D = \gamma + \delta T + g(X - c) + u. \quad (3)$$

式(2)和式(3)分别为 2SLS 估计中的二阶段和一阶段回归等式。其中, δ 反映了顺从者的比率; γ 和 θ 是截距项; $f(\cdot)$ 和 $g(\cdot)$ 分别是指派变量对处理变量和结果变量的影响函数, 可以采用线性和非线性设定; 文献中为了缓解 X 与 T 和 D 之间的多重共线性, 一般对 $f(\cdot)$ 和 $g(\cdot)$ 采用二次型设定; u 和 ϵ 则是随机误差项。可以证明, 当“工具” T 为二值变量时, 我们对处理效应 τ 的估计等价于:

⁵ 由于数据的出生时间信息仅仅具体到出生月份, 因此本文中“箱体”(bin)的最小单位是一个月。

⁶ 顺从者不包括原本就打算生育独生子女或政策推行后仍生育多个子女的个体。本文认为, 样本中不存在“一胎化”政策实施前打算生育一个子女而政策实施后却打算生育多个子女的叛逆者(defier), 也即样本是满足单调性(monotonicity)假定的; 并且, 在当时的制度背景下, 顺从者在人群中的比例应该不低, 这就使得估计结果具有一定的普遍意义。

⁷ 可以证明, 在确定型 RDD 中, $\gamma=0$, $\alpha=0$, 而 $\delta=1$ 。

$$\begin{aligned}\tau &= \frac{\sum_{i=1}^N (T_i - \bar{T})(Y_i - \bar{Y})}{\sum_{i=1}^N (T_i - \bar{T})(X_i - \bar{X})} \\ &= \frac{E[Y | c < X \leq c+h] - E[Y | c-h \leq X \leq c]}{E[D | c < X \leq c+h] - E[D | c-h \leq X \leq c]}.\end{aligned}\quad (4)$$

对比式(4)与式(1)可知, 参数估计只为我们提供了一个近似估计量(不一定是最优估计量)。因此, 更符合直觉的方法则是采用非参数方法中的核密度估计法(kernel density estimation), 通过单边核回归(one-sided kernel regression)来实现对它们的点估计(point estimate), 从而得到 Wald 估计量:

$$\lim_{x \downarrow c} E[Y | X] = \frac{\sum_{i: X_i > c} Y_i \times K((X_i - x)/h)}{\sum_{i: X_i > c} K((X_i - x)/h)}, \quad (5)$$

$$\lim_{x \uparrow c} E[Y | X] = \frac{\sum_{i: X_i < c} Y_i \times K((X_i - x)/h)}{\sum_{i: X_i < c} K((X_i - x)/h)}, \quad (6)$$

$$\lim_{x \downarrow c} E[D | X] = \frac{\sum_{i: X_i > c} D_i \times K((X_i - x)/h)}{\sum_{i: X_i > c} K((X_i - x)/h)}, \quad (7)$$

$$\lim_{x \uparrow c} E[D | X] = \frac{\sum_{i: X_i < c} D_i \times K((X_i - x)/h)}{\sum_{i: X_i < c} K((X_i - x)/h)}. \quad (8)$$

式(5)—式(8)分别从断点两侧对结果变量和处理变量的期望值进行点估计。带宽 h 是依照 Imbens and Kalyanaraman(2012)按最小化均方误差(mean squared error)选择的最优带宽(optimal bandwidth)。 $K(\cdot)$ 为内核函数(kernel function), 文献一般认为三角内核(triangle kernel)是对局部线性回归进行边界估计的最优选择(Fan and Gijbels, 1996; Lee and Lemieux, 2010), 因而本文将使用三角内核进行基准的非参数估计, 同时在稳健性检验中选用矩形内核(rectangular kernel)作为参考。

在 RDD 的框架下, 只要其他控制变量(或称为协变量, covariate, 记为 W)在断点处不发生“跳跃”, 我们在获得 LATE 的一致估计值时便无须控制协变量(Lee and Lemieux, 2010)。然而, 文献中在进行 RDD 估计时加入协变量主要出于两点考虑: 一方面, 加入协变量可以对 RDD 结果的稳健性提供检验, 即如果协变量满足连续性条件, 那么加入协变量将不会显著改变 RDD 的估计结果; 另一方面, 加入协变量将有助于减少估计值的抽样变异性

(sampling variability), 即令 RDD 估计值更接近“真实”值。本文参照文献传统, 将不加入协变量的估计结果作为基准结果, 而将加入协变量的估计结果作为稳健性检验。⁸

DID 与 RDD 的思路稍微不同: 我们使用同样的处理变量 D 将样本分为两组——即独生子女组与非独生子女组, 分别考察这两组人群在“一胎化”政策实施前($T=0$)和后($T=1$)教育程度变化的趋势, 然后比较两组人群教育水平的变化趋势的差距。通过双重差分, 我们去除了两组人群教育水平的时间趋势, 从而得到“一胎化”政策对所有独生子女教育水平的直接(平均)影响——即平均处理效应(average treatment effect, ATE)。

$$Y = \beta_0 + \beta_1 D + \beta_2 T + \beta_3 (D \times T) + \beta_4' W + e. \quad (9)$$

式(9)是具体的回归方程, 其中, $(D \times T)$ 表示处理变量与政策实施前后分组变量的交叉项, 而 β_4 则是控制变量的系数向量。

四、数据与统计描述

本文所使用的数据来自 2005 年全国 1% 人口抽样调查。该调查是在国务院和地方各级政府的领导和组织协调下进行的, 国家统计局负责落实相关调查工作并制订具体计划, 调查经费由中央和地方政府共同承担, 这使数据的样本量和质量获得了较好的保证。该调查的样本以 2000 年全国人口普查为基础, 从中抽取全国 1% 人口作为调查对象, 采取分层次(省级、县级、乡级和社区)、按概率比例、整群随机抽样的方法, 最终获取来自全国 31 个省、自治区和直辖市(除香港、澳门、台湾地区外)的 1300 多万个样本。本文所使用的数据为总体样本量的 1/5, 即从总样本中随机获得的 260 多万个观测个体。该调查的家庭和个人问卷分别收集受访家庭和受访个体的信息, 包括家庭构成和住房情况、人口学特征(出生年月、兄弟姐妹个数等)、社会经济状况以及婚姻、教育、就业、收入、迁移流动等情况。

本文首先将样本限定在 1980 年 10 月前后 5 年内出生于中国(除香港、澳门和台湾地区)的汉族人口⁹, 并在主回归中根据郭志刚等(2003)排除当时没

⁸ 在非参数估计中控制协变量的办法是: 先用结果变量 Y 对协变量 W 进行回归, 然后计算预测值 $\bar{Y}$, 再用结果变量减去预测值得到新的结果变量 $Y^* = Y - \bar{Y}$ 。根据 Lee and Lemieux(2010), 如果协变量 W 不受处理变量 D 的影响(即 W 在断点处连续), 则由协变量 W 引起的 Y 的变化不会影响对 τ 的一致性估计。然而, 如果 W 在断点处发生“跳跃”, 那么 RDD 的估计结果将发生偏误。

⁹ 这里保留断点前后 5 年的样本是为了满足稳健性检验的需要。由于在“证伪”检验中断点位置的移动会导致带宽范围的相应变化, 因此如果仅保留断点前后 3 年的样本会导致证伪检验中所涵盖的数据不足。在主回归中(即以 1980 年 10 月为断点), 我们将样本限定在断点前后的 3 年内。

有普遍推行“一胎化”政策的地区¹⁰；其次，由于政策主要针对城镇居民和国家干部、职工，因此本文进一步将样本限定在城市地区并根据职业和行业信息排除了农业从业者样本；最后，为了排除干扰，本文还根据户口所在地、5年前常住地等信息将部分流动人口排除在外，因为我们无法确定这些流动人口的出生地从而无法确认他们在出生时是否受到“一胎化”政策的影响。最终，本文获得 30 681 个研究样本，其中出生日期距离 1980 年 10 月在 3 年以内的样本有 20 187 个(它们将被用于带宽为 3 年的基准模型分析，而其余样本将参与稳健性检验)。

表 2 是对带宽为 3 年的研究样本进行的描述统计，分为总体样本、非独生子女女子样本和独生子女女子样本。其中，主回归中的结果变量为受访个体的受教育年限¹¹，由问卷所反映的教育程度信息所得，该变量反映了个体的人力资本水平。可以看到，非独生子女样本的平均受教育年限较低(均值为 11.96 年，在统计意义上显著低于独生子女样本的均值即 13.76 年)。由于本文选取的样本较为年轻(22—28 岁)，1980 年 10 月以前出生的人可能会选择工作一段时间再重返学校而导致其连续受教育时间被低估。¹²为了消除“时间趋势”及可能的“季节性”影响，我们通过线性回归去掉了由年龄和出生季节所导致的受教育年限的变化和波动，并将调整后的变量称为“补偿受教育年限”(compensated years of education)，从而更好地保证了断点前后样本的可比性。调整以后，总体样本及子样本的受教育年限在均值意义上均没有发生明显的变化。值得注意的是，“时间趋势”不会对断点处的“跳跃”产生影响，但是“季节性”可能对“跳跃”产生影响(Buckles and Hungerman, 2013; Angrist and Krueger, 1991)，因此，相比于未去除与季节性相关的潜在干扰因素的受教育年限，“补偿受教育年限”更能体现“一胎化”的政策冲击。相应地，在主回归中，我们仍然使用原始的受教育年限作为结果变量，但在加入控制变量的稳健性检验中，我们将引入年份和各个季节的指示变量，以对时间趋势和季节波动进行控制。图 2 中我们以 1 个月为箱体(bin)，描绘了在断点前后出生的样本其“补偿受教育年限”的变化趋势。我们能发现，在断点(1980 年 10 月)附近，“补偿受教育年限”向上发生了较大的“跳跃”。¹³

¹⁰ 根据郭志刚等(2003)，这些地区包括河南、广西、甘肃、河北、内蒙古、云南、青海、宁夏、海南、新疆等。由于这些地区大都施行“允许二孩或二孩以上”的生育政策，因此不属于本文所研究的范围。在后文的安慰剂检验中，我们将使用这部分被剔除的样本进行“证伪”检验。

¹¹ 这里实际上是“连续”受正规学校教育的年限，根据最高学历对应年数计算(在读生视为完成学业)。对于肄业、辍学的样本，我们在计算受教育年限时减去其最高学历阶段对应年限的一半(如大学本科辍学教育年限算为 14 年)。值得注意的是，我们考量一个变量的连续性并不是看变量的取值是否连续，而是看这个变量随指派变量的变化在每个“箱体”所对应的平均水平上是否连续。

¹² 由于 2010 年全国人口普查数据和 2015 年全国人口抽样调查数据都没有反映受访者“是否为独生子女”的信息，并且在微观层面上无法与 2005 年的数据进行合并，因此无法用于本文的研究。

¹³ 我们也对独生子女率进行了“补偿”，即去掉“是否为独生子女”这个变量的时间趋势和季节波动，并绘制了相应的断点图。结果显示，“补偿独生子女率”同样有明显的“跳跃”。

表 2 主要变量的样本描述统计

变量	定义	全样本	非独生子女	独生子女
eduyear	受教育年限(年)	12.66 (2.889)	11.96 (2.894)	13.76* (2.517)
yhat	补偿受教育年限(年)	12.68 (2.883)	11.97 (2.873)	13.81* (2.515)
nosibs	是否为独生子女(是=1,否=0)	0.387 (0.487)	0.000 (0.000)	1.000* (0.000)
healthy	是否健康(是=1,否=0)	0.997 (0.057)	0.996 (0.060)	0.997 (0.051)
married	是否有配偶(是=1,否=0)	0.384 (0.486)	0.435 (0.496)	0.302* (0.459)
male	是否为男性(是=1,否=0)	0.504 (0.500)	0.503 (0.500)	0.506 (0.500)
N	样本量	20 187	12 375	7 812

注：* 表示独生与非独生子女样本的均值在 5% 的置信水平上有显著的区别；括号内为标准差。

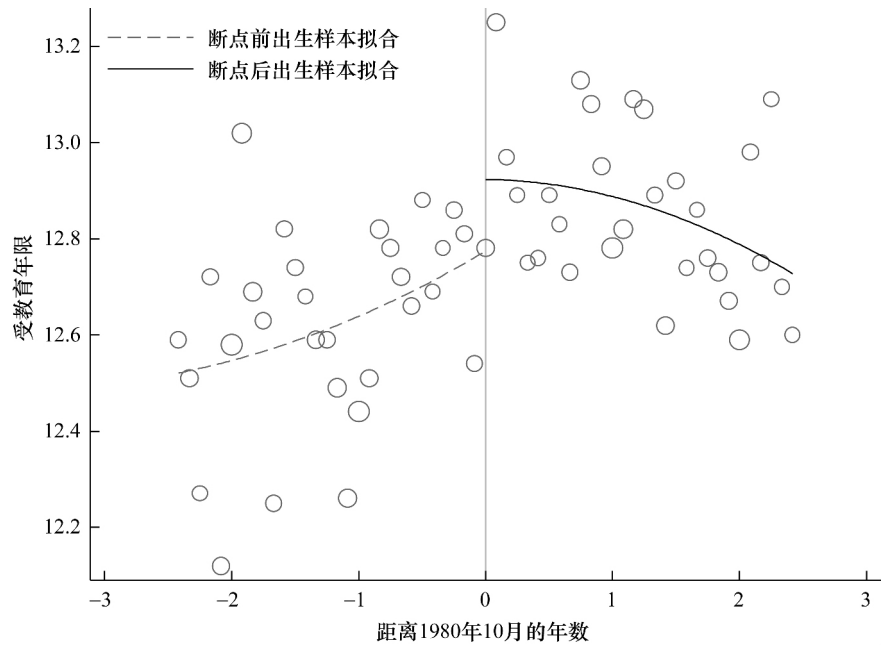


图 2 样本“补偿受教育年限”在断点前后的变化

注：图中散点表示根据每个月出生的样本估计的平均“补偿受教育年限”，即去掉了时间和季节因素后的受教育年限；图中的两条直线分别代表对断点前和断点后两个时期的散点进行非线性回归所得到的拟合线。

为了检验 RDD 估计的稳健性，本文还选取省份、性别、是否有配偶和健康评分等变量作为协变量(covariate)，因为这些变量可能会对个体的人力资本

水平产生直接影响。¹⁴ 其中, 省份变量根据《中华人民共和国行政区划代码》(GB/T2260)构造, 是否有配偶直接来自调查问卷的婚姻状况信息, 而健康评分则是按照雷晓燕等(2010)的方法构造的二值虚拟变量(1 表示健康, 0 表示不健康)。从表 2 还可以看到, 非独生子女样本(对照组)与独生子女样本(处理组)在健康水平和男性比例上均无较大差异。尽管“教育-健康”交互作用(education-health gradient)已经被大量经济学文献证实(Smith, 2004), 但本文并未发现健康状况在两个人群中有明显的区别, 说明样本个体受教育年限的区别不能被健康状况的区别所解释, 这也为 RDD 估计结果的稳健性提供了间接支持。另外, 独生子女样本中的男性比例略高, 这在一定程度上反映了计划生育导致家庭的“性别选择”(Zeng *et al.*, 1993; Das Gupta, 2005; Ebenstein, 2010; Li *et al.*, 2011), 但由于本文所选取的样本是在“一胎化”政策前后较短时间(3 年)内出生的, 这一问题在短期内并不明显; 事实上, 我们发现“一胎化”政策后出生的样本中男性比例更低(这与国家统计局公开发布的分年龄段性别比例的宏观数据一致), 因此我们的估计结果应该较少地受到男性比例增加的影响。然而, 对于是否有配偶的情况, 两组有较大的差异, 独生子女样本的配偶比例显著更低(对照组与实验组有配偶的比例均值分别为 43.5% 和 30.2%), 为了减少婚姻状况的潜在干扰, 本文将使用“熵平衡”(entropy balancing)¹⁵ 的办法计算权重, 对主回归估计结果进行加权调整, 以验证稳健性。接下来, 我们将使用 RDD 方法估计个体人力资本水平与其是否为独生子女之间的关系, 并对估计结果进行多种稳健性检验。

五、估计结果与稳健性讨论

(一) “一胎化”政策与子女教育程度的提高

前文通过图 1 和图 2 展现了独生子女率和受教育年限在断点前后发生跳跃的情况, 本节将通过 RDD 方法更为严格地估计这一跳跃在多大程度上反映了两者之间的因果关系。根据 Lee and Lemieux(2010), 标准的 RDD 估计不需要加入协变量即可根据式(1)与式(5)一式(8)得到一致的 LATE 估计值, 相关结果如表 3 第(1)列所示; 进一步地, 为了控制其他因素对人力资本积累的影响并检验基准 RDD 结果的稳健性, 我们在表 3 第(2)列中提供加入协变量的 RDD 估计结果, 这些协变量包括年份、季节、省份、性别、健康及婚姻状况等; 另外, 为了检验不同内核函数对非参数估计的影响, 我们在表 3 第

¹⁴ 本文认为, 是否有配偶会对个体的“连续”受教育年限产生影响, 因为本文所选取的样本较为年轻, 而相应年龄段的连续学业往往可能因婚姻而暂时中断。

¹⁵ “熵平衡”区别于“倾向得分匹配”方法, 一方面是直接匹配自变量的均值(一阶矩)甚至更高阶矩, 另一方面是通过加权的方式实现匹配, 而不通过样本选择来实现匹配, 具体可参见 Hainmueller(2012)。

(3)列中给出了使用矩形内核(rectangular kernel)进行非参数估计得到的 RDD 估计结果。表 3 第(4)列展示经过“熵平衡”加权调整后的结果,即在回归中设置了使得处理组与对照组有配偶的比例的均值相等的权重。¹⁶

表 3 “一胎化”政策对子女教育水平影响的 RDD 估计结果

估计量及设定	(1)	(2)	(3)	(4)
受教育年限的变化	0.287*** (0.099)	0.295** (0.120)	0.383*** (0.101)	0.251** (0.115)
独生子女概率的变化	0.049*** (0.017)	0.055*** (0.020)	0.0648*** (0.0174)	0.045** (0.021)
加权 LATE	5.888** (2.396)	5.335** (2.560)	5.922*** (1.849)	5.551* (3.031)
是否加入协变量	否	是	否	否
是否“熵平衡”加权	否	否	否	是
内核选择	三角内核	三角内核	矩形内核	三角内核
临界值	1980 年 10 月	1980 年 10 月	1980 年 10 月	1980 年 10 月
最优带宽	2.4 年	2.3 年	1.9 年	1.9 年
样本量	30 681	30 681	30 681	30 681

注:***、**和*分别代表1%、5%和10%水平显著;括号内的标准误差(standard error)是用bootstrap方法计算得到的;加权LATE(即局部Wald估计)根据非参数方法基于不同内核获得;最优带宽的选择依据Imbens and Kalyanaraman(2012)的方法;第(4)列估计进行了加权,而权重是利用处理组与对照组在是否有配偶这个变量上的“熵平衡”计算出来的。

可以看到,表3第(2)—(4)列与表3第(1)列的估计结果没有显著区别:“一胎化”政策的推出使样本成为独生子女的概率在短时间内迅速地提高,提高幅度为5%左右(相当于Wald估计量的分母);而出生在该断点两侧的样本在2005年时所具有的教育水平(以未经调整的受教育年限衡量)同时显示了向上的跳跃,其提高幅度为0.3年左右(即Wald估计量的分子)。根据局部Wald估计方法,我们可以得到“成为独生子女”对受教育年限的影响,即在顺从者中“一胎化”政策对个体人力资本水平的局部平均处理效应(LATE),它可以被解释为人口数量与质量的因果关系(而非简单的相关关系)。

从表3可以看到,对于“一胎化”政策的顺从者而言,成为独生子女将使其受教育年限提高5.5年左右。由此可见,家庭子女数量的减少能够导致子女人力资本水平的提高,这与经典理论“数量-质量”替代的预测一致。根据Becker and Lewis(1973)的理论,在不进行人口控制的情况下,家庭受预算约束等条件的限制,通过最大化目标效用函数决定其最优的子女生育数量。对于顺从者家庭,最优的子女生育数量应该大于1,然而受“一胎化”政策影响,这些家庭只能生育一胎,这就使原来效用函数最大化问题中的“内点解”

¹⁶ 我们同时还做了带宽为1年、2年和3年的回归;各个带宽的结果均为显著且差距较小,说明RDD估计是稳健的。限于空间,这些结果未在本文详细提供,感兴趣的读者可向作者索取。

变为“角点解”，并同时改变了家庭其他决策达到最优的条件。根据边际替代率递减规律，在资源有限的情况下，子女数量的减少使家庭对子女人力资本的最优投资水平相应提高，从而有利于每个子女个体人力资本的上升。当然，我们在理解这个结果时应该注意：（1）该结果是基于特定的人群、历史环境和经济条件得到的，说明中国在二三十年前特定的发展状况下，的确存在“数量-质量”替代，至于具体是什么条件引起了这一关系，在中国当前经济发展水平和人口环境下是否还存在这一关系，还需要进一步的探讨（后文将有所涉及）。（2）这里识别的是“一胎化”政策对于顺从者这一特定人群的影响，而不是对全体人口的影响。正如前文所述，顺从者群体指的是“一胎化”政策实施之前本打算生育两个或两个以上子女但该政策实施后转而只生育一个子女的家庭；RDD 所估计的 LATE 反映的是该群体中生育决策的被动变化所导致的子女平均人力资本的变化率。尽管该结果仅代表“局部影响”而非“全局影响”，但它是在局部受控条件下基于“断点跳跃”思想所得到的因果关系的估计，具有较好的准确性和可信度。可以预期，如果将该“局部影响”推广到全体人群，那么由于“非顺从者”的存在（例如政策前后均打算生育一个子女的家庭、未受政策影响的“超生”家庭等），该政策效果的估计值将有所下降。（3）我们还需要注意以上 LATE 估计值所反映的是“多孩”（两个或两个以上子女）到“一孩”这一生育决策变化所带来的“累积影响”（cumulative effect），而并非子女数量降低一单位的“边际影响”（marginal effect），即它代表“一胎化”政策对于顺从者群体的总效应；可以预期，降低一个子女所对应的平均边际影响将低于此 LATE 估计值。

（二）稳健性检验

在第五部分第（一）节中，我们已经检验了 RDD 估计结果在加入协变量、选择矩形内核与进行“熵平衡”加权这三种情况下的稳健性。接下来，我们还需进行以下几项“例行检查”。

首先，为了检验 RDD 的有效性，我们需要对个体能否精确操控指派进行检验（与 IV 方法不同，RDD 对处理变量本身的外生性要求并不苛刻，而只需假定样本个体无法精确操控指派变量）。具体而言，我们只要求造成断点的政策冲击是不可被“精确预期”的即可。我们承认，一部分有条件的家庭可能通过“内部信息”对政策具有预期，但考虑当时的历史背景，我们认为对于研究样本整体而言，“一胎化”政策仍然是不可被“精确预期”的。¹⁷为了考察是否存在对“一胎化”政策的“精确预期”进而导致的对子女出生时间的

¹⁷ 即使有条件的家庭因为预期到了“一胎化”政策而提前生育了“二胎”，但他们的子女拥有更好的条件接受更多教育，因此虽然断点前独生子女数量被低估（导致 RDD 高估），但断点后非独生子女的受教育程度被高估了（导致 RDD 低估），两者作用方向相反，可以在一定程度上相互抵消各自对 RDD 估计结果的影响。

精确操控，我们按照文献中一般的做法，考察指派变量密度分布的连续性 (McCrary, 2008; Imbens and Lemieux, 2008)：若密度分布不连续，则可能存在对指派的精确操控，否则可认为没有精确操控。按照 McCrary(2008) 和 Lee and Lemieux(2010) 的方法，图 3 给出了指派变量(样本个体出生日期距离断点的时间)的密度分布图。可以看到，无论密度函数曲线还是其置信区间，在断点处都几乎重合，即没有显著的跳跃。按照 McCrary(2008) 的计算方法，我们发现断点前后的密度差异仅仅为 0.012(标准误差为 0.042)，这说明样本个体出生时间(月度数据)的分布是连续和平滑的，即个体没有对指派变量进行精确操控。

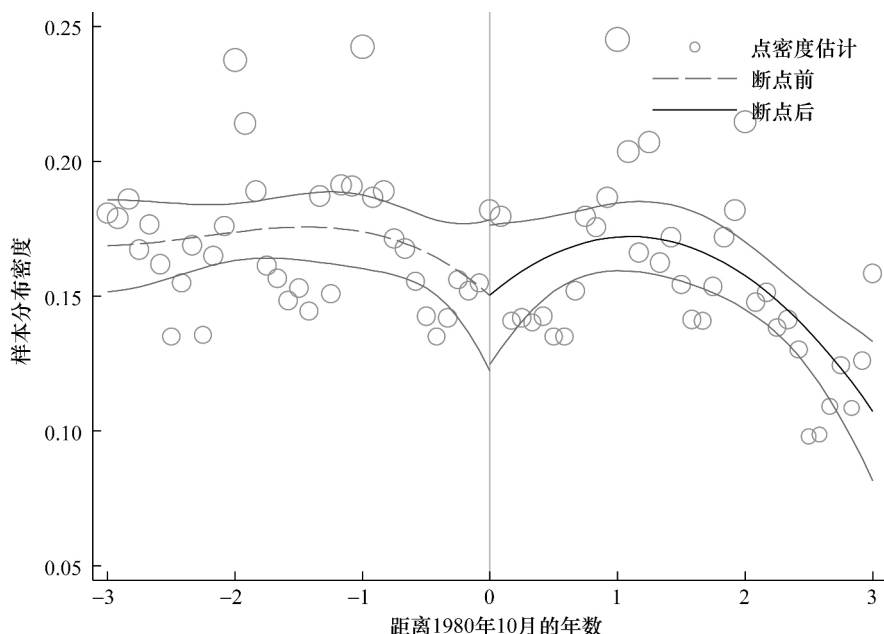


图3 研究样本出生月的密度分布在断点前后的变化

注：本图基于 McCrary(2008) 建议程序 DCdensity 绘制，断点前后拟合曲线上下为 95% 置信区间。

其次，我们需要检验其他控制变量(包括性别、是否有配偶、健康二值虚拟变量和省份分组虚拟变量)在断点处的密度分布连续性：如果这些控制变量在断点处也发生“跳跃”，那么结果变量的“跳跃”就不能被处理变量的“跳跃”完全解释，从而因果推断也就失去效力。为此，我们利用 RDD 中局部线性回归(local linear regression)的思想，将原来的结果变量替换为这些控制变量，并得到表 4 中这些协变量在断点前后的变化值。可以看到，这些变量在断点前后均没有发生显著的变化。

表 4 对主要控制变量的连续性检验(基于 RDD 回归)

主要控制变量	最优带宽	带宽 1 年	带宽 2 年	带宽 3 年
male	-0.031 (0.023)	-0.028 (0.027)	-0.019 (0.020)	-0.000 (0.016)
married	0.039 (0.023)	0.043 (0.026)	0.022 (0.019)	0.001 (0.015)
healthy	0.002 (0.003)	-0.000 (0.003)	-0.001 (0.002)	-0.001 (0.002)
省份分组变量	均不显著			

注：本表的估计均不控制协变量，是将 RDD 的结果变量分别替换为协变量后，在最优带宽下得到的估计结果；显著性标识、标准误差的计算、内核的选择、最优带宽的选择均与表 3 第(1)列相同。

再次，我们将检验 RDD 结果在其他假想临界值点处是否显著(即证伪检验)：如果估计结果在其他临界值处也显著，那么很可能原始 RDD 的结果并非由处理变量的“跳跃”所导致。为此，我们在原断点(1980 年 10 月)前后 1 年内随机选取了其他几个假想时间断点进行回归，表 5 给出了这些假想断点的 RDD 估计结果。可以看到，对于这些时点，LATE 估计值均不显著，这说明表 3 中 RDD 估计结果的显著性并不是由大样本量造成的，从而进一步验证了第五部分第(一)节中主要结论的稳健性和有效性。

表 5 基于其他假想断点的 RDD 估计结果

估计量	1979 年 10 月	1980 年 4 月	1981 年 4 月	1981 年 10 月
受教育年限的变化	0.089 (0.092)	0.086 (0.103)	-0.030 (0.132)	-0.022 (0.113)
独生子女概率的变化	0.028* (0.015)	0.008 (0.017)	0.036 (0.0229)	0.009 (0.020)
加权 LATE	3.136 (3.223)	10.66 (22.78)	-0.820 (3.865)	-2.492 (15.51)
样本量	30 681	30 681	30 681	30 681

注：(1) 假想断点的选择是随机的，我们选取了 1980 年 10 月之前的 6 个月、12 个月和之后的 6 个月、12 个月作为假想的断点；(2) 显著性标识、标准误差的计算、加权 LATE 的计算(内核的选择)、最优带宽的选择、协变量的选择(不加入协变量)均与表 3 第(1)列相同。

另一个值得讨论的问题是：母亲在生下一胎之后往往需要 18—24 个月的生理恢复期，而如果“一胎化”政策恰好在此恢复期内推出，则该断点之前的样本独生子女率也可能受到政策的影响。在这里我们有必要做出相应的解释：一方面，正如表 5 所示，我们在断点之前的各期并未发现样本独生子女概率发生明显的“跳跃”现象，说明这样的样本即使存在也只占少数；另一方面，RDD 方法所估计的是“一胎化”政策在顺从者(complier)之中的影响，而顺从者必须是在没有“一胎化”政策的情况下在断点处能够“自由”进行生育决策的父母，因而那些由于生理或其他原因而无法在断点处生育的父母样本并非 RDD 结果所适用的对象，其存在并不会影响 LATE 的估计结果。

最后，本文使用文献中传统的 2SLS 方法来实现断点回归设计，从而考察

这种基于参数估计的结果与上文中的非参数估计结果是否有显著不同。根据式(2)和式(3)，以样本个体是否出生在1980年10月之后(T)为工具变量，得到2SLS估计值如表6所示(表6的第一行和第二行估计结果分别与表3第(1)、(2)列的第三行估计结果对应，分别表示在不控制和控制协变量的情况下，独生子女率对教育年限的局部平均影响)。可以看出，与表3第(1)、(2)列相比，使用2SLS进行参数估计得到的LATE结果区别不大，并且基本都具有统计显著性。这从另一方面再次证明了主要结果的稳健性。

表6 关于“一胎化”政策影响的RDD估计结果(IV/2SLS估计)

是否加入协变量	带宽 2.4 年 (1)	带宽 1 年 (2)	带宽 2 年 (3)	带宽 3 年 (4)
否	4.729*** (1.777)	5.540* (3.058)	5.685*** (1.662)	4.338*** (1.008)
是	6.034* (3.239)	4.379* (2.452)	7.016** (3.476)	5.965* (3.170)
样本量	16 276	7 040	14 189	20 174

注：***、**和* 分别代表在1%、5%和10%的水平显著；括号内为稳健标准误差(robust standard error)；表中的估计结果即为式(2)中处理变量 D 之前的系数(二阶段回归的主要系数估计值)；第(1)列对应的带宽(2.4年)为依据 Imbens and Kalyanaraman(2012)的方法选取的最优带宽(与表3第(1)列一致)。

(三) 双重差分的思路

另一个可能的估计策略则是通过DID方法来估计“一胎化”政策对于所有独生子女的受教育年限的影响(而不仅仅局限于“顺从者”)。表7提供基于式(9)的回归结果：与表6类似地，我们同样提供加入协变量与不加入协变量的结果；不同的是，这里估计的是交叉项前的系数，该系数表示的是“一胎化”政策使得独生子女相比于非独生子女在平均意义上多提高的受教育年限——该系数不是LATE，而是平均处理效应(ATE)，即将RDD中对于“顺从者”的局部效应平均到了所有的独生子女。可以看到，正如第五部分第(一)节所预测的，若将“局部影响”推广到全体人群，该政策效果的估计值大幅度下降了，但方向不改变，也即“一胎化”政策确实提高了子女的受教育程度。

表7 关于“一胎化”政策影响的DID估计结果

是否加入协变量	带宽 2.4 年 (1)	带宽 2 年 (2)	带宽 3 年 (3)	带宽 5 年 (4)
否	0.411*** (0.086)	0.327*** (0.092)	0.449*** (0.078)	0.616*** (0.065)
是	0.267*** (0.082)	0.189** (0.088)	0.320*** (0.074)	0.504*** (0.063)
样本量	16 276	14 189	20 174	30 681

注：***、**和* 分别代表在1%、5%和10%的水平显著；括号内为稳健标准误差(robust standard error)；表中的估计结果即为式(9)中交叉项 $D \times T$ 之前的系数；第(1)列对应的带宽(2.4年)为依据 Imbens and Kalyanaraman(2012)的方法选取的最优带宽(与表3第(1)列一致)。

(四) 安慰剂 (placebo) 检验

通过以上稳健性检验,我们肯定了 RDD 回归结果的有效性。然而,我们无法排除这一显著结果是否掺杂了其他政策的影响,或者说样本受教育年限的“跳跃”是否主要来自“一胎化”政策。一个潜在的干扰因素是中国在 1999 年施行的“高校扩招”政策:根据邢春冰和李实(2011),该政策显著提高了 1980 年之后出生的人群进入大学的概率,进而有可能导致断点后样本教育水平的提高。为了证明我们的结果并不主要由“高校扩招”政策所致,我们需要进行一组“安慰剂”检验。该检验的基本思路是:考察那些直接受“高校扩招”政策影响但不受“一胎化”政策影响的样本,看其受教育水平在断点前后是否发生“跳跃”;如果不发生“跳跃”,则可间接说明“高校扩招”等其他政策并不是导致样本教育水平变化的主要原因。

第一,我们选用研究样本所在省份的少数民族受访者进行如同表 3 第(1)列的回归。选择该样本的原因是:他们与主回归样本所在的省份相同,受到其他政策影响的情况应该基本一致,只是他们不会受到“一胎化”政策的直接影响(“一胎化”政策不适用于少数民族人口,但“高校扩招”则是适用于所有适龄人口的普惠性政策)。因而,他们受教育水平的跳跃情况能够反映其他政策因素的影响。从表 8 可以看出,该样本的独生子女率或受教育水平在断点前后没有显著跳跃,从而否定“高校扩招”等其他政策导致我们主要结果的推测。

表 8 关于“非一胎化”政策影响的安慰剂检验之一

估计量及设定	带宽 2.3 年 (1)	带宽 1 年 (2)	带宽 2 年 (3)	带宽 3 年 (4)
受教育年限的变化	0.014 (0.208)	0.030 (0.324)	0.009 (0.227)	-0.044 (0.183)
独生子女概率的变化	-0.008 (0.023)	-0.023 (0.035)	-0.013 (0.025)	-0.007 (0.020)
加权 LATE	-1.590 (25.98)	-1.319 (14.63)	-0.678 (18.31)	6.095 (26.37)
是否加入协变量	否	否	否	否
临界值	1980 年 10 月	1980 年 10 月	1980 年 10 月	1980 年 10 月
样本量	10 035	2 306	4 609	6 717

注:(1)这里选取了研究样本所在省份的少数民族受访者进行了与表 3 第(1)列相同的回归;(2)***、**和*分别代表在 1%、5%和 10%水平显著;括号内的标准误差(standard error)是用 bootstrap 方法计算得到的;加权 LATE(即局部 Wald 估计)根据非参数方法基于三角内核(triangle kernel)获得;第(1)列为根据 Imbens and Kalyanaraman(2012)的方法选取最优带宽后得到的估计结果。

第二,我们选取在主回归中剔除的省份地区的汉族样本进行与表 3 第(1)列相同的估计(这些地区在当时多推行的是“允许二孩或二孩以上”的生育政策)。尽管这部分样本与主回归样本所在省份不同,不受“一胎化”政策的直

接影响,但他们应该同样受到“高校扩招”这一全国性教育改革政策的影响。同时,考虑到中国的教育政策往往对少数民族群体有特殊优惠,因此使用这部分地区的汉族样本来进行“安慰剂”检验可能更具说服力。基于表9的估计结果,我们同样没有发现独生子女率或受教育水平有任何“跳跃”现象——这再次表明,本文的RDD结果主要是由“一胎化”政策而非“高校扩招”等其他政策所引起的。综合以上两组“安慰剂”检验结果,我们认为,虽然“高校扩招”政策会对我们的主要回归带来一定程度的干扰,但是它不应该是导致样本受教育年限“跳跃”的主要原因。¹⁸

表9 关于“非一胎化”政策影响的安慰剂检验之二

估计量及设定	带宽 2.4 年 (1)	带宽 1 年 (2)	带宽 2 年 (3)	带宽 3 年 (4)
受教育年限的变化	0.033 (0.066)	0.061 (0.102)	0.022 (0.073)	0.020 (0.060)
独生子女概率的变化	0.008 (0.008)	0.013 (0.012)	0.008 (0.009)	0.008 (0.007)
加权 LATE	4.434 (9.066)	4.849 (8.464)	2.760 (8.914)	2.624 (7.697)
是否加入协变量	否	否	否	否
临界值	1980 年 10 月	1980 年 10 月	1980 年 10 月	1980 年 10 月
样本量	61 417	13 070	26 175	37 996

注:(1)这里选取了研究样本之外省份的汉族受访者进行了与表3第(1)列相同的回归;(2)***、**和*分别代表在1%、5%和10%水平显著;括号内的标准误差(standard error)是用bootstrap方法计算得到的;加权LATE(即局部Wald估计)根据非参数方法基于三角内核(triangle kernel)获得;第(1)列为根据Imbens and Kalyanaraman(2012)的方法选取最优带宽后得到的估计结果。

六、影响机制分析：分样本回归

在上文中,我们发现了顺从者家庭中人口数量与质量的负向因果关系,为了更好地理解这一结果的作用渠道和特定历史条件,以及“数量-质量”替代成立的前提条件,接下来我们将进一步考察这一关系在不同子样本中的具体情况,以试图对影响机制给出一些解释。

(一) 性别分样本回归

我们排除既有男孩又有女孩的家庭,将样本限定在所有子女的性别都相同的家庭(称为“纯性别子女家庭”)。表10显示,“纯性别”子样本的RDD估计结果更加显著,其统计显著水平从5%上升为1%;进一步考察“纯性别

¹⁸事实上,我们利用2000年全国人口普查数据考察了各个年龄段在1999年高校扩招后的在校分布情况,发现在断点前后一年出生的样本中,受“高校扩招”直接影响的高中、中专在校生比例相当。

子女家庭”中的“男孩”子样本和“女孩”子样本可以看到,人口数量与质量的替代关系只在家庭生育的子女全为男孩的情况下显著;而如果子女全为女孩,RDD估计结果则不显著。

表 10 基于子女性别的分样本 RDD 估计结果

估计量	纯性别子女	男孩家庭	女孩家庭
受教育年限的变化	0.393*** (0.114)	0.561*** (0.172)	0.204 (0.156)
独生子女概率的变化	0.071*** (0.020)	0.059** (0.029)	0.082*** (0.030)
加权 LATE	5.500*** (1.910)	9.548** (4.863)	2.504 (1.921)
样本量	16 535	8 750	7 785

注:(1)男孩家庭指家庭中只有男孩的情况,女孩家庭指家庭中只有女孩的情况,纯性别子女家庭指家庭中所有子女性别都相同的情况(包括独生子女家庭);(2)临界值为 1980 年 10 月,带宽为根据 Imbens and Kalyanaraman(2012)方法选取的最优带宽;(3)显著性标识、标准误差的计算、加权 LATE 的计算(内核的选择)、协变量的选择(不加入协变量)均与表 3 第(1)列相同。

一个有趣的现象是,相比“男孩家庭”,“女孩家庭”的独生子女率跳跃更大,并且更为显著,这可能说明“一胎化”政策实施前,更多的家庭在第一胎是女孩时会打算生育二胎,而第一胎是男孩的家庭则更少地打算继续生育,这从侧面反映了“重男轻女”的现象(当然这也与计划生育的政策有关)。另外,有研究指出“一胎化”政策实际上加重了家庭对男孩的“性别选择”(Zeng *et al.*, 1993; Das Gupta, 2005; Ebenstein, 2010; Li *et al.*, 2011),即“一胎化”政策后男孩的比例应该上升。然而,我们的样本主要是在“一胎化”政策刚推行前后 3 年内出生的,通过“性别选择”使男孩数量增加的现象还不明显,并且实际上数据显示这一时间段出生的男孩比女孩要少。因此,“性别选择”不太可能通过造成样本男孩比例偏高而使 RDD 结果被高估。我们的结果表明,“重男轻女”主要体现在对男孩的重视程度上,即不是通过多生男孩而是通过增加对男孩的人力资本投资来体现对男孩的重视。

在中国“重男轻女”的观念造成部分家庭对男孩给予更多的投入,这就有悖于 Becker and Lewis(1973)等提出的“家庭对子女一视同仁”的假设,这对于我们识别 Q-Q 替代关系是不利的。例如,假设一个家庭中有多个女孩和一个男孩,如果此家庭中父母对子女的人力资本投资(如教育开销)主要集中于此男孩,那么他的人力资本水平可能并不低于来自“独子”家庭(即只有一个孩子且为男孩的家庭)的儿子,从而削弱了数量-质量的负向因果关系。表 10 显示,当我们排除这一可能性而只考虑“纯性别子女家庭”时,LATE 的估计值更显著,说明重男轻女现象对这一关系的识别确实造成了一定的干扰。换言之,“重男轻女”现象的存在有可能导致计划生育对人力资本水平的正向

影响被削弱。¹⁹

(二) 地区人均收入水平、信贷水平的分样本回归

接下来,我们再考察人口数量与质量的替代在收入和信贷水平不同的地域上的异质性。本文首先将研究样本所覆盖的地区按人均收入水平划分为“高收入地区”²⁰和“低收入地区”两组。从表11可以看到,高收入地区的RDD估计结果不具有统计显著性,并且其LATE估计值的大小远远低于落后省份,这说明Q-Q替代关系在高收入地区的样本中并不明显。

表11 基于1979—1981年地区人均收入水平及金融机构贷款余额的分样本RDD估计结果

估计量	高收入地区	低收入地区	信贷发达地区	信贷落后地区
受教育年限的变化	0.191 (0.141)	0.346*** (0.130)	0.250* (0.138)	0.293** (0.128)
独生子女概率的变化	0.045* (0.027)	0.051** (0.020)	0.047* (0.025)	0.047** (0.020)
加权LATE	4.220 (3.329)	6.735** (3.136)	5.317 (3.454)	6.225* (3.218)
样本量	15 564	15 117	15 250	15 431

注:(1)临界值为1980年10月,带宽为根据Imbens and Kalyanaraman(2012)方法选取的最优带宽;(2)显著性标识、标准误差的计算、加权LATE的计算(内核的选择)以及协变量的选择(不加入协变量)均与表3相同。

根据Becker and Lewis(1973)等经典理论,人口数量与质量的负向关系还需要在家庭预算约束的假定下才能推出。对于高收入家庭(以高收入地区子样本为代表),其预算约束的限制较为宽松,即家庭能够承担更多的对子女教育的投资,这就降低了人力资本投资与其他家庭决策之间的边际替代斜率,从而使人口数量与质量的负向关系变得不再显著。表11的结果说明,地区人均收入水平(或家庭收入)和可能是第五部分第(一)节所示结果的重要作用渠道,而随着中国今后经济水平的不断提高,家庭对子女的投资能力将得到改善,从而该数量与质量的替代关系将有可能不再显著。换言之,在过去较低经济发展水平的背景下,我们可能高估了计划生育对人力资本积累的正向影响。

然后,我们再将研究样本所覆盖的地区划分为“信贷发达地区”和“信贷落后地区”这两组²¹,以便于考察人口数量与质量的替代在信贷市场发展

¹⁹ 值得一提的是,非独生子女样本可能来自“多胞胎”家庭,而且父母对“多胞胎”子女往往更加“一视同仁”。这样一来,“重男轻女”的现象将在一定程度上被削弱,从而增强了计划生育对人力资本水平的正向影响。

²⁰ “高收入地区”即1979—1981年人均GDP较高的地区(数据来自《新中国五十五年统计资料汇编》),包括北京、天津、上海、辽宁、黑龙江、江苏、广东。

²¹ “信贷发达地区”即1979—1981年金融机构人均贷款余额较高的地区(数据来自《新中国五十五年统计资料汇编》),分别是北京、天津、上海、黑龙江、吉林、江苏、湖北、山西、山东和陕西。

水平不同的地域的异质性。表 11 的结果表明,一方面,信贷发达地区的 RDD 估计结果同样不再具有统计显著性,尽管受教育年限和独生子女率在断点前后的跳跃程度均有一定的显著性;另一方面,其 LATE 估计值的大小也稍低于信贷落后地区。这一结果表明,人口数量与质量的负向关系在信贷发达地区子样本中可能是不成立的。按照经典理论, Q-Q 替代关系还需要信贷市场失灵的条件才能推出。对于金融机构人均贷款水平较高的地区,其金融资源相对丰富,正式与非正式的融资渠道较多,信贷市场相对发达,这就使得父母在对子女进行投资时较少地受到信贷约束的影响,从而削弱了这些地区的 Q-Q 替代。这说明,地区金融发展水平和信贷市场的健全程度可能是第五部分第(一)节所示结果的重要作用渠道,而随着中国今后金融发展水平的不断提高,受到家庭对子女投资能力改善的影响,这一数量与质量的负向关系将有可能不再存在。换句话说,基于落后的金融制度下的历史数据,我们将一定程度地高估计划生育对人力资本积累的正向影响。

在以上三个分样本回归中,本文分别对经典理论的“父母对子女一视同仁”“家庭面临预算约束”和“信贷市场失灵”这三条假定进行了检验。结果表明,人口数量与质量的替代在缺少任何一条假定的情况下都将被削弱, Q-Q 替代关系在不同维度均存在异质性,这些结论将为我们如何阐释和推广本文的主要实证结果提供有益的借鉴。需要注意的是,受限于数据和研究条件,本文没有对这三条假定背后的具体作用机制进行更深入的分析。我们只能说,缺少“预算约束”和“信贷市场欠发达”这两个条件“数量-质量”替代将难以显现,但不能说“预算约束”和“信贷市场欠发达”就是引起“数量-质量”替代的直接原因——这两条假定的作用机制可能还需要更为详细的微观数据进行揭示,这也为后续研究提供了可能的方向。

七、结论与政策建议

人口数量与质量的替代是重要的经济学理论命题,它对一个国家的人口政策和长期经济发展也具有重要的实践价值和指导作用。自 20 世纪 70 年代关于家庭人力资本投资的经典理论建立以来,经济学家们一直试图通过实证研究来检验人口数量与质量的因果关系,然而受到数据可获得性和统计方法的制约,相关实证文献还比较有限。本文根据中国的历史和制度特点,选用具有全国代表性的人口抽样调查数据,利用人类计划生育史中一次不可多得的自然实验机会,结合新的计量经济学方法——断点回归设计(RDD),估算了人口数量对人口质量(教育水平)的单向因果影响。这是 RDD 方法在相关研究中的新尝试,对该领域的研究在方法论上有所补充。同时,本文的估计结果还在一定程度上揭示了中国家庭人力资本的投资与决定机制,有助于为中国未来的人口政策提供微观经济学理论基础。

随着中国人口老龄化的加剧，人口政策的放松成为社会各界关注的热点话题，十八届三中全会提出的“单独二孩”政策以及随后在十八届中央委员会第五次全体会议提出的“全面二孩”政策更是引发了学界和媒体的激烈讨论。本研究的结果对这一问题的讨论可以提供以下几点启示：第一，基于全样本的多种计量回归结果表明，从整体而言，中国过去三十多年的计划生育政策在控制人口数量和提升人口质量方面均有积极意义，对于这一政策的评估趋向正面。第二，分样本结果显示，“数量-质量”替代关系在不同维度存在异质性，尤其在经济发达、信贷市场较为活跃的省份，这一结论不再稳健。这对我们现阶段的人口政策调整具有现实意义，我们在将本文的主要结论推广到现有政策制定时要非常慎重：过去三十多年的“数量-质量”替代并不意味着在现阶段仍然成立；当经济发展到一定阶段时，家庭的收入和信贷约束有所缓解，子女的培养成本对家庭不再构成负担，这时候人口数量和质量的替代很可能不再显著。第三，通过加快落后地区的信贷市场发展速度、建立低息助学贷款机制等方式缓解家庭对子女的投资约束，将可能减少计划生育的全面放宽对人力资本积累所产生的负面影响，这将有利于中国当前的人口政策调整，降低“人口数量提高导致人口质量下降”的风险。第四，分样本回归结果还显示出中国家庭中普遍存在“重男轻女”的性别歧视现象，这同样不利于中国人口素质的全面提升；因此，进一步贯彻男女平等的理念，并推行更加公平的教育政策也将为进一步放松人口管控提供积极有效的支持。

参 考 文 献

- [1] Angrist, J. D., and A. B. Krueger, "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics*, 1991, 106(4), 979—1014.
- [2] Angrist, J. D., and V. Lavy, "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement", *Quarterly Journal of Economics*, 1999, 114(2), 553—575.
- [3] Angrist, J. D., V. Lavy, and A. Schlosser, "Multiple Experiments for the Causal Link between the Quantity and Quality of Children", *Journal of Labor Economics*, 2010, 28(4), 773—824.
- [4] Banister, J., *China's Changing Population*. Stanford, CA: Stanford University Press, 1987.
- [5] Becker, G. S., and H. G. Lewis, "On the Interaction between the Quantity and Quality of Children", *Journal of Political Economy*, 1973, 82(2), 279—288.
- [6] Black, S. E., P. J. Devereux, and K. G. Salvanes, "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education", *Quarterly Journal of Economics*, 2005, 120(2), 669—700.
- [7] Buckles, K. S., and D. M. Hungerman, "Season of Birth and Later Outcomes: Old Questions, New Answers", *Review of Economics and Statistics*, 2013, 95(3), 711—724.
- [8] Cameron, L., N. Erkal, L. Gangadharan, and X. Meng, "Little Emperors: Behavioral Impacts of China's One-Child Policy", *Science*, 2013, 340(6130), 272—273.

- [9] Chen, Y., A. Ebenstein, M. Greenstone, and H. Li, "Evidence on the Impact of Sustained Exposure to Air Pollution on Life Expectancy from China's Huai River Policy", *PNAS*, 2013, 110 (32), 12936—12941.
- [10] Cooney, R. S., J. Wei, and M. G. Powers, "The One-Child Certificate in Hebei Province, China: Acceptance and Consequence, 1978—1988", *Population Research and Policy Review*, 1991, 10(2), 137—155.
- [11] Das Gupta, M., "Explaining Asia's 'Missing Women': A New Look at the Data", *Population and Development Review*, 2005, 31(3), 529—535.
- [12] Ebenstein, A., "The 'Missing Girls' of China and the Unintended Consequences of the One-Child Policy", *Journal of Human Resources*, 2010, 45 (1), 87—115.
- [13] Fan, J., and I. Gijbels, *Local Polynomial Modelling and Its Applications*. London, New York and Melbourne: Chapman and Hall, 1996.
- [14] Glick, P. J., A. Marini, and D. E. Sahn, "Estimating the Consequences of Changes in Fertility on Child Health and Education in Romania: An Analysis Using Twins Data", Cornell University, Food and Nutritio Policy Program Working Paper No. 183, 2005.
- [15] 郭志刚、张二力、顾宝昌、王丰, "从政策生育率看中国生育政策的多样性", 《人口研究》, 2003 年第 5 期, 第 1—10 页。
- [16] Hainmueller, J., "Entropy Balancing: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies", *Political Analysis*, 2012, 20(1), 25—46.
- [17] Hatton, T. J., and R. M. Martin, "Fertility Decline and the Heights of Children in Britain, 1886—1938", *Explorations in Economic History*, 2010, 47(4), 505—519.
- [18] 黄少安、孙涛, "人口负担与人口红利的权衡——论中国现阶段不能放松计划生育政策", 《学术月刊》, 2013 年第 7 期, 第 66—77 页。
- [19] Imbens, G., and K. Kalyanaraman, "Optimal Bandwidth Choice for the Regression Discontinuity Estimator", *Review of Economic Studies*, 2012, 79(3), 933—959.
- [20] Imbens, G., and T. Lemieux, "Regression Discontinuity Designs: A Guide to Practice", *Journal of Econometrics*, 2008, 142(2), 615—635.
- [21] Lee, D. S., and T. Lemieux, "Regression Discontinuity Designs in Economics", *Journal of Economic Literature*, 2010, 48(2), 281—355.
- [22] 雷晓燕、谭力、赵耀辉, "退休会影响健康吗?", 《经济学》(季刊), 2010 年第 4 期, 第 1539—1558 页。
- [23] Li, H., J. Zhang, and Y. Zhu, "The Quantity-Quality Trade-Off of Children in a Developing Country: Identification Using Chinese Twins", *Demography*, 2008, 45(1), 223—243.
- [24] Li, H., J. Yi, and J. Zhang, "Estimating the Effect of the One-Child Policy on the Sex Ratio Imbalance in China: Identification Based on the Difference-in-Differences", *Demography*, 2011, 48 (4), 1535—1557.
- [25] Li, J., and R. S. Cooney, "Son Preference and the One Child Policy in China: 1979—1988", *Population Research and Policy Review*, 1993, 12(3), 277—296.
- [26] Li, J., "China's One-Child Policy: How and How Well Has It Worked? A Case Study of Hebei Province, 1979—1988", *Population and Development Review*, 1995, 21(3), 563—585.
- [27] 梁中堂, "艰难的历程:从'一胎化'到'女儿户'", 《开放时代》, 2014 年第 3 期, 第 11—44 页。

- [28] Liu, H., "The Quality-Quantity Trade-off: Evidence from the Relaxation of China's One-Child Policy", *Journal of Population Economics*, 2014, 27(2), 565—602.
- [29] 刘永平、陆铭,“放松计划生育政策将如何影响经济增长——基于家庭养老视角的理论分析”,《经济学》(季刊),2008年第4期,第1271—1300页。
- [30] Lordan, G., and P. Frijters, "Unplanned Pregnancy and the Impact on Sibling Health Outcomes", *Health Economics*, 2013, 22(8), 903—914.
- [31] 马骁骁,“从人口老龄化看中国未来计划生育政策——人口老龄化原因的实证分析”,《经济视角》,2010年第5期,第46—49页。
- [32] McCrary, J., "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test", *Journal of Econometrics*, 2008, 142(2), 698—714.
- [33] Park, C., and W. Chung, "Sib-Ship Size, Birth Order, and Children's Education in Developing Countries: Evidence from Bangladesh", *Hitotsubashi Journal of Economics*, 2012, 53(1), 1—23.
- [34] Qian, N., "Quantity-Quality and the One Child Policy: The Only-Child Disadvantage in School Enrollment in Rural China", NBER Working Paper No. 14973, 2009.
- [35] Rosenzweig, M. R., and J. Zhang, "Do Population Control Policies Induce More Human Capital? Twins, Birth Weight and China's 'One-Child' Policy", *Review of Economic Studies*, 2009, 76(3), 1149—1174.
- [36] Smith, J. P., "Unraveling the SES-Health Connection", *Population and Development Review*, 2004, 30, 108—132.
- [37] Van der Klaauw, W., "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach", *International Economic Review*, 2002, 43(4), 1249—1287.
- [38] Willis, R. J., "A New Approach to the Economic Theory of Fertility Behavior", *Journal of Political Economy*, 1973, 81(2), S14—S64.
- [39] 徐剑,《中国人口政策效果分析》,吉林大学博士学位论文,2010年。
- [40] 邢春冰、李实,“扩招‘大跃进’、教育机会与大学毕业生就业”,《经济学》(季刊),2011年第4期,第1187—1208页。
- [41] 杨发祥,《当代中国计划生育史研究》,浙江大学博士学位论文,2003年。
- [42] 余静文、王春超,“新‘拟随机实验’方法的兴起——断点回归及其在经济学中的应用”,《经济学动态》,2011年第2期,第125—131页。
- [43] Zeng, Y., P. Tu, B. Gu, Y. Xu, B. Li, and Y. Li, "Causes and Implications of the Recent Increase in the Reported Sex Ratio at Birth in China", *Population and Development Review*, 1993, 19(2), 283—302.
- [44] Zhang, J., and B. G. Spencer, "Who Signs China's One-Child Certificate, and Why?" *Journal of Population Economics*, 1992, 5(3), 203—215.
- [45] 周俊山、尹银,“中国计划生育政策对居民储蓄率的影响——基于省级面板数据的研究”,《金融研究》,2011年第10期,第61—73页。

The Impact of Birth Control Policy on Offspring's Education Level: Microeconomic Evidence from China

XUEZHENG QIN RUDAI YANG

(Peking University)

CASTIEL CHEN ZHUANG*

(University of Washington)

Abstract The effect of the recent “two-child policy” has attracted a new wave of attention. Using the 2005 China’s inter-census 1% population survey data and multiple econometric methods, we identify the causal relationship between quantity and quality of population in China. The results suggest that being a single child increased his/her educational attainment, but this effect was only evident in regions with lower income and less developed credit markets. With China’s economic development, relaxing households’ constraints on investment in offspring’s human capital may ease the negative impacts of a country-wide relaxation of birth control and thus may promote sustainable economic development.

Key Words birth control, education level, regression discontinuity design

JEL Classification J13, D13, I20

* Corresponding Author; Castiel Chen Zhuang, UW Economics, Savery Hall Room 305, Seattle, WA 98195, United States. Tel: 1-206-487-6429; E-mail: zogcee@gmail.com, cczhuang@uw.edu