

# 灾难之殇:负向外生冲击的教育损失

——以1976年唐山大地震为例

王 骏,李 波

**[摘 要]** 文章使用2005年全国1%人口抽样调查数据,运用双重差分(DiD)模型和工具变量下的局部平均处理效应(LATE-IV)方法,研究了唐山大地震对个体受教育水平和在劳动力市场中表现的影响。研究发现,唐山大地震使得个体受教育年限减少0.14-0.21年,可看作地震的短期教育损失;此外,唐山大地震使得个体受教育年限的减少进而带来的收入损失约为3.51%-4.77%,带来的人力资本损失约占2005年唐山地区生产总值的0.30%-0.41%,可看作地震的长期教育损失。

**[关键词]** 唐山大地震;人力资本;双重差分模型;工具变量;局部平均处理效应

## 一、引言

地震,这种突发性的自然灾害不仅给社会经济发展造成不可估量的损失,也可能对整个社会的人力资本(诸如教育、营养与健康等)和福利水平(诸如收入和消费)带来持久和深远的影响(Baez et al., 2010)。相关研究表明,地震造成了人均收入的暂时性下降,并降低了消费水平(Baez and Santos, 2008; Gignoux and Menéndez, 2014)。也有学者发现地震显著降低了个体的受教育水平。比如,地震降低了入学率并提高了儿童进入劳动力市场的可能性(Santos, 2007),通过导致个体早期营养不良进而影响认知发展和学业成就,降低受教育年限(Bustelo et al., 2012)。还有文献从身体和心理健康等方面探讨地震的影

**[收稿日期]** 2016-04-01

**[作者简介]** 王骏,北京师范大学经济与工商管理学院/首都教育经济研究院,电子邮箱地址:wangj@mail.bnu.edu.cn;李波,北京师范大学经济与工商管理学院/首都教育经济研究院,电子邮箱地址:doctorleebo@sina.com。

**[致 谢]** 作者感谢北京师范大学经济与工商管理学院刘泽云教授和杨娟副教授对本文的指导和建议。谨以此文纪念唐山大地震四十周年,献给这座凤凰涅槃、浴火重生的城市。

响。比如,地震显著提高了人口死亡率,并通过摧毁医疗设施提高了传染性疾病的发病率(Bathia,2008),造成人们抑郁、焦虑等精神病态行为的增加(Chen et al.,2002;Roussos et al.,2005;Şalcioğlu and Başoğlu,2008)。最近的一些研究还关注地震对人力资本的代际影响,这些研究表明,遭受地震影响的父母一辈的子女一代的人力资本也偏低(Caruso,2015;Caruso and Miller,2015)。除此之外,还有许多文献发现诸如干旱、洪涝、台风、海啸等自然灾害显著提高了营养不良的发生率,恶化了身体发育和健康状况,严重阻碍了学业进展,从而降低了受教育水平和劳动生产率(Hoddinott and Kinsey,2001;Baez and Santos,2007;Aderman et al.,2009;Maccini and Yang,2009;Neelsen and Stratmann,2011;Tan et al.,2014)。

唐山大地震是二十世纪世界死亡人数最多的地震<sup>①</sup>。1976年7月28日凌晨3时42分53.8秒,中国河北省唐山市(东经118.2°,北纬39.6°)发生强度为里氏7.8级地震,震中烈度XI度,震源深度为23千米,地震持续时间约为12秒,地震产生的能量相当于400颗广岛原子弹爆炸产生的威力。地震共造成242469人死亡,175797人重伤,4204名儿童成为孤儿。其中,唐山地区震亡148022人,81630人重伤,全家震亡的7218户,2652名儿童成为孤儿。682267间民用建筑中有656136间倒塌和受到严重破坏。唐山大地震对学校的破坏极为严重,严重破坏和倒塌的建筑物面积为42.72万平方米,破坏率高达92.3%。地震造成的唐山市财产损失91亿多元,文、教、科、卫的财产损失超过12亿元,占损失总额的13.5%,损失率高达88.4%。目前,有关唐山大地震对人力资本影响的研究仅有一篇。Xu(2011)基于2000年全国人口普查数据建立县级面板数据,采用双重差分模型研究了唐山大地震的长期影响。研究结果表明,地震显著降低了地区的平均受教育水平,震级较强县区的初中毕业率下降了25.3%。地震造成了受教育群体性别比例的失衡,男女初中毕业率之比提高了96%。但他发现受地震影响的一代人在未来劳动力市场中的失业率并没有显著上升。然而,该研究没有足够关注相关政策在特定历史背景下的城乡差异性,采用的结果变量为人口规模、毕业率和失业率等宏观指标,也没有基于人力资本理论深入研究唐山大地震造成的个体受教育水平的下降与当前收入水平的关系。

本文使用2005年全国1%人口抽样调查中随机抽取的20%的样本,研究

<sup>①</sup> 20世纪地震灾害造成全球160—175万人震亡,约占1万多次重大自然灾害死亡总人数的一半,其中唐山大地震占全球地震灾害震亡总人数的15%左右(于山等,2003)。

了唐山大地震对个体受教育水平和在未来劳动力市场中表现的影响。首先,本文应用双重差分模型,将1961—75年出生组作为实验组,并将1951—60年和1976—85年出生组作为前、后非实验组,同时,为剔除其他外生冲击对估计的干扰,将承德、保定和邯郸作为控制组。估计结果表明,唐山大地震使得个体受教育年限减少0.14—0.21年,可将此看作唐山大地震的短期教育损失。其次,本文将1961—75年出生组的虚拟变量作为受教育年限的工具变量,应用LATE-IV方法,发现受唐山大地震潜在影响的一代人的教育回报率为20.16%—21.79%。为剔除其他外生冲击对估计的干扰,本文将1961—75年出生组的虚拟变量及其与唐山的交互项作为工具变量,估计的教育回报率为20.93%—27.85%。最后,利用唐山大地震的短期教育损失和估计得到的教育回报率,本文估计了唐山大地震的长期教育损失。结果显示,唐山大地震使得个体受教育年限的减少进而带来的收入损失约为3.51%—4.77%,唐山大地震带来的人力资本损失约占2005年唐山GDP的0.30%—0.41%,可将此看作唐山大地震的长期教育损失。

本文的创新之处在于如下四点。首先,本文以唐山大地震为例,为自然灾害对人力资本和福利水平的影响为主题的研究提供了来自中国的新的证据。其次,使用微观调查数据而非宏观层面数据,更为细致地估计了唐山大地震对个体受教育水平的影响。第三,将个体所在出生组作为工具变量,识别了教育与收入的因果关系,估计了受唐山大地震影响一代的教育回报率。最后,本文试图构建唐山大地震损失核算的基本框架,并对唐山大地震的短期和长期教育损失进行了估算。

本文余下部分的内容安排如下:第二部分对数据进行说明,详细介绍本文的样本框和研究设计。第三部分利用双重差分模型探讨唐山大地震对个体受教育水平的影响。第四部分使用LATE-IV方法估计受唐山大地震影响一代的教育回报率。第五部分是对第三、四部分的整合,在给出唐山大地震教育损失核算框架的基础上,计算唐山大地震的短期和长期教育损失。第六部分报告本文的结论,并指出研究的不足之处。

## 二、样本与数据

本研究所用数据为2005年全国1%人口抽样调查中随机抽取的20%的样本,数据处理前的样本量为2585481个。首先,我们需要确认个体受教育水平最有可能受到唐山大地震影响的出生组。这里,本文将唐山大地震对个体受教

育情况的影响期限设定在1976—82年<sup>①</sup>,将受影响的教育阶段设定为小学和初中<sup>②</sup>。当时的入学年龄为7岁,小学学制为5年或6年,初中学制为3年或4年,故本文将1961—75年出生组作为实验组,同时分别将1951—60年和1976—85年出生组作为前、后非实验组。更为详细的样本框设计见附表1。其次,本文将样本限制为具有农业户口的个体,同时剔除调查时居住地在其他县(市、区)和国外、户口登记地为其他县(市、区)和户口不详的个体。第三,本文将样本限制在居住在河北省的家庭户个体。第四,考虑到地理位置的远近,本文将承德、保定和邯郸作为对照组<sup>③</sup>(如图1所示)。最后,由于用到收入数据,本文只保留主要生活来源为劳动收入的个体,删除了个人月收入为0或缺失、未工作、就业身份是家庭帮工的个体。最终数据的样本量为13759,其中唐山为2891。

在变量处理方面,个体受教育年限由受教育水平折算得到。收入用取对数的月收入衡量。根据调查问卷中的变量“与户主关系”,将调查时与父母同住或户口与父母在同一户籍的个体的父母的受教育年限匹配到相对应的个体上。相关变量的描述统计见附表2。可以看出,各地区男性比重略高于女性,平均年龄约为37岁,平均受教育年限为



图1 河北省地图

<sup>①</sup> 已有史料没有对唐山大地震的影响期限进行明确界定,本文只能进行合乎实际的推算。据《唐山大地震震后救援与恢复重建》一书记载,《唐山城市总体规划》于1977年获国务院批准,又在1978年3月和1979年9月进行了两次修订。当时提出的目标是“一年准备初步开展,三年大干,一年扫尾,到1982年全部建成”。故较为合理的猜测是,唐山大地震的影响期限为1976—82年,即唐山大地震对个体受教育水平的影响期限为6年。但实际上,受到文化大革命拨乱反正、地质构造复杂等因素的影响,重建的过程要远比预计的漫长,正式的重建工作到1986年底才基本完成。

<sup>②</sup> 20世纪70年代,中国平均受教育水平非常低。世界银行数据显示,1975年、1980年和1985年,中国的平均受教育年限分别为3.97年、4.75年和5.25年。《新中国五十年统计资料汇编(1949—98)》显示,1976—80年,河北省学龄儿童入学率分别为98.0%、97.4%、96.7%和96.5%,农村地区则会大幅度低于河北省平均水平。《新河北60年(1949—2009)》显示,1980年河北省小学毕业生升学率仅为82.3%,农村地区和初中毕业生升学率会更低。鉴于此,本文将唐山大地震影响的教育阶段设定为小学和初中。

<sup>③</sup> 本文没有选择河北省的省会石家庄作为控制组,是因为省会城市在特定历史时期的特殊性可能会影响估计结果;没有选择廊坊和沧州,是因为这两个地区邻近唐山大地震的震中(即路南区、丰南区一带),也遭受了地震带来的巨大损失。据《唐山大地震震后救援与恢复重建》一书记载,天津、廊坊和沧州震亡约2万余人。

8.31年,唐山为8.46年,略高于样本平均水平。平均收入约为425.25元,唐山为558.60元,大幅度高于样本平均水平。从样本结构来看,1951—60年和1976—85年出生组分别占全部样本的27%和25%左右,1961—75年出生组所占样本比例约为47%,对唐山而言,两个年龄组样本占全部样本的比例基本相同。唐山和其他地区的样本分别占全部样本的21.01%和78.99%。

### 三、唐山大地震对教育的影响

#### (一) 平均受教育年限随出生年变化的变动趋势

图2给出了各地区不同出生组平均受教育年限的变化情况。这种变化与发生在我国的重大灾害(“三年自然灾害”、唐山大地震)、重大政治运动(文化大革命)、重大教育政策(农村基础教育普及政策、高等教育扩招政策)和法律(《义务教育法》)的出台与实施关系密切。从图2可看出,1951—60年出生组的平均受教育年限均随出生年份的变化呈现大幅度上升的趋势,这主要是受到了农村基础教育普及政策<sup>①</sup>的影响。而1976—85年出生组的平均受教育年限随出生年份的变化平稳地保持在9年左右。这部分个体,特别是1980—85年出生组,在一定程度上受到了高等教育扩招政策<sup>②</sup>的影响。总之,对1951—

<sup>①</sup> 农村基础教育普及政策最初始于1967年。1971年4月的《全国教育工作会议纪要》(以下简称《纪要》),正式提出了基础教育普及政策,在教师、办学主体、教育经费及学制等方面给出了相应的规定。《纪要》明确指出:“争取在第四个五年计划期间,农村普及小学五年教育,有条件的地区,普及七年教育。”1974年5月,基础教育进一步偏向中学,国务院科教组(1975年被撤销)发出《关于1974年教育事业计划(草案)的通知》,要求逐步在大中城市普及十年教育,农村有条件的地区普及七年教育。直到1978年,国家教育委员会认为《纪要》不切实际,并对农村中小学进行了整顿和调整,大批农村中小学被取消,宣告了基础教育普及政策的结束。Sicular和Yang(2015)发现,对于1948—65年出生的个体,受到农村基础教育普及政策影响的个体的受教育年限比未受到政策影响的个体高0.4年以上。由于该政策颁布并实施于文化大革命期间,和城镇地区不同,文化大革命在一定程度上促进了农村地区基础教育的发展,而非阻碍(Meng and Gregory, 2002, 2007)。

<sup>②</sup> 高等教育扩招政策始于1999年。教育部出台《面向21世纪教育振兴行动计划》,提出到2010年,高等教育规模有较大扩展,入学率接近15%。这一政策极大地提高了高等教育招生规模和毛入学率。各年度《全国教育事业统计公报》(教育部规划司)显示:1999年本科生数和研究生数比1998年分别增长47.4%和21.4%,高等教育毛入学率从1998年的9.76%大幅度增长至2007年的23%,高等教育从精英教育阶段迅速进入大众化阶段。但相关研究表明,高等教育扩招政策是一项更加倾向于城镇地区的政策(邢春冰和李实,2011)。故在图2中,我们并没有发现1976—85年出生的个体的平均受教育年限显著增长。

60年和1976—85年出生组而言,唐山地区和除唐山外其他地区的平均受教育年限随出生年份变化的变动趋势基本相同。

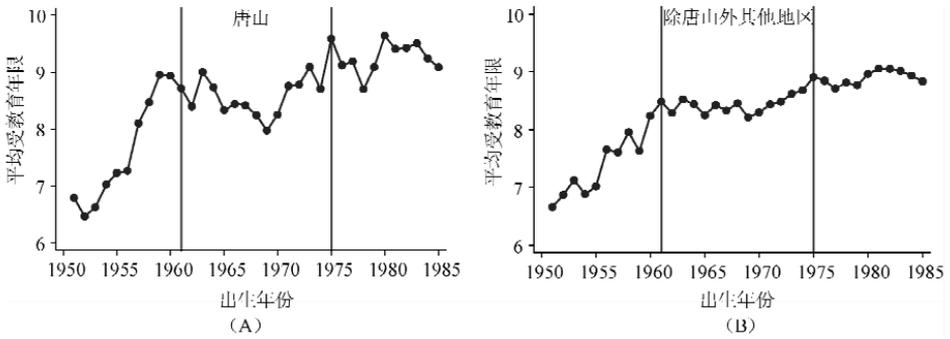


图2 受教育年限随个体出生年份的变化

图2还展示了1961—75年出生组平均受教育年限随出生年份变化的变动趋势。如图2(B)所示,对除唐山外的其他地区而言,各出生组的平均受教育年限大致相同。这主要是农村基础教育普及政策宣告结束后,基础教育学校数和学生数都大幅削减,在一定程度上阻碍了农村教育的发展。但对于唐山而言,20世纪60年代末出生的个体的平均受教育年限比20世纪60年代初出生的个体的差距超过一年。这种平均受教育年限随出生年份的变化迅速下降的趋势可能在很大程度上源于唐山大地震。需要指出的是,对1970年以后出生的个体而言,全部地区均呈现出平均受教育年限随出生年份变化而上升的趋势。这主要是受到1986年《义务教育法》颁布并在河北省实施<sup>①</sup>的影响。因此,唐山1961—75年出生组的受教育水平,不仅受到唐山大地震的负面影响,也受到《义务教育法》颁布和实施的正面影响,但对其他地区而言,受教育水平没有受到唐山大地震的负面影响。

为更加形象地反映实际受教育年限随出生年份变化的变动情况,解释上述灾害、政策和法律等外生冲击对个体受教育水平的影响,本文参照Ichinoand和Winter-Ebmer(2004)的方法,通过对受教育年限关于年龄多项式进行回归,得到受教育年限的残差,这一过程被称为去趋势(detrending)。具体而言,本文分

<sup>①</sup> 1986年,我国颁布《义务教育法》并在各省份逐步实施。该法律明确规定:凡年满六周岁的儿童,不分性别、民族、种族,应当入学接受规定年限的义务教育,条件不具备的地区,可以推迟到七周岁入学。相关研究表明,《义务教育法》的颁布和实施显著影响了个体的教育获得,且对农村地区的影响更大(Fang et al., 2012)。

地区性别别受教育年限关于年龄的三次多项式<sup>①</sup>进行回归,从而得到受教育年限的残差。地区受教育年限的残差随出生年份变化的变动趋势见图3。一方面,我们发现,受到农村基础教育普及政策结束的影响,所有地区1961—75年出生组受教育年限的残差绝大部分小于0;另一方面,相较于其他地区,唐山受教育年限残差的绝对值更大,说明唐山的某些外生冲击确实降低了个体的受教育水平。

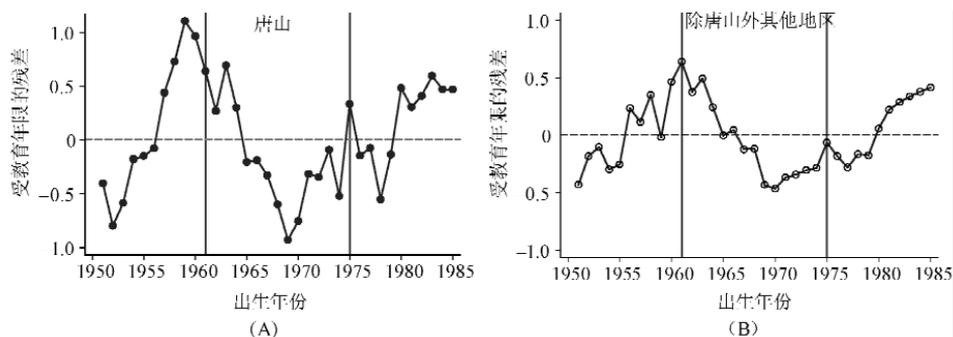


图3 受教育年限的残差随出生年份的变化

更进一步,本文利用回归方程计算了各个地区不同出生组受教育年限的差异。回归方程的被解释变量均为受教育年限的残差,解释变量包括1951—60年出生组和1976—85年出生组的虚拟变量,基准组为1961—75年出生组,同时控制性别,但不包括常数项,回归结果如表1所示。可以看出,唐山1961—75年出生组的受教育年限比1951—60年出生组低0.18年(10%的水平上统计显著),比1976—85年出生组低0.25年(1%的水平上统计显著)。邯郸、保定和承德三个地区1961—75年出生组的受教育水平比1951—60年出生组低0.05年(统计不显著),比1976—85年出生组低0.15年(1%的水平上统计显著)。

## (二) 双重差分模型及估计结果

虽然唐山1961—75年出生组的受教育水平均显著低于前后两个年龄组,但由于受到若干外生冲击的影响,表1估计出的这一差距并非完全为唐山大地震所致。因此,只有排除唐山大地震之外的外生冲击,才能准确识别出唐山

<sup>①</sup> 当工具变量与出生年份(或年龄)相关时,Bound and Jaeger(1996)强调了剔除年龄趋势对收入研究的重要性。年龄多项式次数的选取原则是:随着年龄多项式次数的增加,各估计系数均具有统计显著性,调整的 $R^2$ 上升。在本文中,当趋势用年龄的三次多项式刻画时,各项估计系数均在1%的水平上统计显著,且调整的 $R^2$ (超过0.9)最大,故本文采取年龄的三次多项式剔除受教育年限的收入年龄趋势。

表1 不同出生组的受教育年限差异

基准组:	(1)	(2)	(3)	(4)	(5)
出生于1961—75年	唐山	邯郸	保定	承德	除唐山外
出生于1951—60年	0.1759*	0.0082	0.0652	-0.0835	0.0529
	(0.0954)	(0.0821)	(0.0750)	(0.1558)	(0.0571)
出生于1976—85年	0.2477***	0.1089**	0.1038**	0.2161*	0.1479***
	(0.0771)	(0.0521)	(0.0525)	(0.1117)	(0.0425)
样本量	2891	4236	5147	1485	11985

注:上述模型的被解释变量均为受教育年限的残差,受教育年限的残差由分地区分性别的受教育年限关于年龄三次多项式的回归得到;上述模型均控制性别,但不包括常数项;第(5)列还控制地区虚拟变量;括号内为异方差稳健标准误;\*、\*\*和\*\*\*分别表示在1%、5%和10%的显著性水平上统计显著。

大地震对个体受教育水平的影响。双重差分模型(difference-in-differences,简称DiD)可以帮助我们识别这一影响<sup>①</sup>。具体而言,在DiD模型的研究设计(如表2)中,处理组是唐山,控制组是邯郸、保定和承德,实验组是1961—75年出生组,非实验组是1951—60年出生组和1976—85年出生组。若回归方程为

$$edu = \alpha_0 + \alpha_1 TS + \alpha_2 dummy61_75 + \alpha_3 TS \cdot dummy61_75 + Z\gamma + \mu \quad (1)$$

其中, $edu$ 表示受教育年限的残差, $TS$ 是一个二值虚拟变量,唐山为1,否则为0, $dummy61_75$ 也是一个二值虚拟变量,出生于1961—75年为1,否则为0, $TS \cdot dummy61_75$ 表示 $dummy61_75$ 与 $TS$ 的交互项, $Z$ 为其他控制变量,包括性别和地区虚拟变量, $\mu$ 为随机误差项。那么, $\alpha_3$ 识别的就是唐山大地震与个体受教育水平的因果关系。

方程(1)的估计结果见表3。第(1)—(3)列分别报告了将邯郸、保定和承德作为控制组的估计结果,第(4)列报告了同时将这三个地区作为控制组的估计结果。可以看出,在剔除其他外生冲击后,唐山大地震确实给唐山1961—75年出生组的受教育水平带来了负面影响,使其平均受教育年限显著减少0.16—0.26年。

<sup>①</sup> Chen and Zhou(2007)、Meng and Qian(2009)、Xu(2011)和Caruso et al.(2015)都采用DiD模型研究自然灾害对人力资本的影响。

表2 双重差分模型

	实验组(1)	非实验组(2)	(1)-(2)
处理组 ①	$E(edu dummy61_75 = 1, TS = 1)$ $= \alpha_0 + \alpha_1 + \alpha_2 + \alpha_3$	$E(edu dummy61_75 = 0, TS = 1)$ $= \alpha_0 + \alpha_1$	$\alpha_2 + \alpha_3$
控制组 ②	$E(edu dummy61_75 = 1, TS = 0)$ $= \alpha_0 + \alpha_2$	$E(edu dummy61_75 = 0, TS = 0)$ $= \alpha_0$	$\alpha_2$
①-②	$\alpha_1 + \alpha_3$	$\alpha_1$	$\alpha_3$

表3 唐山大地震对教育的影响:DiD模型估计结果

	(1)	(2)	(3)	(4)
	唐山和邯郸	唐山和保定	唐山和承德	全部地区
出生于1961—75年	-0.0926 (0.0646)	-0.1241** (0.0595)	-0.0330 (0.1292)	-0.0993** (0.0417)
唐山×出生于1961—75年	-0.1982** (0.1004)	-0.1666* (0.0971)	-0.2578* (0.1503)	-0.1914** (0.0874)
样本量	7127	8038	4376	13759

注:上述模型的被解释变量均为受教育年限的残差,受教育年限的残差由分地区分性别的受教育年限关于年龄三次多项式的回归得到;上述模型均控制性别,且包括常数项,模型(4)还包括地区虚拟变量;括号内为异方差稳健标准误;\*、\*\*和\*\*\*分别表示在1%、5%和10%的显著性水平上统计显著。

### (三) 稳健性分析

由于双重差分模型的估计结果对接下来的分析至关重要,故本文实施一系列检验以保证估计结果的稳定性。首先,双重差分模型应满足平行趋势假定(common trend assumption),即唐山和其他地区的平均受教育年限在未接受处理时随出生年的变化率应一致。本文采用通常的做法,以受教育年限的残差为因变量,以非实验组各出生年(1951—60年和1976—85年)的虚拟变量、处理组虚拟变量(唐山)以及二者的交互项为自变量进行回归,如果交互项的估计系数不具有统计显著性,则说明模型满足平行趋势假定。估计结果如附表4所示,可以看出,交互项的估计系数除极个别出生年外均统计不显著。但对交互项的联合显著性检验结果表明,前非实验组和后非实验组的F值分别为2.71和0.95,前者在1%的水平上统计显著,后者则在10%的显著性水平上统计不显著<sup>①</sup>。这说明,对前非实验组而言,处理组和控制组的变动趋势存在显著性

① 事实上,如果F检验不包括前非实验组的1959年和1960年虚拟变量与唐山的交互项,前非实验组的F值为1.38,全部非实验组的F值为1.13,均在10%的显著性水平上统计不显著,即满足平行趋势假定。

差异。图4较为直观地报告了按地区分样本回归后出生年虚拟变量的估计系数。我们发现,除1959和1960年外,各地区出生年虚拟变量的估计系数差异不大。故有理由怀疑,发生于1958—61年的三年自然灾害对人力资本的冲击存在地区差异,从而导致DiD模型设计在一定程度上违背平行趋势假定。因此,我们在模型中加入1958—61年出生组虚拟变量及其与唐山虚拟变量的交互项。估计结果表明,剔除三年自然灾害对人力资本的异质性影响后,1961—75年出生组虚拟变量与唐山虚拟变量交互项的估计系数下降至 $-0.1620$ ,但依然在10%的水平上统计显著。

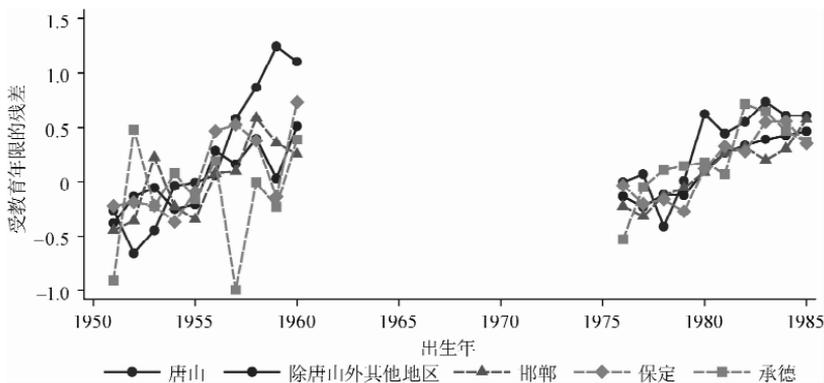


图4 按地区分样本回归出生年虚拟变量的估计系数

其次,为证实唐山1961—75年出生组受教育水平的下降确实由唐山大地震所致,而非同时期其他冲击的影响,本文参考Moser and Voena(2012)使用的证伪检验(falsification test)方法,选择了20世纪70—80年社会经济发展水平与河北省相近的辽宁省、湖南省、湖北省和安徽省,并按照同样的标准在各个省份中选取四个地级市作为本文四个地区的参照组<sup>①</sup>。模型设定仍然为式(1),但将被解释变量替换为参照组城市的受教育年限的残差,重新进行估计。估计结果如附表5所示。可以看出,所有模型交互项的估计系数均在10%的水平上统计不显著。更进一步,本文使用三重差分模型(difference-in-difference-in-differences,简称DiDiD)控制不可观测的因素对被解释变量的影响,即在DiD模型的基础上引入省份虚拟变量及其与1961—75年出生组变量、唐山虚拟变量的三重交互项,估计结果如表6所示。可以看出,所有模型三重交互项的估计系数均统计显著。以上两个检验结果均表明,在排除不可观测的影响因素

<sup>①</sup> 各省份选择的的地级市(按先后顺序分别对应河北省的唐山、邯郸、保定和承德)如下:辽宁省的大连、营口、阜新和朝阳,湖北省的宜昌、黄石、襄樊和孝感,湖南省的湘潭、株洲、衡阳和邵阳,安徽省的芜湖、宣城、安庆和黄山。

后,唐山大地震确实显著降低了个体的受教育水平。更重要的是,该结论在模型设定违背平行趋势假定的情况下依然成立。

第三,本文进行两种安慰剂检验(placebo test)。一方面,通过虚拟唐山大地震的发生时点,构造与样本类似的反事实验证因果关系。本文分别将唐山大地震的发生时间提前2年和错后2年,即假定唐山大地震发生于1974年和1978年,其他模型设定不变。附表7的估计结果表明,所有模型交互项的估计系数均不显著。另一方面,随着地震影响期限的延长,唐山大地震对个体受教育水平的影响应当越来越小直至消失。故本文将地震的影响期限分别设定为1976—78年、1976—80年和1976—84年(基准模型为1976—82年),估计结果如附表8所示。可以看出,随着影响期限的延长,交互项的估计系数的绝对值减小,由在1%的水平上统计显著变为统计不显著。以上两个检验充分说明,唐山大地震对个体受教育水平的影响确实存在。

最后,我们考察唐山大地震对个体受教育水平影响的动态异质性(dynamic heterogeneity)。由于实验组出生年的跨度长达15年,故不同出生年份个体的受教育水平遭受的影响可能存在差异。因此,本文将式(1)中的dummy61\_75替换为各个出生年的虚拟变量,并分别与唐山虚拟变量交互。估计结果如图5所示,其中圆点表示估计系数值,直线表示置信水平为90%时的置信区间。可以看出,唐山大地震的影响对不同出生组而言确实存在动态异质性。

地震对1965—70年出生组的影响最为显著,这可能是因为,该出生组的小学和初中教育阶段均受到了唐山大地震的影响,累积的教育损失更大。但同时也可以看出,地震对1961—64年和1971—75年出生组的影响比较微弱。

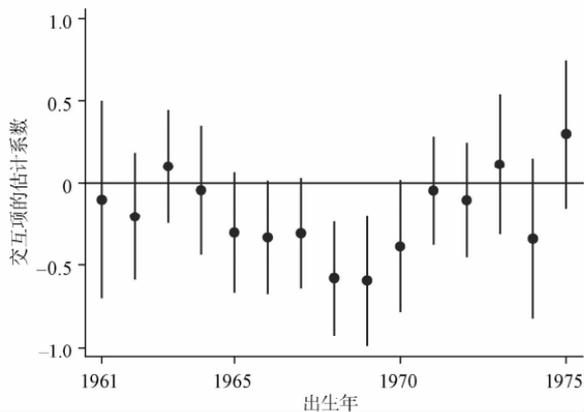


图5 唐山大地震对受教育水平影响的动态异质性

## 四、唐山大地震对收入的影响

### (一) 平均收入随出生年变化的变动趋势

图6展示了各地区不同出生组平均收入(取对数)的变化情况。唐山的平均收入随出生年份变化呈现波动上升趋势。但对除唐山外其他地区而言,平均

收入随出生年份的变化呈现倒U型关系,表明1961—75年出生组的平均收入略高于前后两个非实验组,这与标准的“年龄—收入”曲线一致。

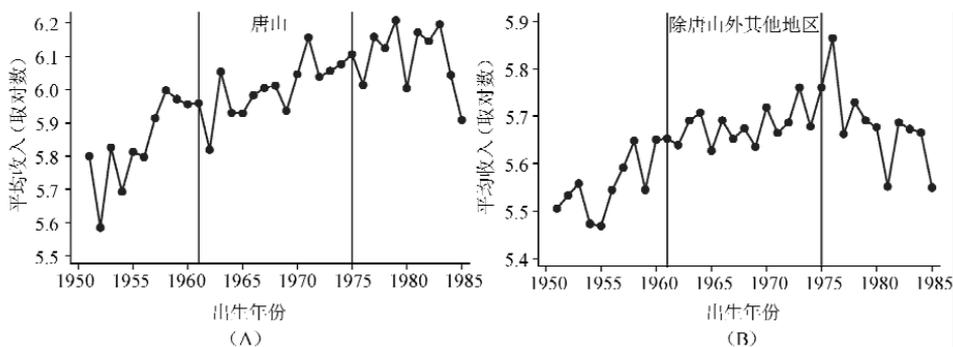


图6 收入随个体出生年份的变化

鉴于年龄与收入的关系,对个体收入剔除年龄趋势显得格外重要。这里,本文延续给受教育年限剔除年龄趋势的方法。具体而言,我们分地区分性别对收入关于年龄的三次多项式进行回归,从而得到收入的残差。各地区收入的残差随出生年份变化的变动趋势见图7。可以看出,唐山1961—75年出生组的收入的残差绝大部分小于0。

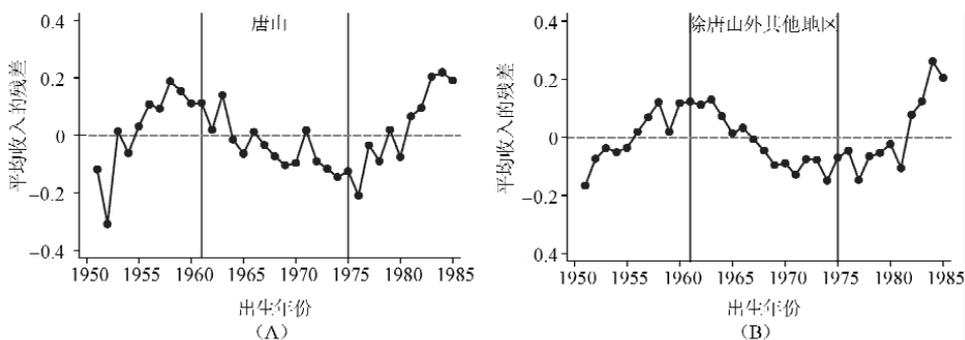


图7 收入的残差随个体出生年份的变化

和表1的做法类似,本文利用回归方程计算了各个地区不同出生组收入的差异。回归方程的被解释变量均为收入的残差,解释变量包括出生于1951—60年、出生于1976—85年两个虚拟变量,基准组为1961—75年出生的个体,同时控制性别,但不包括常数项,回归结果见表4。唐山1961—75年出生组的平均收入比1951—60年出生的个体低4.04% (统计不显著),比1976—85年出生的个体低5.16% (10%的水平上统计显著)。对除唐山外其他地区而言,1961—75年出生组的平均收入比1951—60年出生组低1.67% (统计不显著),比1976—85年出生组低3.81% (5%的水平上统计显著)。

表4 不同出生组的收入差异

基准组:	(1)	(2)	(3)	(4)	(5)
出生于 1961 — 75 年	唐山	邯郸	保定	承德	除唐山外
出生于 1951 — 60 年	0.0404 (0.0285)	0.0156 (0.0213)	0.0124 (0.0219)	-0.0206 (0.0391)	0.0167 (0.0167)
出生于 1976 — 85 年	0.0516* (0.0311)	0.0204 (0.0208)	0.0410* (0.0225)	0.0190 (0.0501)	0.0381** (0.0170)
样本量	2891	4236	5147	1485	10868

注:上述模型的被解释变量均为对数化月收入的残差,对数化月收入的残差由分地区分性别的月收入关于年龄三次多项式的回归得到;上述模型均控制性别,但不包括常数项;模型(5)还控制地区虚拟变量;括号内为异方差稳健标准误;\*、\*\*和\*\*\*分别表示在1%、5%和10%的显著性水平上统计显著。

## (二)教育对收入的影响

正如图3、图7、表1和表4所示,唐山1961—75年出生组的平均受教育年限和收入均低于其他两个出生组。那么,受教育年限的减少和收入的损失是否具有因果关系?工具变量(instrumental variable,简称IV)下的局部平均处理效应(local average treatment effect,简称LATE)(Imbens and Angrist,1994)能够帮助我们识别这一关系<sup>①</sup>。在一系列假设(后文详细讨论,此处省略)成立的前提下,将1961—75年出生组虚拟变量作为个体受教育年限的工具变量,由于这一变量与误差项中不可观测的因素(如能力)无关,且与个体的受教育年限有关(见表1),故这一方法能够估计教育与收入的因果关系。同时,由于教育回报的异质性,IV估计量是对局部平均处理效应的估计,即1961—75年出生组的教育回报率。具体而言,结构方程为

$$\ln income = \beta_0 + \beta_1 edu + M\lambda + \varepsilon \quad (2)$$

其中, $\ln income$ 表示取对数的月收入的残差, $edu$ 表示受教育年限的残差, $M$ 为其他控制变量,包括性别。由于1961—75年出生组虚拟变量为工具变量,故第一阶段回归为

$$edu = \rho_0 + \rho_1 dummy_{61-75} + M\omega + \nu \quad (3)$$

估计结果如表5第(1)列所示。OLS方法估计的受教育年限的估计系数为0.0620,且在1%的水平上统计显著。对LATE-IV而言,用广义矩估计

<sup>①</sup> 近期的许多研究应用LATE-IV方法识别教育与收入的因果关系,进而估算外生冲击的教育收益与成本,比如Ichino and Winter-Ebmer(2004)、Meng and Gregory(2007)、Fang et al.(2012)和Sicular and Yang(2015)。

(generalized method of moments,简称 GMM)方法估计的 1961—75 年出生组的教育回报率为 21.80% (在 5% 的水平上统计显著),是 OLS 估计值的 3.5 倍。内生性检验 C 统计量的  $\text{Chi}^2$  值为 2.74,说明受教育年限确实具有内生性。弱工具变量检验 Cragg—Donald Wald 统计量的 F 值为 14.32 (大于 10),说明 1961—75 年出生组的虚拟变量不是一个弱工具变量 (Staiger and Stock, 1997)。

表 5 教育回报率的估计结果:基准回归

样本	(1)	(2)	(3)	(4)	(5)
	唐山	唐山和邯郸	唐山和保定	唐山和承德	全部地区
OLS					
受教育年限	0.0620*** (0.0065)	0.0469*** (0.0039)	0.0555*** (0.0039)	0.0671*** (0.0052)	0.0527*** (0.0029)
LATE—IV(2SLS)					
受教育年限	0.2180** (0.1030)	0.2273** (0.0966)	0.2392*** (0.0923)	0.2148** (0.1024)	0.2349*** (0.0876)
LATE—IV(LIML)					
受教育年限		0.2280** (0.0971)	0.2408*** (0.0933)	0.2165** (0.1037)	0.2357*** (0.0881)
LATE—IV(GMM)					
受教育年限	0.2180** (0.1030)	0.2280** (0.0966)	0.2383*** (0.0922)	0.2150** (0.1024)	0.2349*** (0.0876)
第一阶段					
出生于 1961—75 年	-0.2910*** (0.0769)	-0.0926 (0.0646)	-0.1241** (0.0595)	-0.0330 (0.1292)	-0.0993** (0.0417)
唐山 × 出生于 1961—75 年		-0.1982** (0.1004)	-0.1666* (0.0971)	-0.2578* (0.1503)	-0.1914** (0.0874)
内生性检验	2.74*	4.39**	4.95**	2.47	5.47**
弱工具变量检验	14.32***	8.20***	9.34***	7.19***	10.01***
样本量	2891	7127	8038	4376	14876

注:上述模型的被解释变量和解释变量分别为对数化月收入的残差和受教育年限的残差,二者均由分地区分性别的月收入关于年龄三次多项式的回归得到;上述模型均控制性别,且包括常数项,模型(5)还控制地区虚拟变量;对于唐山样本,工具变量为 1961—75 年出生组虚拟变量,对于全部地区样本,工具变量为 1961—75 年出生组虚拟变量及其与唐山虚拟变量的交互项;内生性检验报告的是 C 统计量( $\text{Chi}^2$  值),弱工具变量检验报告的是 Cragg—Donald Wald 统计量(F 值);括号内为异方差稳健标准误;\*、\*\*和\*\*\* 分别表示在 1%、5% 和 10% 的显著性水平上统计显著。

但表5第(1)列的LATE-IV估计的教育回报率是否是由于唐山大地震而遭受教育损失的群体的教育回报率?换言之,该结果是否能够直接作为解释唐山大地震长期教育损失的依据?答案是否定的。按照LATE的假设(Angrist et al., 1996),这个教育回报率仅为1961—75年出生组的教育回报率,并非是受唐山大地震影响的群体的教育回报率,除非1961—75年出生组的收入损失仅受由唐山大地震所带来的受教育年限减少的影响。但正如前文所述,文化大革命期间农村基础教育普及政策的结束以及《义务教育法》的颁布与实施都与个体受教育水平密切相关,这些因素造成的个体受教育年限的变化所引起的收入的变化被混淆在了表5第(1)列的LATE-IV估计中。故我们有必要进行非混淆检验(unconfoundedness test),以剔除这些因素对估计的干扰。

因此,我们将唐山与其他地区的数据合并。除唐山外的其他地区并没有受到唐山大地震的影响,但均受到了文化大革命期间农村基础教育普及政策结束和《义务教育法》的颁布与实施等干扰因素的影响。通过增加其他地区的样本,对照组的质量得以提升,因为此时的对照组不仅包括了同一地区不同出生组的个体,也包括了不同地区同一出生组的个体。进一步,我们在回归模型中控制地区的虚拟变量,并将1961—75年出生组的虚拟变量及其与唐山的交互项作为受教育年限的工具变量。具体而言,结构方程为

$$\ln income = \delta_0 + \delta_1 edu + X\eta + \xi \quad (4)$$

其中, $\ln income$ 表示取对数的月收入的残差, $edu$ 表示受教育年限的残差, $X$ 为其他控制变量,包括性别和地区的虚拟变量。由于1961—75年出生组虚拟变量及其与唐山虚拟变量的交互项为工具变量,故第一阶段回归为

$$edu = \phi_0 + \phi_1 dummy61\_75 + \phi_2 TS \cdot dummy61\_75 + X\tau + \varsigma \quad (5)$$

这种设计的核心就是在第一阶段应用双重差分模型,即通过剔除地区间共同的干扰因素,识别出唐山大地震对1961—75年出生组受教育年限的影响。因此,LATE-IV估计的 $\Delta$ 识别的就是1961—75年出生组由于唐山大地震而遭受教育损失的一代的教育回报率。回归结果如表5所示。可以看出,LATE-IV估计的教育回报率约为23.49%,且在1%的水平上统计显著,同时模型通过了内生性检验和弱工具变量检验。如果将唐山和其他三个地区分别组合成为三个独立样本,用LATE-IV方法估计的三个样本的教育回报率约为21.50%—23.83%。但弱工具变量检验Cragg-Donald Wald统计量的F值均处于5—10之间,故存在弱工具变量的嫌疑,故表5同时报告了对弱工具变量不太敏感度的有限信息最大似然法(limited information maximum likelihood,简

称LIML)的估计结果。可以看出,GMM和LIML的估计结果基本相同。<sup>①</sup>

由于唐山大地震对个体受教育水平的影响对不同出生组而言存在动态异质性,因此本文还将1961—75年各个出生年的虚拟变量(针对唐山样本)以及这些变量与唐山虚拟变量的交互项(针对全部样本)作为工具变量组,对1961—75年出生组的教育回报率进行估计。估计结果如附表9所示。可以看出,使用工具变量组估计的教育回报率约为20.77%—22.31%,与表5的估计结果基本一致。但由于弱工具变量检验Cragg-Donald Wald统计量的F值较小,故难以排除弱工具变量的怀疑。

### (三) 稳健性分析

为保证教育回报率估计的稳健性,本文从模型设定、样本选择、变量选取和样本框设计等诸多方面对估计结果进行稳健性分析<sup>②</sup>。

#### 1. 是否满足LATE定理条件

LATE定理共包括独立性条件(independence condition)、排他性条件(exclusion condition)、第一阶段条件(first stage condition)和单调性条件(monotonicity condition)(Angrist and Pischke,2008)。独立性条件要求,工具变量既与潜在结果有关,也与潜在处理状态无关。由于唐山大地震是不可预测的巨大自然灾害,是一个纯粹外生的负向冲击,故满足独立性假设显而易见。排除性条件要求,工具变量只能通过唯一的已知途径对被解释变量产生影响,即唐山大地震只能通过影响个体受教育水平对收入造成影响。具体而言,对于受教育水平受到唐山大地震影响的个体,受教育年限的减少应当是影响收入的唯一途径。当然,地震可能会造成心理扭曲和营养不良,从而影响个体在未来劳动力市场的表现。在本研究的样本中,我们观测到的是个体在2005年的劳动收入,唐山大地震等外生冲击不太可能直接影响个体在震后30年的收入水平。第一阶段假设条件,工具变量必须显著影响处理状态,即唐山大地震必须显著影响个体受教育水平,这在表5的第一阶段回归中已经得到验证。单调性条件要求,对于那些受工具变量影响的个体,这种影响必须以相同的方式发生,即不存在某些人因唐山大地震发生受教育年限增加的情况。在引言中,本文已经介绍了唐山大地震的灾难性影响,故单调性假设不证自明。因此,本研究设计满足

<sup>①</sup> 为便于对不同估计方法下估计结果的比较,表6还报告了用两阶段最小二乘法(two-stage least squared,简称2SLS)的结果,与GMM估计结果基本相同。

<sup>②</sup> 受限于文章篇幅,部分变量的估计结果并未在正文和附表中报告,如有需要可向作者索取。

LATE 定理条件,表5的估计结果识别的确实为局部平均处理效应。

## 2. 排除一般均衡效应

所谓一般均衡效应,是指在不同出生年份的劳动者不能完全替代的情况下,受到供给效应的影响,具有较小人口规模和较低受教育水平的一代人,可能获得更高的教育回报(Berger,1985;Card and Lemieux,2001)。为排除一般均衡效应的影响,我们在模型中加入了受地震影响一代的人口规模、受地震影响一代的平均受教育年限和未受地震影响一代的平均受教育年限,这三个指标均基于2005年中国1%抽样调查数据的一个20%的随机样本,在按照性别(2类)、行业(20类)和城市(4类)的人口分组中计算得到。估计结果如附表10所示。受地震影响一代的人口规模的估计系数显著为负,说明人口规模对收入有抑制作用。受地震影响一代的平均受教育年限和未受地震影响一代的平均受教育年限的估计系数为正,说明个体所在群体的平均受教育水平对收入有促进作用(未报告)。使用唐山样本估计的受唐山大地震影响一代的教育回报率为20.16% - 21.75%,使用全部样本估计的教育回报率为24.41% - 27.85%,这与表5的估计结果基本一致。

## 3. 纠正样本选择偏差

样本选择(sample selection)问题在本研究中主要体现在三个方面。首先,自然灾害与失业率、劳动供给规模、劳动供给时间和工作决策等密切相关<sup>①</sup>(Meng and Qian,2009;Xu,2011),而个体是否进入劳动力市场的决策又会影响教育回报率的估计。本文将有收入个体视为进入劳动力市场,将收入为0的个体视为未进入劳动力市场。故在本文使用的样本中,未进入劳动力市场的样本量为2017,占全部样本的11.94%。我们根据Wooldridge(2010)提供的方法<sup>②</sup>,采用含有内生解释变量的Heckman两阶段模型纠正这种样本选择偏误。估计结果如附表11第(1)列所示<sup>③</sup>。逆米尔斯比的估计系数统计显著,说明模型确实存在样本选择偏差(未报告)。IV估计结果表明,纠正样本选择偏差后,使用

① 比如,Meng和Qian(2009)发现,受三年自然灾害影响的个体在成年后的劳动供给更低;Di Pietro和Mora(2011)则发现拉奎拉地震显著降低了劳动参与率,但并未影响就业率;Xu(2011)也没有发现受唐山大地震影响的个体在成年后的失业率更高。

② 这种方法已经在多项研究中使用过,如Arabsheibani和Mussurov(2007)、Chen和Hamori(2009)等。

③ 使用Heckman两阶段模型纠正样本选择偏差时,需要引入识别变量。本文的识别变量包括家庭总人口中0-5岁人口的比例、6-15岁人口的比例和60岁及以上人口的比例,分别衡量家庭抚养学龄前儿童、未成年学龄儿童和老人的负担(Zhang et al.,2005)。

唐山和全部地区样本估计的受地震影响一代的教育回报率分别为 9.88% 和 9.53%,且均在 1% 的水平上统计显著。

其次,使用明瑟收入方程估计教育回报率时,解释变量应为个体的工资性收入。但由于无固定职业和无固定工作场所的个体在农村普遍存在,农村样本的工资性收入难以被界定。相比于女性,男性个体的收入可能在更大程度上体现为工资性收入。使用男性样本的另一好处在于,男性个体通过婚嫁娶等方式从其他地区农村流动到唐山农村的概率比女性更低,故男性个体始终居住在本地农村的可能性就比女性更大。使用男性样本,就更能保证个体的确是在样本市的农村地区接受的义务教育。附表 11 第(2)列报告了对男性教育回报率的估计结果。可以看出,地震使男性个体的受教育年限降低了 0.19 年,受地震影响的男性一代的教育回报率约为 19.35%,这与表 5 基本一致。

最后,本文通过调查问卷提供的个体户口登记地的省、市、县信息,将在这四个地级市农村地区以农业户口身份流动到其他地区(包括省内市外和省外)的个体纳入到本研究的样本中,在一定程度上降低了样本选择偏误。如附表 11 第(3)列所示,样本调整前后的估计结果大致相同。然而遗憾的是,我们无法识别其他两类可能在唐山大地震影响期间居住在农村的个体。一是受城市化影响由农业户口变为非农户口的个体,这属于户口特征的自然变动;二是在本研究所涉及的四个地级市农村地区以农业户口身份接受义务教育但进入城镇劳动力市场并将户口变为非农户口的个体。这部分个体往往具有较高的受教育水平。由此带来的样本选择偏误,我们无法纠正。

#### 4. 其他稳健性分析

除上述稳健性分析,附表 12 还报告了其他稳健性检验的结果。模型(1)控制了少数民族虚拟变量、目前的居住地类别(城市、县镇和农村)、父母的受教育年限<sup>①</sup>等个体特征和后非实验组的虚拟变量。估计结果与表 5 一致。模型(2)和(3)分别将个体是否接受完初中阶段教育(包括毕业和肄业)、是否接受高中阶段教育作为受教育水平的衡量指标,考察唐山大地震如何通过影响个体的教育决策降低其受教育水平进而影响未来的收入水平。可以看出,地震显著降低了个体完成初中阶段教育和接受高中阶段教育的概率,但 LATE-IV 估

<sup>①</sup> 父母受教育水平缺失的样本占全部样本的近 85%,若仅对父母受教育水平不缺失的样本回归,则会损失大量样本并存在严重的同住性样本选择偏差,故本文将父母受教育水平缺失的个体的父母平均受教育年限赋值为 0,并在模型中控制父母受教育水平是否均缺失的虚拟变量。

计的教育回报率非常高,且估计结果也在一定程度上受到了弱工具变量的影响。模型(4)的实验组仅包括小学和初中教育阶段均受地震影响的个体,前后非实验组保持不变。估计结果表明,地震使得这部分个体的平均受教育年限减少0.37年,接近样本平均水平(0.19年)的2倍,LATE-IV估计的教育回报率为16.88%,略低于样本平均水平(23.49%)。

## 五、唐山大地震的教育损失

第三部分和第四部分分别考察了唐山大地震对个体受教育水平的影响和受唐山大地震影响的一代人的教育回报率。基于上述估计结果,本文通过建立成本核算框架计算唐山大地震的短期和长期教育损失。

唐山大地震的短期教育损失是指唐山大地震给个体受教育水平造成的损失。本文采用两种方式定义唐山大地震的短期教育损失。一是针对唐山样本计算1961—75出生组的平均受教育年限相对于其他出生组的差异,记为 $SCOST1$ ,等同于方程(3)中的估计系数 $\rho_1$ 。二是针对全部地区样本,剔除各地区共同的外生冲击对个体受教育水平的影响后,计算唐山1961—75年出生组的平均受教育年限相对于其他出生组的差异,记为 $SCOST2$ ,等同于方程(1)中的估计系数 $\alpha_3$ ,也等同于方程(5)中的估计系数。

唐山大地震的长期教育损失是指唐山大地震使得个体受教育水平减少进而对其在未来劳动力市场中的收入造成的损失。本文采用三种方式定义唐山大地震的长期教育损失。一是用LATE-IV估计得到的教育回报率衡量,其意义是指对受唐山大地震影响的一代而言,多接受一年正规教育收入增加的百分比(或者说少接受一年正规教育收入减少的百分比)。本文将使用唐山样本和全部地区样本估计的唐山大地震的长期教育损失分别记为 $SCOST1_A$ 和 $SCOST1_B$ ,二者分别等价于方程(2)中的 $\beta_1$ 和方程(4)中的 $\delta_1$ 的IV估计量。二是将唐山大地震的短期教育损失与受地震影响的一代人的教育回报率相结合,即在考虑唐山大地震对个体受教育年限的影响的前提下,进一步考察个体在未来劳动力市场中的收入损失。对于唐山样本和全部地区样本,唐山大地震的长期教育损失分别记为 $SCOST2_A$ 和 $SCOST2_B$ ,前者等同于估计系数 $\rho_1$ 与 $\beta_1$ 的乘积,后者等同于 $\phi_2$ 和 $\delta_1$ 的乘积。 $SCOST1$ 、 $SCOST2$ 、 $LCOST1$ 和 $LCOST2$ 的关系如式(6)和式(7)所示:

$$LCOST2_A = SCOST1 \cdot LCOST1_A \quad (6)$$

$$LCOST2_B = SCOST2 \cdot LCOST1_B \approx LCOST2_A \cdot \frac{SCOST2}{SCOST1} \quad (7) \textcircled{1}$$

三是参照 Ichino 和 Winter-Ebmer(2004)的方法,计算受唐山大地震影响的一代人的收入损失占有所有样本平均收入的百分比,这与唐山大地震造成的唐山 GDP 的损失十分接近。对于唐山样本和全部样本,唐山大地震的长期教育损失分别记为  $LCOST3_A$  和  $LCOST3_B$ ,即

$$LCOST3_A = \frac{E(\ln income | dummy61_75_i = 1, TS_i = 1) \cdot LCOST2_A \cdot \Pr(dummy61_75_i = 1 | TS_i = 1)}{E(\ln income | TS_i = 1)} \quad (8)$$

$$LCOST3_B = \frac{E(\ln income | dummy61_75_i = 1, TS_i = 1) \cdot LCOST2_B \cdot \Pr(dummy61_75_i = 1)}{E(\ln income)} \quad (9)$$

上述指标的计算结果如表6所示。唐山样本的计算结果表明,唐山大地震使得个体受教育年限减少 0.27 - 0.29 年,受地震影响一代的教育回报率为 20.16% - 21.80%,故唐山大地震使得个体受教育年限的减少进而带来的收入损失约为 5.56% - 6.32%,由唐山大地震带来的人力资本损失约占 2005 年唐山 GDP 的 0.51% - 0.58%。但仅使用唐山样本会高估唐山大地震的教育损失,因为 1961 - 75 年出生组的受教育水平既可能潜在受到唐山大地震的影响,也有可能受到其他外生冲击的影响。使用全部地区样本的计算结果表明,唐山大地震使得个体受教育年限减少 0.14 - 0.21 年,可将此看作唐山大地

① 本文在计算的过程中采用了近似处理。

$$LCOST2_B = SCOST2 \cdot SCOST2_B = \frac{E(\ln income_i | IV_i = 1) - E(\ln income_i | IV_i = 0)}{E(educ_i | IV_i = 1) - E(educ_i | IV_i = 0)} \cdot SCOST2$$

$$= \frac{E(\ln income_i | dummy61_75_i = 1, TS_i = 1) - E(\ln income_i | dummy61_75_i = 0, TS_i = 0)}{E(educ_i | dummy61_75_i = 1, TS_i = 1) - E(educ_i | dummy61_75_i = 0, TS_i = 0)} \cdot SCOST2$$

要想得到式(7),必须保证以下两个条件同时成立

$$E(educ_i | dummy61_75_i = 0, TS_i = 0) = E(educ_i | dummy61_75_i = 0, TS_i = 1)$$

$$E(\ln income_i | dummy61_75_i = 0, TS_i = 0) = E(\ln income_i | dummy61_75_i = 0, TS_i = 1)$$

即唐山地区和非唐山地区的前后非实验组的平均受教育年限和收入水平不应该具有显著性的差异。本文利用单因素方差对上述两个条件进行检验。巴特利特检验(Bartlett's test)结果表明单因素方差分析满足等方差假定。单因素方差检验结果表明,两个样本平均受教育年限和收入水平不存在显著性差异。由于,

$$SCOST1 = E(educ_i | dummy61_75_i = 1, TS_i = 1) - E(educ_i | dummy61_75_i = 0, TS_i = 1)$$

则

$$LCOST2_B = [E(\ln income_i | dummy61_75_i = 1) - E(\ln income_i | dummy61_75_i = 0)] \cdot$$

$$\frac{SCOST2}{SCOST1} = LCOST2_A \cdot \frac{SCOST2}{SCOST1} \circ$$

震的实际短期教育损失。受地震影响一代少接受一年正规教育,收入减少 20.93% - 27.85%。因此,唐山大地震使得个体受教育年限的减少进而带来的收入损失约为 3.51% - 4.77%,唐山大地震带来的人力资本损失约占 2005 年唐山 GDP 的 0.30% - 0.41%,可将此看作唐山大地震的实际长期教育损失。相比于全部地区样本的计算结果,唐山样本的计算结果高估了唐山大地震的教育损失。

表 6 唐山大地震的短期和长期教育损失

	<i>SCOST</i>	<i>LCOST1</i>	<i>LCOST2</i>	<i>LCOST3</i>
基准回归				
唐山样本	-0.29	21.80%	-6.34%	0.58%
全部样本	-0.19	23.49%	-4.50%	0.39%
稳健性检验				
唐山样本	[-0.29, -0.27]	[20.16%, 21.79%]	[-6.32%, -5.56%]	[0.51%, 0.57%]
全部样本	[-0.21, -0.14]	[20.93%, 27.85%]	[-4.77%, -3.51%]	[0.30%, 0.41%]

注:稳健性检验仅包括附表 10、附表 11 的模型(3)和附表 12 的模型(1)。

## 六、结论

诸如地震、海啸等突发性自然灾害不仅给社会经济发展造成了不可估量的损失,也可能对整个社会的人力资本和社会福利带来持久和深远的影响。本文以唐山大地震为例为该领域的研究提供了新的证据。本文使用 2005 年全国 1% 人口抽样调查中随机抽取的一个 20% 的样本,运用 DiD 模型和 LATE - IV 方法研究了唐山大地震对个体受教育水平和在未来劳动力市场中表现的影响。

本文主要得到如下结论:首先,双重差分模型的估计结果表明,剔除不同地区受到的共同外生冲击的影响后,唐山大地震使得个体受教育年限减少 0.14 - 0.21 年,可将此看作唐山大地震的短期教育损失。其次,本文估计了受唐山大地震影响一代的教育回报率。LATE - IV 估计结果表明,个体少接受一年正规教育,收入水平下降 20.93% - 27.85%。最后,利用唐山大地震的短期教育损失和估计得到的教育回报率,计算了唐山大地震的长期教育损失。唐山大地震使得个体受教育年限的减少进而带来的收入损失约为 3.51% - 4.77%,人力资本损失约占 2005 年唐山 GDP 的 0.30% - 0.41%,可将此看作唐山大地震的实际长期教育损失。鉴于此,当诸如自然灾害等剧烈的负向外生冲击发生时,我们不仅应关注和弥补它们给社会经济方面造成的巨大损失,同时更应积极进

行教育等人力资本方面的救助,尽可能减少负向外生冲击对个体人力资本水平的影响,并对人力资本积累过程受到严重冲击的居民进行持续性的补偿,从而为个体远期福利水平提供基本保障。

当然,本文还存在如下不足。首先,在唐山大地震的自然场景下应用LATE-IV方法是否满足排除性假定仍有待考察。虽然从理论上讲,对于受教育水平受到唐山大地震影响的个体,受教育年限的减少是影响收入的唯一途径,但确实也存在地震通过恶化个体早期营养健康状况影响其在未来劳动力市场的表现的可能性。其次,受限于数据,本文无法对唐山大地震对个体受教育水平的影响机制进行深入探讨。再次,本文将受唐山大地震影响的地区仅设定为唐山,而没有考虑其他诸如河北廊坊、北京和天津等受唐山大地震影响的地区,故本文估计的仅是唐山大地震对唐山农村居民人力资本和福利水平的影响,这可能会低估唐山大地震的教育损失。最后,由于缺少县级数据,本文的估计结果未能刻画不同震级地区居民唐山大地震的教育损失,也没能在更为宏观的区县层面探讨唐山大地震与人力资本和福利水平的关系。

### [参考文献]

- 吴要武、赵泉,2010:《高校扩招与大学毕业生就业》,《经济研究》第9期。
- 邢春冰、李实,2011:《扩招“大跃进”、教育机会与大学毕业生就业》,《经济学(季刊)》第4期。
- 颜敏,2012:《能力偏误、教育溢价与我国工资收入差异:基于微观计量方法的实证研究》,东北财经大学博士论文。
- 于山、苏幼坡、刘天适、刘瑞兴,2003:《唐山大地震震后救援与恢复重建》,北京:中国科学技术出版社。
- 宗树兴,2010:《1986年〈中华人民共和国义务教育法〉立法和实施研究》,河北大学博士论文。
- Angrist, J. D., Imbens, G. W. and Rubin, D. B., 1996, “Identification of causal effects using instrumental variables”, *Journal of the American Statistical Association*, 91(434): 444-455.
- Angrist, J. D. and Pischke, J. S., 2008, *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Alderman, H., Hoogeveen, H. and Rossi, M., 2009, “Preschool nutrition and subsequent schooling attainment: longitudinal evidence from Tanzania”, *Economic Development and Cultural Change*, 57(2): 239-260.
- Arabsheibani, G. R. and Mussurov, A., 2007, “Returns to schooling in Kazakhstan”, *Economics of Transition*, 15(2): 341-364.

- Baez, J. , de la Fuente, A. and Santos, I. , 2010, "Do natural disasters affect human capital? An assessment based on existing empirical evidence", IZA Discussion Paper, No. 5164, Institute for the Study of Labor.
- Baez, J. E. and Santos, I. V. , 2007, "Children's vulnerability to weather shocks: A natural disaster as a natural experiment", World Bank Research.
- Baez, J. E. and Santos, I. V. , 2008, "On shaky ground: The effects of earthquakes on household income and poverty", Unpublished manuscript.
- Berger, M. C. , 1985, "The effect of cohort size on earnings growth: A reexamination of the evidence", *Journal of Political Economy*, 93(3) : 561-573.
- Bhatia, K. G. , 2008, "Foundations for industrial machines and earthquake effects", *Journal of Earthquake Technology*, 45(1) : 13-29.
- Bound, J. and Jaeger, D. A. , 1996, "On the validity of season of birth as an instrument in wage equations: A comment on Angrist & Krueger's 'Does compulsory school attendance affect schooling and earnings?'" , NBER Working Paper, No. 5835.
- Bustelo, M. , Arends-Kuenning, M. and Lucchetti, L. , 2012, "Persistent impact of natural disasters on child nutrition and schooling: Evidence from the 1999 Colombian earthquake", IZA Discussion Paper, No. 6354, Institute for the Study of Labor.
- Card, D. and Lemieux, T. , 2001, "Can falling supply explain the rising return to college for younger men? A cohort-based analysis", *The Quarterly Journal of Economics*, 116(2) : 705-746.
- Caruso, G. , 2015, "The Legacy of Natural Disasters: The Intergenerational Impact of 100 Years of Natural Disasters in Latin America", Available at SSRN 2560891.
- Caruso, G. and Miller, S. , 2015, "Long run effects and intergenerational transmission of natural disasters: A case study on the 1970 Ancash Earthquake", *Journal of Development Economics*, 117(11) : 134-150.
- Chen, G. and Hamori, S. , 2009, "Economic returns to schooling in urban China: OLS and the instrumental variables approach", *China Economic Review*, 20(2) : 143-152.
- Chen, S. H. , Lin, Y. H. , Tseng, H. M. and Wu, Y. C. , 2002, "Posttraumatic stress reactions in children and adolescents one year after the 1999 Taiwan Chi-Chi earthquake", *Journal of the Chinese Institute of Engineers*, 25(5) : 597-608.
- Chen, Y. and Zhou, L. A. , 2007, "The long-term health and economic consequences of the 1959-1961 famine in China", *Journal of Health Economics*, 26(4) : 659-881.
- Di Pietro, G. and Mora, T. , 2015, "The effect of the L' Aquila earthquake on labor market outcomes", *Environment and Planning C: Government and Policy*, 33(2) : 239-255.
- Fang, H. , Eggleston, K. , Rizzo, J. A. , Rozelle, S. and Zeckhauser, R. J. , 2012, "The returns to education in china evidence from the 1986 Compulsory Education Law", NBER Work-

- ing Paper, No. 18189.
- Gignoux, J. and Menéndez, M. , 2014, “Benefit in the wake of disaster: Long-run effects of earthquakes on welfare in rural Indonesia”, PSE Working Papers n2014-27.
- Hoddinott, J. and Kinsey, B. , 2001, “Child growth in the time of drought”, *Oxford Bulletin of Economics and Statistics*, 63(4) : 409-436.
- Ichino, A. and Winter-Ebmer, R. , 2004, “The long run educational cost of World War II”, *Journal of Labor Economics*, 22(1) : 57-86.
- Imbens, G. W. and Angrist, J. D. , 1994, “Identification and estimation of local average treatment effects”, *Econometrica*, 62(2) :467-476.
- Maccini, S. and Yang, D. , 2009, “Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall”, *The American Economic Review*, 99(3) : 1006-1026.
- Meng, X. and Gregory, R. G. , 2002, “The Impact of Interrupted Education on Subsequent Educational Attainment: A Cost of the Chinese Cultural Revolution”, *Economic Development and Cultural Change*, 50(4) , 935-959.
- Meng, X. and Gregory, R. , 2007, “Exploring the impact of interrupted education on earnings the educational cost of the Chinese Cultural Revolution”, IZA Discussion Paper, No. 2548, Institute for the Study of Labor.
- Meng, X. and Qian, N. , 2009, “The long term consequences of famine on survivors: Evidence from a unique natural experiment using China’s great famine”. NBER Working Paper, No. 14917.
- Moser, P. and Voena, A. ,2010, “Compulsory Licensing: Evidence from the Trading with the Enemy Act”, *The American Economic Review*, 102(1) :396-427
- Neelsen, S. and Stratmann, T. , 2011, “Effects of prenatal and early life malnutrition: Evidence from the Greek famine”, *Journal of Health Economics*, 30(3) : 479-488.
- Roussos, A. , Goenjian, A. K. , Steinberg, A. M. , Sotiropoulou, C. , Kakaki, M. , Kabakos, C. , Karagianni, S. and Manouras, V. , 2005, “Posttraumatic stress and depressive reactions among children and adolescents after the 1999 earthquake in Ano Liosia”, Greece, *The American Journal of Psychiatry*, 162(3) : 530-537.
- Şalcioğlu, E. and Başoğlu, M. , 2008, “Psychological effects of earthquakes in children: prospects for brief behavioral treatment”, *World Journal of Pediatrics*, 4(3) : 165-172.
- Santos, I. , 2007, “Disentangling the effects of natural disasters on children: 2001 earthquakes in El Salvador”, Kennedy School of Government, Harvard University.
- Sicular, T. and Yang, J. , 2015, “The returns to schooling in rural china evidence from the Cultural Revolution education expansion”, CIBC Working Paper. Centre Imperial Bank of Commerce.

- Staiger, D., Stock, J. and Watson, M., 1997, "The NAIRU, unemployment and monetary Policy", *Journal of Economic Perspectives*, 11(1): 33-49.
- Tan, C. M., Tan, Z. and Zhang, X., 2014, "Sins of the fathers: The intergenerational legacy of the 1959-61 Great Chinese Famine on children's cognitive development", IFPRI Discussion Paper, No. 1351. International Food Policy Research Institute.
- Wooldridge, J. M., 2010, *Econometric Analysis of Cross Section and Panel Data*, MIT press.
- Xu, G., 2011, "Long-run consequences of natural disasters: Evidence from Tangshan", Discussion Paper, No. 1117, DIW Berlin, German Institute for Economic Research.
- Zhang, J., Liu, P. W. and Yung, L., 2007, "The Cultural Revolution and returns to schooling in China: Estimates based on twins", *Journal of Development Economics*, 84(2): 631-639.
- Zhang, J., Zhao, Y. H., Park, A. and Song, X., 2005, "Economic returns to schooling in urban China, 1988-2001", *Journal of Comparative Economics*, 33(4): 730-752.

## The Educational Cost of the Exogenous Shock: Evidence from Tangshan Earthquake in 1976

WANG Jun, LI Bo

(Business School/Capital Institute for Economics of Education, Beijing Normal University)

**Abstract:** Based on a random sample from 1 percent population survey data of 2005, this paper studies the impacts of Tangshan Earthquake on the educational attainment and subsequent labor market outcomes of affected cohorts using the methodology of difference-in-differences model and the local average treatment effect interpretation of instrumental variables technique. We find that the schooling years of the cohorts potentially affected by Tangshan Earthquake is 14% - 21% of a year of schooling lower than the non-earthquake cohorts, which can be considered as the short-term educational cost of Tangshan Earthquake. The rate of return to years of schooling for the earthquake cohorts is 20.93% - 27.85%. The earthquake leads 3.51% - 4.77% loss of average income through the reduction of schooling years. A loss of 0.30% - 0.41% of GDP in 2005 can be attributed to the lower educational attainment of the earthquake cohorts, which can be considered as the long-term educational cost of Tangshan Earthquake.

**Key words:** Tangshan Earthquake; human capital; difference-in-differences model; instrumental variable; local average treatment effect

## 附表

附表1 样本框

	出生年份	2005年时年龄	入学年份		被影响的年数	
			小学	初中	小学	初中
前非实验组	1951	54	1958	1964	0	0
	1952	53	1959	1965	0	0
	1953	52	1960	1966	0	0
	1954	51	1961	1967	0	0
	1955	50	1962	1968	0	0
	1956	49	1963	1969	0	0
	1957	48	1964	1970	0	0
	1958	47	1965	1971	0	0
	1959	46	1966	1972	0	0
	1960	45	1967	1973	0	0
实验组	1961	44	1968	1974	0	1
	1962	43	1969	1975	0	2
	1963	42	1970	1976	0	3
	1964	41	1971	1977	1	3
	1965	40	1972	1978	2	3
	1966	39	1973	1979	3	3
	1967	38	1974	1980	4	3
	1968	37	1975	1981	5	2
	1969	36	1976	1982	6	1
	1970	35	1977	1983	6	0
	1971	34	1978	1984	5	0
	1972	33	1979	1985	4	0
	1973	32	1980	1986	3	0
	1974	31	1981	1987	2	0
	1975	30	1982	1988	1	0

续表

	出生年份	2005年时年龄	入学年份		被影响的年数	
			小学	初中	小学	初中
后非实验组	1976	29	1983	1989	0	0
	1977	28	1984	1990	0	0
	1978	27	1985	1991	0	0
	1979	26	1986	1992	0	0
	1980	25	1987	1993	0	0
	1981	24	1988	1994	0	0
	1982	23	1989	1995	0	0
	1983	22	1990	1996	0	0
	1984	21	1991	1997	0	0
	1985	20	1992	1998	0	0

注:假设唐山大地震影响个体受教育水平的年份为1978—82年;假定学生7岁上小学,13岁上初中。

附表2 描述统计

	唐山	邯郸	保定	承德	除唐山外	全部
男性	54.00%	52.20%	53.74%	57.37%	53.63%	53.71%
出生年份	1967.01	1968.54	1967.68	1966.19	1967.81	1967.64
年龄	37.83	36.34	37.18	38.66	37.05	37.22
少数民族	2.63%	0.38%	1.48%	40.61%	6.40%	5.60%
受教育水平	8.46	8.30	8.36	7.95	8.28	8.31
收入	558.60	335.68	440.10	369.70	389.78	425.25
父母的平均受教育年限	6.22	5.96	6.11	4.99	5.93	5.98
父母的受教育水平均缺失	82.39%	75.05%	76.18%	81.55%	76.47%	77.72%
居住在城市	10.17%	14.02%	5.15%	0.27%	7.94%	8.41%
居住在县镇	15.77%	6.80%	14.88%	14.95%	11.74%	12.59%
居住在农村	74.06%	79.18%	79.97%	84.78%	80.32%	79.00%
出生于1951—60年	28.19%	25.57%	27.63%	30.57%	27.23%	27.43%
出生于1961—75年	49.60%	45.28%	46.01%	52.19%	46.57%	47.21%
出生于1976—85年	22.21%	29.15%	26.36%	17.24%	26.21%	25.37%
样本量	2891	4236	5147	1485	10868	13759
样本所占比重	21.01%	30.79%	37.41%	10.79%	78.99%	100.00%

附表3 1951—85年出生组可能经历的重大外生冲击

重大外生冲击	持续时间	受影响的 出生组	2005年 时年龄	资料来源
三年自然灾害	1959—61	1954—61	44—51	Mengand Qian(2009)
文化大革命	1966—76	1945—64	41—60	Dengand Treiman(1997)
		1947—63 <sup>a</sup>	42—58	Mengand Gregory(2002)
		1947—61 <sup>a</sup>	44—58	Mengand Gregory(2007)
		1950—69	45—54	Zhang et al.(2007)
农村基础教育普及政策	1971—78	1953—60	45—52	Sicularand Yang(2015)
义务教育法	1986—今	1970—71 <sup>b</sup>	32—35	Fanget al.(2012)
高等教育扩招	1999—2007	1981—83 <sup>c</sup>	22—24	吴要武和赵泉(2010)
		1980—84 <sup>c</sup>	21—25	邢春冰和李实(2011)
		1981—85 <sup>c</sup>	20—24	颜敏(2012)

注:a. 通过在模型中加入个体被中断的教育阶段的虚拟变量, Mengand Gregory(2002、2007)分析了文化大革命对个体受教育水平的影响。由于唐山大地震仅影响1976—82年间处于小学或初中阶段的个体,根据两篇文章的样本框设计,受唐山大地震影响的出生组分别为出生于1951—63年和1951—61年的个体,对应的2005年时的年龄分别为42—54岁和44—54岁。

b. 由于中国各省份实施义务教育法的有效期限不同, Fanget al.(2012)将受义务教育法影响的群体定义为各省在实施义务教育法时年龄为15—16岁的个体。河北省于1986年5月在第六届人民代表大会第四次会议上修订1985年《河北省普及九年制义务教育实施方案》(试行),制定并通过了《河北省普及义务教育条例》,故经笔者计算,受义务教育影响的群体为1970—71年出生组。资料来源:宗树兴. 1986年《中华人民共和国义务教育法》立法和实施研究. 河北大学博士论文, 2010。

c. 吴要武和赵泉(2010)、邢春冰和李实(2011)均使用2005年1%抽样调查数据研究了高等教育扩张、教育机会获得、大学生就业与失业的关系。两项研究分别将22—24岁和21—25岁的个体作为处理组。严敏(2012)使用2006年中国综合社会调查(China General Social Survey,缩写为CGSS)数据研究了高等教育扩张对大学教育溢价的影响。她将2006年21—25岁的个体即在2002—2005年获得最高学位的个体作为处理组,故2005年这些个体的年龄为20—24岁。严格来说,这一群体并非涵盖受高等教育扩招影响的全部个体。若高等教育扩招政策从1999年持续至2007年,受高等教育政策影响的年龄组应为1980—88年出生的个体。

附表4 唐山大地震对教育的影响:共同趋势假定检验

	(1)	(2)	(3)	(4)	(5)	(6)
	唐山	邯郸	保定	承德	除唐山外	全部地区
出生于1951年	-0.2622 (0.2389)	-0.4464** (0.2253)	-0.2225 (0.1962)	-0.9041* (0.4948)	-0.3819*** (0.1431)	-0.3818** (0.1705)
出生于1952年	-0.6568*** (0.2419)	-0.3587 (0.2385)	-0.1873 (0.1871)	0.4793 (0.3976)	-0.1355 (0.1382)	-0.1355 (0.1483)
出生于1953年	-0.4469 (0.2780)	0.2260 (0.2152)	-0.2192 (0.1925)	-0.2246 (0.4289)	-0.0541 (0.1370)	-0.0540 (0.1766)
出生于1954年	-0.0369 (0.2344)	-0.2274 (0.2111)	-0.3701** (0.1695)	0.0789 (0.3406)	-0.2489** (0.1232)	-0.2489 (0.1520)
出生于1955年	-0.0094 (0.2162)	-0.3445* (0.1883)	-0.0882 (0.1911)	-0.1594 (0.3691)	-0.2055 (0.1267)	-0.2055 (0.1631)
出生于1956年	0.0624 (0.2142)	0.0781 (0.1767)	0.4687*** (0.1590)	0.1915 (0.3587)	0.2841** (0.1130)	0.2841** (0.1349)
出生于1957年	0.5803*** (0.2111)	0.0966 (0.1940)	0.5258*** (0.1827)	-0.9943** (0.4130)	0.1635 (0.1276)	0.1635 (0.1558)
出生于1958年	0.8675*** (0.2131)	0.5869*** (0.1890)	0.3779** (0.1804)	-0.0050 (0.3465)	0.3968*** (0.1224)	0.3969*** (0.1454)
出生于1959年	1.2433*** (0.2668)	0.3579* (0.2173)	-0.1384 (0.2051)	-0.2278 (0.3588)	0.0289 (0.1375)	0.0289 (0.1667)
出生于1960年	1.1051*** (0.2402)	0.2534 (0.2253)	0.7337*** (0.1897)	0.3855 (0.3557)	0.5116*** (0.1341)	0.5117*** (0.1502)
出生于1976年	-0.0054 (0.2588)	-0.2245 (0.2278)	-0.0318 (0.1879)	-0.5289 (0.6945)	-0.1344 (0.1442)	-0.1342 (0.1154)
出生于1977年	0.0664 (0.2854)	-0.3159 (0.2152)	-0.2035 (0.1810)	-0.0474 (0.4773)	-0.2306* (0.1343)	-0.2306** (0.1035)
出生于1978年	-0.4109 (0.2689)	-0.1280 (0.2055)	-0.1634 (0.1787)	0.1080 (0.4072)	-0.1164 (0.1288)	-0.1164 (0.0949)
出生于1979年	0.0071 (0.2551)	-0.0541 (0.1971)	-0.2730 (0.2041)	0.1467 (0.4618)	-0.1257 (0.1370)	-0.1257 (0.1118)
出生于1980年	0.6212** (0.2607)	0.0858 (0.1870)	0.1170 (0.1749)	0.1808 (0.5043)	0.1070 (0.1256)	0.1070 (0.1053)



续表

	(1)	(2)	(3)	(4)	(5)	(6)
	唐山	邯郸	保定	承德	除唐山外	全部地区
唐山×出生于1976年						0.1287 (0.2673)
唐山×出生于1977年						0.2970 (0.2701)
唐山×出生于1978年						-0.2948 (0.3433)
唐山×出生于1979年						0.1325 (0.1873)
唐山×出生于1980年						0.5144** (0.2581)
唐山×出生于1981年						0.1757 (0.2240)
唐山×出生于1982年						0.2160 (0.2235)
唐山×出生于1983年						0.3471 (0.2239)
唐山×出生于1984年						0.1822 (0.1749)
唐山×出生于1985年						0.1435 (0.1594)
样本量	2891	4236	5147	1485	10868	13759

注:上述模型的被解释变量为受教育年限的残差,由分性别的受教育年限关于年龄三次多项式的回归得到;上述模型均控制性别和地区虚拟变量,且包括常数项;括号内为异方差稳健标准误;\*、\*\*和\*\*\*分别表示在1%、5%和10%的显著性水平上统计显著。

附表5 唐山大地震对教育的影响:证伪检验结果

	(1)	(2)	(3)	(4)
	辽宁	湖北	湖南	安徽
出生于1961—75年	-0.2900*** (0.0616)	-0.2071*** (0.0597)	-0.1784*** (0.0503)	0.0015 (0.0901)

续表

	(1)	(2)	(3)	(4)
	辽宁	湖北	湖南	安徽
处理组 × 出生于 1961 — 75 年	-0.0262 (0.1215)	-0.0441 (0.1048)	-0.0672 (0.1073)	-0.1164 (0.2105)
处理组	大连	宜昌	湘潭	芜湖
样本量	5685	9426	8989	4505

注:上述模型的被解释变量均为受教育年限的残差,由分地区分性别的受教育年限关于年龄三次多项式的回归得到;上述模型均控制性别和地区虚拟变量,且包括常数项;括号内为异方差稳健标准误;\*、\*\*和\*\*\*分别表示在1%、5%和10%的显著性水平上统计显著。

附表6 唐山大地震对教育的影响:DiDiD模型估计结果

	(1)	(2)	(3)	(4)
	河北和辽宁	河北和湖北	河北和湖南	河北和安徽
河北	-0.0783 (0.0491)	-0.0514 (0.0491)	-0.0311 (0.0429)	0.0546 (0.0720)
河北 × 出生于 1961 — 75 年	0.1904*** (0.0669)	0.1113* (0.0637)	0.0802 (0.0603)	-0.0875 (0.0912)
河北 × 唐山 × 出生于 1961 — 75 年	-0.1506** (0.0742)	-0.1372* (0.0720)	-0.1414* (0.0725)	-0.1567* (0.0804)
样本量	19444	23185	22748	18264

注:上述模型的被解释变量均为受教育年限的残差,由分地区分性别的受教育年限关于年龄三次多项式的回归得到;上述模型均控制性别和地区虚拟变量,且包括常数项;括号内为异方差稳健标准误;\*、\*\*和\*\*\*分别表示在1%、5%和10%的显著性水平上统计显著。

附表7 唐山大地震对教育的影响:安慰剂检验结果 A

	(1)	(2)
	提前两年	错后两年
实验组	-0.0908** (0.0433)	-0.1779*** (0.0408)
唐山 × 实验组	-0.0086 (0.0893)	-0.1128 (0.0856)

续表

	(1)	(2)
	提前两年	错后两年
实验组	1959 — 73 年	1963 — 77 年
前非实验组	1949 — 58 年	1953 — 62 年
后非实验组	1974 — 83 年	1978 — 87 年
样本量	13667	13674

注:上述模型的被解释变量均为受教育年限的残差,由分地区分性别的受教育年限关于年龄三次多项式的回归得到;上述模型均控制性别和地区虚拟变量,且包括常数项;括号内为异方差稳健标准误;\*、\*\*和\*\*\*分别表示在1%、5%和10%的显著性水平上统计显著。

附表8 唐山大地震对教育的影响:安慰剂检验结果 B

影响期限	(1)	(2)	(3)	(4)
	1976 — 78 年	1976 — 80 年	1976 — 82 年	1976 — 84 年
实验组	-0.0578 (0.0459)	-0.0920** (0.0436)	-0.0993** (0.0417)	-0.1657*** (0.0402)
唐山 × 实验组	-0.2597*** (0.0935)	-0.2132** (0.0903)	-0.1914** (0.0874)	-0.1161 (0.0839)
实验组	1961 — 71 年	1961 — 73 年	1961 — 75 年	1961 — 77 年
前实验组	1951 — 60 年	1951 — 60 年	1951 — 60 年	1950 — 60 年
后实验组	1972 — 81 年	1974 — 83 年	1976 — 85 年	1978 — 88 年
样本量	12330	13070	13759	14641

注:上述模型的被解释变量均为受教育年限的残差,由分地区分性别的受教育年限关于年龄三次多项式的回归得到;上述模型均控制性别和地区虚拟变量,且包括常数项;括号内为异方差稳健标准误;\*、\*\*和\*\*\*分别表示在1%、5%和10%的显著性水平上统计显著。

附表9 教育回报率的估计结果

样本	(1)	(2)	(3)	(4)	(5)
	唐山	唐山和邯郸	唐山和保定	唐山和承德	全部地区
LATE - IV(2SLS)					
受教育年限	0.1378*** (0.0392)	0.1470*** (0.0283)	0.2206*** (0.0343)	0.1284*** (0.0308)	0.2077*** (0.0264)

续表

样本	(1)	(2)	(3)	(4)	(5)
	唐山	唐山和邯郸	唐山和保定	唐山和承德	全部地区
LATE - IV (LIML)					
受教育年限		0.1597*** (0.0321)	0.2710*** (0.0473)	0.1429*** (0.0383)	0.2231*** (0.0295)
LATE - IV (GMM)					
受教育年限	0.1383*** (0.0388)	0.1460*** (0.0279)	0.2207*** (0.0339)	0.1187*** (0.0301)	0.2092*** (0.0263)
内生性检验	4.23**	14.38**	29.41**	2.92*	44.50**
弱工具变量检验	6.91***	5.91***	5.22***	4.85***	7.53***
样本量	2891	7127	8038	4376	14876

注:上述模型的被解释变量和解释变量分别为对数化月收入的残差和受教育年限的残差,二者均由分地区分性别的月收入关于年龄三次多项式的回归得到;上述模型均控制性别,且包括常数项,模型(5)还控制地区虚拟变量;对于唐山样本,工具变量为1961—75年各个出生年虚拟变量,对于全部地区样本,工具变量为1961—75年各个出生年虚拟变量及其与唐山虚拟变量的交互项;内生性检验报告的是C统计量( $\text{Chi}^2$ 值),弱工具变量检验报告的是Cragg-Donald Wald统计量(F值);括号内为异方差稳健标准误;\*、\*\*和\*\*\*分别表示在1%、5%和10%的显著性水平上统计显著。

附表10 教育回报率的估计结果:排除一般均衡效应的影响

	(1)	(2)	(3)
唐山样本			
LATE - IV			
受教育年限	0.2020** (0.0948)	0.2175** (0.1049)	0.2016* (0.1049)
第一阶段			
出生于1961—75年	-0.2804*** (0.0757)	-0.2825*** (0.0732)	-0.2759*** (0.0737)
内生性检验	4.56**	4.30**	3.62*
弱工具变量检验	13.74***	14.90***	14.03***
样本量	2888	2888	2884
全部样本			
LATE - IV			
受教育年限	0.2785*** (0.0856)	0.2562*** (0.0802)	0.2441*** (0.0834)

续表

	(1)	(2)	(3)
第一阶段			
唐山	-0.1110*** (0.0412)	-0.1304*** (0.0400)	-0.1277*** (0.0401)
唐山 × 出生于 1961 — 75 年	-0.1711** (0.0852)	-0.1661* (0.0835)	-0.1436* (0.0838)
内生性检验	15.17***	12.28***	10.33***
弱工具变量检验	10.57***	13.48***	11.86***
样本量	13747	13747	13728

注:模型(1) - (3)分别控制了受地震影响的一代的人口规模、受地震影响的一代的平均受教育年限和未受地震影响的一代的平均受教育年限,为进一步保证估计结果的稳健性,将分组变量中的行业替换成职业(6类),估计结果基本不变;上述模型的被解释变量和解释变量分别为对数化月收入的残差和受教育年限的残差,二者均由分地区分性别的月收入关于年龄三次多项式的回归得到;上述模型均控制性别和地区虚拟变量,且包括常数项;对于唐山样本,工具变量为 1961 — 75 年出生组虚拟变量,对于全部地区样本,工具变量为 1961 — 75 年出生组虚拟变量及其与唐山虚拟变量的交互项;内生性检验报告的是 C 统计量( $\text{Chi}^2$  值),弱工具变量检验报告的是 Cragg-Donald Wald 统计量(F 值);括号内为异方差稳健标准误;\*、\*\*和\*\*\*分别表示在 1%、5%和 10%的显著性水平上统计显著。

附表 11 教育回报率的估计结果:纠正样本选择偏误

	(1)	(2)	(3)
唐山样本			
LATE - IV			
受教育年限	0.0988*** (0.0223)	0.1936* (0.1131)	0.2146** (0.1079)
第一阶段			
出生于 1961 — 75 年	-1.7633*** (0.0952)	-0.3578*** (0.0990)	-0.2710*** (0.0737)
内生性检验	3.03*	1.47	2.65
弱工具变量检验	304.70***	13.07***	13.84***
样本量	3284	1561	2931
全部样本			
LATE - IV			
受教育年限	0.0953*** (0.0226)	0.1927*** (0.0893)	0.2520*** (0.0839)

续表

	(1)	(2)	(3)
第一阶段			
唐山	-0.6593*** (0.0514)	-0.1413*** (0.0523)	-0.1190*** (0.0372)
唐山 × 出生于 1961 — 75 年	-0.1876*** (0.0817)	-0.2165* (0.1119)	-0.1553* (0.0853)
内生性检验	3.85***	2.54	8.01***
弱工具变量检验	116.49***	10.19***	11.49***
样本量	15650	7390	15441

注:模型(1)使用 Heckman 两阶段方法纠正了样本选择偏差,模型(2)仅使用男性样本,模型(3)使用的样本还包括户口所在地位于本研究样本地区且为农业户口但居住地为其他地区的个体,且控制了户口所在地和居住地是否一致的虚拟变量;上述模型的被解释变量和解释变量分别为对数化月收入的残差和受教育年限的残差,二者均由分地区分性别的月收入关于年龄三次多项式的回归得到;上述模型均控制性别和地区虚拟变量,且包括常数项;对于唐山样本,工具变量为 1961 — 75 年出生组虚拟变量,对于全部地区样本,工具变量为 1961 — 75 年出生组虚拟变量及其与唐山虚拟变量的交互项;内生性检验报告的是 C 统计量( $\text{Chi}^2$  值),弱工具变量检验报告的是 Cragg-Donald Wald 统计量(F 值);括号内为异方差稳健标准误;\*、\*\* 和 \*\*\* 分别表示在 1%、5% 和 10% 的显著性水平上统计显著。

附表 12 教育回报率的估计结果:其他稳健性检验

	(1)	(2)	(3)	(4)
唐山样本				
LATE - IV				
受教育年限	0.2179** (0.1028)	1.6600* (0.9123)	1.4735** (0.7252)	0.1690** (0.0768)
第一阶段				
实验组	-0.2899*** (0.0767)	-0.0382** (0.0152)	-0.0431*** (0.0119)	-0.4874*** (0.0951)
内生性检验	2.92*	3.81*	3.53**	2.17
弱工具变量检验	14.26***	6.35**	13.06***	26.27***
样本量	2891	2891	2891	2891
全部样本				
LATE - IV				
受教育年限	0.2093** (0.1012)	1.9893** (0.9009)	1.6697** (0.6629)	0.1688** (0.0664)

续表

	(1)	(2)	(3)	(4)
第一阶段				
唐山	-0.0424 (0.0517)	-0.0083 (0.0080)	-0.0120** (0.0056)	-0.1148*** (0.0526)
唐山 × 实验组	-0.2122** (0.0833)	-0.0299* (0.0172)	-0.0310** (0.0132)	-0.3726*** (0.1085)
内生性检验	3.16*	6.70***	5.83**	3.13*
弱工具变量检验	11.51***	3.70**	8.80***	15.56***
样本量	13759	13759	13759	13759

注:模型(1)控制了少数民族虚拟变量、目前的居住地类别、父母的平均受教育年限以及父母的平均受教育年限是否均缺失;模型(2)和(3)的被解释变量分别为是否完成初中教育和是否接受高中教育;模型(4)的实验组为1964—69年出生组,前、后非实验组分别为1951—60年和1976—85年;上述模型的被解释变量和解释变量分别为对数化月收入的残差和受教育年限的残差,二者均由分地区分性别的月收入关于年龄三次多项式的回归得到;上述模型均控制性别和地区虚拟变量,且包括常数项;对于唐山样本,工具变量为1961—75年出生组虚拟变量,对于全部地区样本,工具变量为1961—75年出生组虚拟变量及其与唐山虚拟变量的交互项;内生性检验报告的是C统计量( $\text{Chi}^2$ 值),弱工具变量检验报告的是Cragg-Donald Wald统计量(F值);括号内为异方差稳健标准误;\*、\*\*和\*\*\*分别表示在1%、5%和10%的显著性水平上统计显著。