

寻找阿基米德的“杠杆” ——“出生季度”是个弱工具变量吗？

吴要武*

摘 要 A K91 用出生季度作工具变量估计教育回报率，在引起众多质疑的同时，也推动了微观经济计量学的一个活跃领域——探讨弱工具问题。本文使用 2005 年中国人口 1% 抽样调查数据，重新讨论了教育回报率的 2SLS 估计中，“出生季度”是否为弱工具变量问题：在发达国家，完成高中阶段教育者在队列人口中比例过高，导致出生季度对教育变异的影响很小，因此，出生季度可能是个弱工具变量；在发展中国家，能够进入高中阶段学习者，在队列人口中不到一半，出生季度对教育变异的影响很大，因此，是个强工具变量。重新估计的结果显示，2SLS 估计及其改进形式得到的教育系数显著高于 OLS 估计值。除了“数据质量”，模型误设可能是 A K91 遭遇弱工具的另一个原因。

关键词 工具变量，弱工具，教育收益率

一、引 言

教育收益率一直是劳动经济学家关注的重要问题，使用观测性数据时，用工具变量方法估计教育收益率占据突出地位。Angrist and Krueger (1991) (下文简称 A K91) 估计美国劳动力市场教育收益率的方法有着重要的影响。其基本方法是：发达国家或早或迟都颁布了《义务教育法》，在满 16 周岁（或 17 周岁）之前退学为法律所禁止，而人们的出生季度是随机分布的。如果 6 岁队列中不同出生季度者同时入学，那些出生在上半年的学生，到中学

* 中国社会科学院人口与劳动经济研究所。通信地址：北京市建国门内大街 5 号人口与劳动经济研究所；100732；电话：(010) 85196067；E-mail: wuyw@cass.org.cn。感谢蔡（昉）师长长期的教诲与耐心的指导，师门十年，他一招一式地教会了我作研究；感谢 Scott Rozelle 教授对本文的激励性评述。两位匿名审稿人精彩的评议对改进本文弥足珍贵，绝非客套。2007—2008 年度，在斯坦福大学经济学系 Seema Jayachandran, Nick Bloom, Luigi Pistaferri 等教授的课堂学习与讨论中受益良多，并在此期间完成了本文的初稿。本文得到中国社会科学院重点课题《中国城镇非正规就业问题》的资助。在此一并致谢。作者文责自负。

2001 年以来，刘学军博士在理论、方法和情感上，都给了我兄弟般的帮助与支持，此文是对这位训练有素却英年早逝的青年经济学家的纪念。如果说经济学训练过程如穿越浩瀚沙漠般漫长，刘学军和我就是那两匹结伴前行的骆驼。

的某个学年结束时, 已满 16 周岁, 可以退学 (进入劳动力市场); 下半年出生的学生, 却因未满 16 周岁而必须“滞留”在学校里多接受一年教育, 其结果, 与下半年出生的同一队列人口相比, 第一季度出生者因更早退学而有更短的平均受教育年限; 更短的受教育时间会导致更低的工资。由于后三季度出生者的教育优势是《义务教育法》带来的, 这就相当于一个自然实验, 利用此外生冲击产生的变异可以解决忽略变量带来的估计偏差问题。

A K91 的估计结果显示, 2SLS 与 OLS 方法的估计结果几乎没有显著差异。两位经济学家谨慎地推断: 用 OLS 估计得到的教育回报率存在的偏差可能并不大。但其他经济学家使用同一个数据发现, “出生季度”是个弱工具变量: 人们的出生季度与其受教育年限变异之间, 只有微弱的相关性。存在弱工具问题时, 2SLS 估计不仅难以矫正 OLS 估计的偏差, 反因有更大的标准差而有更低的效率, 导致“治疗比疾病本身更坏” (Bound, Jaeger, and Baker, 1995) (下文简称 BJB95)。

尽管 A K91 的结论受到质疑, 但很多流行的经济计量学教科书都把“出生季度”这个工具变量以及对它的质疑同时吸纳进来 (Wooldridge, 2005; Stock and Watson, 2006)。一个有趣的现象是, 当“出生季度”被其他学者检验出为弱工具后, 两位作者在后来的文章中仍然把“出生季度”作为工具变量的一个案例来介绍 (Angrist and Krueger, 1999; 2001)。这意味着, A-K 一直坚持“出生季度”是个有效的工具。从今天的学术关注度看, A K91 在微观经济计量学上开辟了一个非常活跃的研究领域: 弱工具问题。简要回顾一下相关文献, 经济学家的关注点集中在如何识别弱工具 (Staiger and Stock, 1997; Stock and Yogo, 2004); 探讨弱工具条件下 2SLS 估计的偏差及决定因素, 如 X 与 Z 的相关性 (第一阶段的 R^2)、结构方程与约简型方程的误差项 u 与 v 的相关性 ()、观测值数量 (n)、工具变量的数目 (K) 等 (Hahn and Hausman, 2002; 2003; Murray, 2006); 在存在弱工具的条件下, 如何改进计量方法以获得稳健一致的估计量 (Baum *et al.*, 2007; Hansen *et al.*, 2006)。但尚未见到探讨“出生季度”成为弱工具原因的文献。在作者看来, 这是因为研究者都使用了同一个数据。中国 2005 年的 1% 人口抽样调查数据支持本文作新的探索。本文的研究目的有两个: 第一, 开辟不同的研究路径, 寻找产生弱工具的直接原因和社会经济根源; 第二, 弄清中国城镇劳动力市场上的教育回报率。

二、什么因素影响了工具变量的相关性?

BJB95 使用同样的数据复制了 A K91 的研究, 根据 2SLS 第一阶段回归的 F 值和 R^2 值, 判断出生季度与受教育年限之间只有很小的相关性, 认定出生季度是个弱工具。但他们并没有解释, 出生季度与受教育程度之间为什么会

弱相关。或者更严格表述为“为什么‘出生季度’在美国是个弱工具”？在开始讨论中国劳动力市场问题之前，本文先回答这个问题。

根据 A K91 报告的信息，在 1980 年美国人口普查数据中，1930—1939 年出生人口队列，高中及以上者所占比例已达到 77 %——后来的文献几乎都使用这个队列的数据；1940—1949 年出生队列中这一比例则达到了 86 %。A K91 发现，在 1950 年和 1960 年以后出生队列里，《义务教育法》失效了：不同出生季度者之间显示不出受教育差异。在作回归分析时，他们干脆舍弃了 1950 年以后出生队列样本，理由是一部分人尚处于受教育阶段。作者发现，1950 年以后出生队列高中及以上者所占比例超过了 90 %。由此推测，“出生季度”与“受教育年限”相关性低，是美国高中入学率“过高”导致的。受教育年限在出生季度间的变异过小导致了弱工具出现。

一般说来，16 岁人口队列通常处于初中或高中阶段，美国高中阶段的教育已得到普及，几乎每个学生都能从初中直接升入高中（8 年级升入 9 年级），从美国劳动年龄人口中高中及以上者所占比例，可以看出这个事实。为了说明高中阶段的升学率如何影响了“出生季度”作为工具变量的有效性，本文以中国为例作阐释。与美国相比，中国的特殊环境是，高中阶段的教育远未普及，初中升高中时存在严格的入学考试，升学率也远低于美国。

中国的 16 岁人口队列常常处于初中三年级升高中的阶段。从经验直觉可知：如果进入高中的升学率为 0，那么，无论是出生在上半年还是出生在下半年的学生，都无法进入高中，其结果，出生季度与受教育年限之间没有相关性；如果进入高中的升学率为 100 %，所有初三学生都可以进入高中，忽略掉辍学因素，那么，出生季度与受教育年限之间也没有相关性。可是，一旦初中升高中的升学率介于 0—100 % 之间时，出生季度就会影响学生的选择，从而影响不同季度出生者的最终受教育年限。

本文作出一些理想化的假定，模拟并说明升学率如何在出生季度与受教育年限之间扮演了重要角色。这是我们理解的中国和美国之所以出现差异的根本原因。假定如下：

1. 假定所有人口队列都在 16 岁时升入高中（9 月份入学）。

2. 如果第一年没有考上高中，假设决定学生选择的因素只有一个：是否满 16 周岁。前三个季度出生者，到 9 月份入学时已经或即将满 16 周岁，他们选择加入劳动力市场；第四季度出生者，因不满 16 周岁，无法进入劳动力市场，他们选择回初中三年级复读。

3. 假定初中应届生和复读生升高中的比例是一样的。

4. 假定没有辍学。在中国“严进宽出”的教育体制下，这是一个可以接受的假设。但在 A K91 那里，“辍学”却是导致受教育年限在不同出生季度者之间产生差异的主要原因。

然后，我们对不同升学率下，出生在第四季度者的受教育优势进行模拟。

图 1 显示: 如果入学率为 (0,100), 则第四季度出生者升入高中的比例就会高于其他三个季度出生者。纵轴代表某个队列中, 第四季度出生者比其他季度出生者“多出的高中生人数”在队列总人口中的比例。显然, 当高中入学率为 50 % 时, 第四季度出生者升入高中的优势最大: “多”升入高中者占队列总人口的比例为 6.25 %。¹ 在一个大样本数据中, 这个差异应该能够显著地呈现出来。

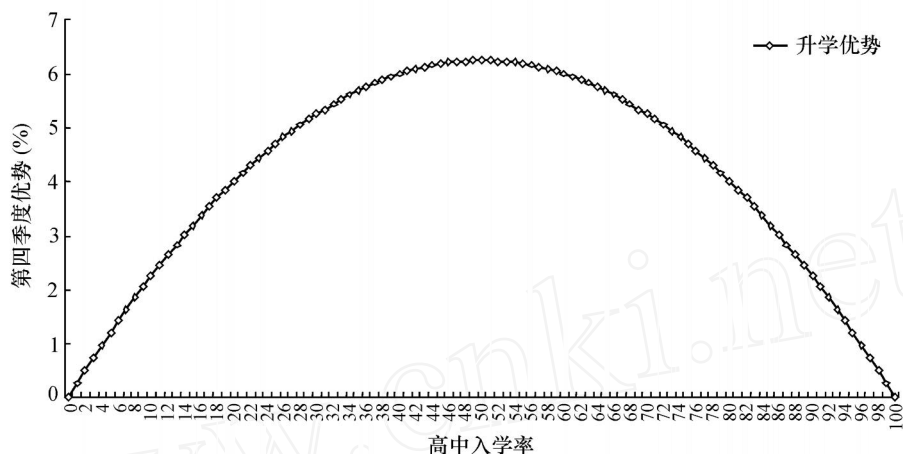


图 1 出生在第四季度者的受教育优势

高中入学率越是接近 50 %, 出生季度对升学, 从而对受教育年限的影响越大, 这时, 出生季度可能成为强工具变量; 高中升学率越是接近 0 或者 100 %, 出生季度与高中升学率的相关性越低, 那么, 出生季度更可能成为弱工具。美国有 80 % 以上的高中升学率, 却没有类似中国这样的升学竞争机制, 出生季度对受教育选择的影响途径仅仅来自“辍学”。一般说来, 只有少数患有身心疾病、距离学校过远和格外厌倦学习的学生才可能中途“辍学”, 因此, 不同出生季度者之间的受教育差异可能非常小, 这时, 出生季度与受教育年限之间只有微弱的相关性。由于美国完成高中阶段教育者在人口队列中的比例越来越接近 100 %, A-K 不可能找到一个更好的数据来重新检验工具变量的强弱了。

基于以上分析, 本文提出一个可以检验的假说:

中国人口队列的高中入学率在持续提高, “出生季度”作为工具变量时, 会随人口队列演进而得到增强, 由弱工具变量转化为强工具变量。

¹ 具体的模拟方法是, 假定人们的决策分为两期: 第一期, 不同出生季度的 16 岁队列, 升入高中的比例一样; 第二期, 将队列总人口减去升入高中的人口, 然后除以 4, 得到第四季度出生但第一期没有升入高中的人口, 然后将这个人口数乘以高中入学率, 得到第二期升入高中的人数; 再将第二期升入高中的人数 (只有第四季度出生者选择了复读) 除以队列总人口。而且假定人们只上到高中。

三、数据与识别

(一) 数据

2005年11月1日，国家统计局作了全国1%人口抽样调查（俗称“小普查”），问卷中设计了劳动者的工作特征和收入等问题，使我们能利用这个数据深入分析“出生季度-受教育年限-工资水平”之间的关系。从最终回收问卷看，实际调查了全国人口的1.31%，由于在不同省市抽样比不一样，国家统计局提供了各个省市的抽样比和每个观测值代表的权重。从最终数据中，以家庭为单位抽取了1/5的样本，提供给部分研究机构，共计2 585 481个观测值，741 226户，其中有723 697个家庭户，17 529个集体户。

由于历次人口普查或者“小普查”的问卷都在变动，本研究需要的一些信息在以前的人口普查问卷中存在，而2005年的问卷中却没有设计，所以，本文还辅助使用了1990年和2000年的人口普查数据。如2000年问卷中设计了“受教育程度”的分类——“是否成人学历”。在中国，“成人学历”通常是毕业后才申请，与普通升学考试无关。在识别出生季度可能影响升学时，应该考虑“成人学历”这个因素，但这个信息在2005年调查时缺失。

受“文革”的影响，2005年小普查数据中，高中及以上者在出生队列中的比例有个大的波动。由于本文更关注城镇劳动力市场，作者只选取城镇人口作为研究对象，这样一来，受教育状况在人口队列中基本上是稳定上升的。中国的法定退休年龄一直为60岁，停留在劳动力市场上的最大人口队列为1945年出生者。1945年及以后出生队列人口中，高中及以上者的比例从约20%上升到约50%。中国于1999年开始扩大高等教育规模，高校录取人数8年间增长了近5倍，大学毕业生供给的急剧增长或许会影响教育回报率。高校扩招主要影响了1980年以后的出生队列，到2005年人口抽样调查时，这些年轻人口队列中，相当大一部分人还在接受教育，未进入劳动力市场。1998年国企改革加速导致了严重的就业冲击，不少地方政府鼓励大龄劳动者退出劳动力市场，大龄队列的劳动参与率下降。考虑了上述因素后，本文排除接近退休的大龄人口队列和尚未完成教育的年轻人口队列，截取1950—1979年出生队列作为分析对象。

在本文所限定的研究群体中，高中及以上者的比例为36.1%，而初中及以上者的比例则达到78.9%，这意味着，样本人口初中毕业后，有54%的人没有升入高中。中国城镇人口在初中和高中阶段有更大的变异，出生季度与升学选择之间因之有更大的相关性，用出生季度作工具变量时，可能不会遭遇美国那样的弱工具风险。但中国也同样面临一些可能导致弱工具的因素。在识别部分，作者评估这些因素对出生季度与受教育年限之间相关性的影响。

(二) 识别

由于中国缺少美国那样完备的法律体系和严格的执法环境,对“出生季度-受教育年限-市场表现”的传导机制必须进行更加谨慎的识别。本文的识别分两部分:第一,出生季度与受教育年限变异之间是否存在因果关系。作者观察出生在前三季度的人,受教育年限是否显著低于第四季度出生者;然后,继续观察同一队列中,不同出生季度者的收入变异是否与受教育年限变异相一致。第二,不同出生季度者之间的受教育年限变异可能因一系列因素的扰动而被弱化,对这些因素要进行分析并作出评估。由于观测性截面数据是包含了各种影响后的最终结果,因此,本文先讨论对识别起弱化作用的因素,然后再考察“出生季度-受教育年限-小时工资”之间的关系。使季度间受教育年限变异弱化的因素在中国可能以其他方式存在,可以把它们视为特殊的约束条件,包括:

1. 《义务教育法》

中国最早的《义务教育法》颁布于 1986 年,但该法并没有美国那样的“满 16 周岁或 17 周岁才能离开学校”的规定,而仅规定,实行 9 年制义务教育,“满 6 周岁入学”是建议而非强制性的。在现实生活中,人们可以把孩子的进入小学推迟到 8 岁队列,也可以在进入 6 岁队列但不满 6 周岁时入学:面对“提前入学”,教育部门的官员不会公开赞扬,但也决不会施加惩罚。从 1990 年 7 岁队列各月份出生者的在校率看出,变化是平缓的,7、8 月份出生者与 9、10 月份出生者几乎没有差别。可以确信:中国的小学实际入学年龄与美国没有差异。由于《义务教育法》从 1986 年 7 月开始实施,对本文观察的 1950—1979 年出生队列来说,基本上不受其“入学年龄”规定的影响。²作者推断:如果“出生季度-受教育年限差异-收入差异”这个传导机制在中国存在,一定与《义务教育法》没有关系。

是什么因素导致了“出生季度”对受教育年限产生影响呢?是中国严厉的就业政策。从建国到 2002 年,就业压力一直是中国政府面临的严峻挑战。无论城乡,就业岗位都是极其稀缺的,正因为如此,政府对“入职年龄”有着严格的限制:劳动年龄人口的法定标准是“16 周岁”,“除文艺、体育和特种工艺部门外,均不能使用未满 16 周岁的未成年人”。确保“就业政策”得到严格执行的一个重要因素是城乡家庭之间的相互监督:如果邻居家的孩子未满 16 周岁就被“招工”、“招干”或者进入生产队“当社员,领工分”,对

² 与 2006 年的修订版相比,1986 年版本的《义务教育法》显得温和,一直在使用“应当”这个“劝告性”词汇,而不是体现法律严肃性的“必须”和惩罚。第五条规定:“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。”义务教育法的目的是鼓励教育,“不满 6 周岁”的孩子提前入学不会被教育部门的官员劝阻。“条件不具备”是非常有弹性的,可以解读为允许儿童在“6 岁—8 岁”队列时入学。

自己的孩子就是不公平的：抢占了极其稀缺的就业机会。无论是计划经济时代还是改革开放以后，无论是城镇还是农村，不满16周岁，都无法加入劳动者行列并获得工作岗位。

改革开放以后，农村实行了联产承包责任制，城镇则出现了劳动力市场，但即使在城镇新兴部门，雇佣16周岁以下的劳动者也是违法行为。由于劳动力丰富和价格低廉，中国城镇劳动力市场上很少有童工。从社会习俗的角度看，中国在历史上就将16周岁视为男性成年的界限，1963年颁布的刑法（33稿），将“承担完全刑事责任”的年龄规定为16周岁³（李学武，2003）。进入劳动力市场的年龄限制，与发达国家义务教育法的作用一样，把一些想提前退学者阻留在学校。

A K91 在分析美国的义务教育法影响受教育年限时，指出两个途径：（1）不满16周岁退学要受到专门执法官员的惩罚；（2）不满16周岁无法获得进入劳动力市场的许可证（work certificate or work permit）。⁴《义务教育法》和“就业许可证”这“两块等高的木板”共同构成中学生合法退学的“闸门”，使不同出生季度者在10—11年級的边际上出现受教育变异。后来学者在讨论工具变量发生机制时，都注意到了《义务教育法》，却忽视了由学校发放的“就业许可证”这个同样重要的条件。在中国，不满16周岁退学不被严厉禁止，但在限制进入劳动力市场方面，严厉程度却不亚于美国。尽管A K91声称：“如果没有《义务教育法》，不同出生季度群体之间就不会出现受教育的差异”（原文第989页）。当他们在解释《义务教育法》为什么会被遵守时，立即求助于《公平劳动标准法》（原文第992—993页）。我们下面会看到，在不受义务教育法影响的中国大龄人口队列中，出生季度间的受教育变异同样是显著的。因为劳动力政策发挥了同样的作用。

2. 升高中时的年龄

美国的初中和高中之间没有门槛，但中国有。初中毕业升入高中要经过非常严格的考试，“中考”是比“高考”覆盖范围更广、参与人数更多的社会性考试。无论城镇还是农村，初中毕业生如果考不上高中，他们面临的选择有两个：继续复读或者等待就业。但是，按照1986年《义务教育法》建议的6岁队列入学，在初中毕业时，这个队列可能处于15岁队列。即使出生在第

³ 中国《刑法》第十七条规定，“已满十六周岁的人犯罪，应当负刑事责任。”……“因不满十六周岁不予刑事处罚的，责令他的家长或者监护人加以管教；在必要的时候，也可以由政府收容教养。”《民法通则》第十一条规定，“十八周岁以上的公民是成年人，具有完全民事行为能力，可以独立进行民事活动，是完全民事行为能力人。十六周岁以上不满十八周岁的公民，以自己的劳动收入为主要生活来源的，视为完全民事行为能力人。”1994年的《劳动法》第九十四条规定，“用人单位非法招用未满十六周岁的未成年人的，由劳动行政部门责令改正，处以罚款；情节严重的，由工商行政管理部门吊销营业执照。”显然，无论政府还是家庭，都把“满16周岁”视为迈入成人行列的重要门槛。

⁴ 参见A K91的第992—993页。其他学者也指出了类似的影响：义务教育法规定达到某个年龄的儿童必须入学接受教育；只有达到某个年龄才能得到“工作许可证”（Lleras-Muney，2002）。

一季度的人,也同样因无法进入劳动力市场而选择复读或等待。按照本文的假设,16岁队列的人才会因出生季度的约束而选择进入市场或者继续升学。那么,15岁队列是个升入高中的惯常年龄吗?如果是,出生季度与受教育年限之间的相关性就会弱化。

2000年的人口普查和2005年1%抽样调查数据的标准时间都是当年11月1日0时,正处于新学年第一学期的期中,两个数据中的16岁队列分别对应着1984年和1989年出生的人口。1990年的普查标准时间为6月1日0时,是某学年第二学期的末尾。调查时点的在校生,通常是上个年度入学或者升级的,为了控制这个特殊变化,本文同时观察几个“适龄”出生队列在高中和初中的分布状况,1990年多观察一年,见表1。

表 1 不同调查年份在校中学生的分布

单位: %

1990 年			2000 年			2005 年		
出生年份	初中	高中	出生年份	初中	高中	出生年份	初中	高中
1971	27.9	72.1	1982	18.8	81.2	1987	16.2	83.8
1972	46.7	53.3	1983	34.4	65.6	1988	30.4	69.6
1973	69.9	30.1	1984	60.3	39.7	1989	56.8	43.2
1974	89.9	10.1	1985	87.1	12.9	1990	86.3	13.7
1975	97.9	2.1						

注:1990年人口普查的标准时间为6月1日0时,正是中国学年的期末,观察到的在校生入学年份为上一年,分别对应着14—18岁队列。2000年人口普查和2005年1%抽样调查的标准时间为11月1日,两个数据中的出生年份分别对应着15—18岁队列。

资料来源:1990年第四次人口普查1%数据;2000年第五次人口普查0.95%长表抽样数据;2005年1%人口抽样调查数据。

先看1990年的15—19岁队列,15岁队列的就学状况相当于2000年和2005年数据的14岁队列。1974年出生者上一年为15岁队列,这个队列的中学在校生只有10.1%为高中生,其他90%还是初中生;1973年出生者为上一年的16岁队列,进入高中的比例上升到30.1%,但在初中就读者仍然占70%;直到17岁队列,读高中的人数才超过了读初中的人数。

接着观察2000年和2005年的数据。中国城乡青少年的升学年龄似乎略有下降,但并未改变基本态势:15岁队列升入高中的比例在2000年为12.9%,2005年略微提升到13.7%;16岁队列进入高中的比例迅速提高,2000年,有近40%为高中生;2005年则超过43%。这两个观测年份同样是17岁队列的高中生比例超过初中生。

这证明本文的担心不是一个严重问题:绝大多数高中及以上者是在16岁队列及以后进入高中的。进一步的识别可参见附表1: DID分析结果显示,当人口队列从15岁向16岁演进的时候,第四季度出生者在校率的变化显著为正,即16岁队列的第四季度出生者更不容易离开学校。可以得出结论:“出生在第四季度”对16岁队列的孩子“选择继续升学”有显著的正效应。

3. 初中以后的职业技术教育

1999 年以来，中国政府开始建立劳动预备制度，职业技术教育的招生规模扩大，为那些初、高中毕业后进入劳动力市场的年轻劳动者提供职业技术教育。可以推断，16 岁队列的上半年出生者已满 16 周岁，这一年初中毕业并直接进入职业技术学校，那么，他们的“受教育年限”仍然在增长，“中专/技校”相当于“高中阶段”的教育。这时，出生季度对教育的影响就会弱化。

如果有这种“选择性偏差”，应该从不同季度出生者的“受教育类别”上显示出来。表 2 报告了 2000 年人口普查数据中⁵，中专毕业生在三个人口队列内不同季度出生者群体中所占的比例。在所有的三组出生队列中，第四季度出生者都有略高的中专毕业生比例。这意味着，第一季度出生者在初中毕业后并没有更多地进入职业技术教育体系。显然，“劳动预备制度”尚未影响到目标人口队列，作者担心的使工具变量弱化的这个风险不存在。

表 2 不同人口队列的中专毕业生比例 单位：%

	1950—1959	1960—1969	1970—1979
季度 1	5.55 (22.9)	5.67 (23.1)	10.85 (31.1)
季度 2	5.05 (21.9)	5.45 (22.7)	10.73 (30.95)
季度 3	5.35 (22.5)	5.35 (22.5)	11.38 (31.75)
季度 4	5.82 (23.4)	5.66 (23.1)	11.72 (32.16)
总计	5.46 (22.7)	5.54 (22.9)	11.20 (31.53)

注：括号内为标准差。
资料来源：2000 年人口普查长表 0.95 % 抽样数据。

4. 成人高等教育状况

在中国的高等教育体系中，成人学历占相当大的比例。2000 年人口普查数据显示，在 1950—1979 年出生队列（城镇人口）中，大专毕业生的 49 % 是成人学历；本科毕业生中成人学历者占 22.1 %。拥有成人学历者被调查时会报告最高学历。如果第四季度出生者有更高的比例升入高中，从而比前三个季度出生者有更长的受教育年限，一旦前三个季度出生者选择成人高等教育，那么，就会缩小与第四季度出生者的受教育年限差异，从而弱化出生季度对受教育年限的影响。

⁵ 在 2005 年 1 % 人口抽样调查问卷中，对“受教育程度”的分类没有区分“高中”和“中专”。本文用 2000 年人口普查数据来分析。

另一方面,作者还担心,中国的成人高等教育并不是完全面向完成了高中阶段教育的成年人。比如,如果初中毕业的前三季度出生者,通过成人教育获得了大专文凭,就会高估其实际受教育年限。弄清前三个季度出生者是否有更多的成人高等教育学历,有助于数据质量评估。表 3 显示了城镇出生人口队列为 1950—1979 年成人教育学历的分布。可以看出,出生在第一、二季度者,其成人学历的比例要比第三、四季度出生者高出 1—1.5 个百分点。成人教育缩小了第四季度出生者的受教育优势。

表 3 成人教育学历的分布

单位: %

出生季度	高中	中专	大专	大学本科
季度 1	6.4	24.5	49.7	22.9
季度 2	6.1	24.7	49.5	22.8
季度 3	6.0	23.4	48.6	21.7
季度 4	6.0	23.2	48.2	21.4
总计	6.2	23.9	48.9	22.1

资料来源:2000 年人口普查长表 0.95 % 抽样数据。

5. 辍学与进入市场

在美国,出生季度对受教育年限的影响路径为辍学,中国同样存在辍学现象。但问题是,辍学会成为不同出生季度者受教育差异的直接原因吗? 2005 年调查数据显示,在 1950—1979 年出生人口队列中,初中和高中“辍学”的比例分别是 1.1 % 和 0.34 %。进一步的观察显示,初中和高中阶段的辍学者,四个出生季度的比例分别为 0.84 %、0.84 %、0.82 % 和 0.87 %。没有显著的季度差异。显然,“辍学影响受教育差异”的机制在中国不存在。出生季度对受教育年限的影响并非来自前三个季度出生者的辍学率高,那么,只能来自第四季度出生者的升学率高。

进入劳动力市场的衡量指标是劳动参与率:在 1990 年人口普查数据中,1973 年四个季度出生(城镇)人口的劳动参与率分别为 37.3 %、36.2 %、32.3 % 和 28.0 %;1974 年四个季度出生人口的劳动参与率分别为 23.2 %、19.5 %、15.5 % 和 13.5 %。在进入劳动力市场的合法边际年龄上,出生季度显示出非常显著的差异。满 16 周岁对进入劳动力市场有着显著影响。

6. 选择性偏差

当 A K91 把出生季度作为工具变量时,遇到的一个批评是选择性偏差。同一个队列(6 岁)中,那些下半年出生的孩子,或许会在第二年入学,一些父母相信入学年龄增大会让孩子有更好的学校表现甚至延长了受教育年限。如果存在这样的选择性,受教育年限的变异就可能不是外部冲击导致的。如果真有这样的优势,第四季度出生者都可以选择推迟一年入学。A K91 已经回

答了这个问题：观察孩子的入学年龄。如果第四季度出生的孩子跟随下一个人口队列入学，最终会发现一个结果——这个季度的孩子入学年龄更大（接近7岁）。但事实为，第四季度出生者的入学年龄平均为6.07岁，第一季度出生者的入学年龄平均为6.45岁，第二和第三季度出生者分别为6.28岁和6.08岁。第四季度出生者实际入学年龄最小。

从后来的文献看，研究者都关注工具变量的强弱而不怀疑其有效性。中国的数据无法观察人们的入学年龄，作者假定中国家庭在孩子入学年龄（生月大小）上的选择与美国没有差异。

四、描述性统计结果

（一）不同出生季度者的受教育年限

表4的描述性统计结果显示，在所观察的30个出生队列中，第四季度出生者的平均受教育年限显著高于前三个季度出生者，只有3个队列的结果不显著（5%水平）：1958年、1964年和1965年。在总样本中，前三个季度出生者平均受教育年限比第四季度出生者少0.19年，样本平均受教育年限为9.74年，这意味着，第四季度出生者受教育年限比总样本高出约1.96%，美国1930—1939年和1940—1949年出生队列中，相应的差异只有约0.5%。

我们分人口队列继续比较出生季度与受教育年限之间的关系：1950—1959年出生队列的平均受教育年限为8.74年，前三个季度出生者的平均受教育年限为8.69年，比第四季度出生者低0.184年；1960—1969年出生队列的平均受教育年限为9.87年，前三季度出生者的平均受教育年限为9.83年，比第四季度出生者低0.143年；1970—1979年出生队列的平均受教育年限为10.34年，前三季度出生者的受教育年限为10.28年，比第四季度出生者低了0.224年。在三个分组队列中，受教育年限的季度差异都在1%水平上统计显著。

美国1930—1939年出生队列中，与第四季度出生者相比，前三个季度出生者的受教育年限分别少0.124年、0.086年和0.015年；1940—1949年出生队列中，前三季度出生者的受教育年限分别少0.085年、0.035年和0.017年。显然，每个人口队列分组的数据，季度间受教育年限的差异都大于美国，都大于其最好的1930—1939年队列数据。由此可以断言：出生季度在中国导致的受教育年限变异大于美国。在一个巨大的样本里，出生季度间更大的受教育年限变异应该带来一致的估计结果。

表 4 不同人口队列中受教育年限的差异

单位:年

出生年份	前三季度	第四季度	差异	出生年份	前三季度	第四季度	差异
1950	7.759 (0.037)	7.929 (0.059)	- 0.170 (0.070)	1965	9.724 (0.024)	9.747 (0.039)	- 0.023 (0.046)
1951	7.890 (0.037)	8.078 (0.059)	- 0.188 (0.070)	1966	9.624 (0.024)	9.811 (0.042)	- 0.187 (0.048)
1952	7.923 (0.033)	8.100 (0.055)	- 0.177 (0.064)	1967	9.573 (0.026)	9.826 (0.042)	- 0.253 (0.049)
1953	8.183 (0.033)	8.425 (0.053)	- 0.242 (0.062)	1968	9.766 (0.022)	9.893 (0.037)	- 0.127 (0.043)
1954	8.507 (0.031)	8.750 (0.050)	- 0.242 (0.058)	1969	9.805 (0.024)	9.899 (0.038)	- 0.093 (0.045)
1955	8.722 (0.031)	9.010 (0.052)	- 0.289 (0.060)	1970	9.880 (0.022)	10.069 (0.037)	- 0.189 (0.043)
1956	9.041 (0.031)	9.197 (0.049)	- 0.156 (0.058)	1971	9.996 (0.023)	10.190 (0.039)	- 0.194 (0.046)
1957	9.275 (0.029)	9.454 (0.049)	- 0.179 (0.057)	1972	10.069 (0.023)	10.224 (0.038)	- 0.155 (0.045)
1958	9.466 (0.029)	9.550 (0.048)	- 0.084 (0.056)	1973	10.153 (0.024)	10.308 (0.038)	- 0.155 (0.045)
1959	9.699 (0.032)	9.905 (0.049)	- 0.206 (0.059)	1974	10.221 (0.024)	10.389 (0.039)	- 0.168 (0.046)
1960	9.939 (0.029)	10.186 (0.053)	- 0.247 (0.060)	1975	10.309 (0.025)	10.589 (0.040)	- 0.280 (0.047)
1961	10.063 (0.034)	10.255 (0.052)	- 0.192 (0.062)	1976	10.431 (0.026)	10.739 (0.042)	- 0.308 (0.049)
1962	9.972 (0.025)	10.251 (0.036)	- 0.280 (0.044)	1977	10.654 (0.027)	10.873 (0.042)	- 0.219 (0.050)
1963	9.970 (0.021)	10.114 (0.037)	- 0.144 (0.043)	1978	10.736 (0.026)	11.021 (0.039)	- 0.285 (0.047)
1964	9.954 (0.024)	9.926 (0.039)	0.028 (0.046)	1979	10.726 (0.025)	11.025 (0.041)	- 0.299 (0.048)

注:括号内为标准差。

(二) 对第四季度出生者受教育优势的进一步考察

在中国城镇劳动力市场上, 由于受教育水平的持续提高, 年轻劳动人口的受教育年限高于年老者, 在作同一个出生队列内不同出生季度者的教育比较时, 这种受教育变化趋势会带来误差。A K91 建议用移动平均的方法消除这种趋势性误差: 通过将某个出生季度前面两个季度与后面两个季度(- 2, + 2)的受教育年限进行移动平均得到预测值, 然后, 用所有这个季度出生者的实际受教育年限减去预测值得到一个变差。这个平均变差的波动可显示某出生季度真实的受教育优势。沿着同样的技术路线, 本文计算了不同出生季度者的受教育年限变差, 发现所有第四季度出生者都是正值。见图 2。与 A K91 的

结果相比，存在一个有趣的差异：在可观测的29个出生队列中，美国第一季度出生者的受教育变差，有27个小于0；中国第一季度出生者的变差却有18个大于0，但所有29个第四季度的变差都大于0。从中可以得出两个结论：第一，第四季度出生者的确有显著的受教育优势；第二，导致出生季度间教育年限变异的机制，在中美之间是显著不同的。

图2还回答了BJB95的一个挑战性的质疑：出生季度或许具有选择性，富裕家庭更愿意生孩子时避开冬季。如果BJB95这个论断成立，那么，第二和第三季度出生者应该有更大的比例出生在富裕家庭，而父母的财富与子女受教育水平之间是强正相关的，因此，第二、三季度出生者应该有更长的受教育年限。图2显示的信息恰恰相反：第二和三季度出生者的受教育变差几乎都小于0。结合图1与表4，可以断言，在检验弱工具方面，中国的数据质量优于美国。⁶

（三）出生季度间的教育差异与工资差异

前面已经看到，第四季度出生者的确有显著的受教育优势，这个优势能转化为工资优势吗？图3报告了1950—1979年出生队列人口中，不同出生季度者的受教育状况与小时工资之间的关系。可以看出，受教育年限在各个人口队列内的变化，基本上与对数小时工资的变化相一致。绝大多数第四季度出生者的对数小时工资，都高于前三个季度出生者，而且，小时工资与受教育年限之间的变异方向都是一致的。这意味着，第四季度的受教育优势，正确地转化为市场收入优势了。当我们进一步观察大专及以上学历群体的小时工资时，发现季度之间不再有显著差异。⁷

由以上分析可以判断：就业政策设定的进入劳动力市场门槛为16周岁，导致出生季度影响受教育年限变异进而影响了小时工资的变异，尽管只有一小部分人受到出生季度的影响，由于样本巨大且数据质量高，还是可以看出，“出生季度 受教育优势 更高的工资”这个传导机制在中国城镇劳动力市场上清晰存在。

⁶ 1950—1979年人口队列中，高中毕业生所占比例从19%上升到48%，每个年龄队列的季度差异像受教育年限差异一样显著。结合图1可以比较中美两国人口队列的高中毕业生分别处于19%—48%和77%—95%+阶段。从AK91的图1-3看出，在其观察的1930—1959年出生队列样本中，1947—1959年出生队列人口的季度间教育变异已经不明显了，于是，AK91不再使用1950—1959队列。从数据的变异性和显著性判断，中国的数据质量优于美国。高中毕业生在人口队列中的详细分布信息，报告于附表2中。

⁷ 经过t检验，三个出生队列的“大专及以上学历”人口中，前三季度出生者与第四季度出生者之间，虚拟假设“工资差异不等于0”的t值分别为0.41、0.18和0.47。根据一位匿名审稿人的建议：如果出生季度对受教育年限的影响只会持续到高中阶段，那么，可以检验大专及以上学历群体中，第四季度出生者是否还有显著的受教育优势。考虑到各种社会因素的“干扰”，作者观察1990年、2000年和2005年数据的20岁和21岁队列中“第四季度出生的本专科在校生”的比例，然后与“高中及以上者”群体比较，发现“本专科在校生”中第四季度出生者的比例还略低于“高中及以上群体”，这说明出生季度对受教育的影响没有超过高中阶段。这两个结果可以排除工具变量与结构性方程的误差项相关，因此，工具的外生性是可信的。

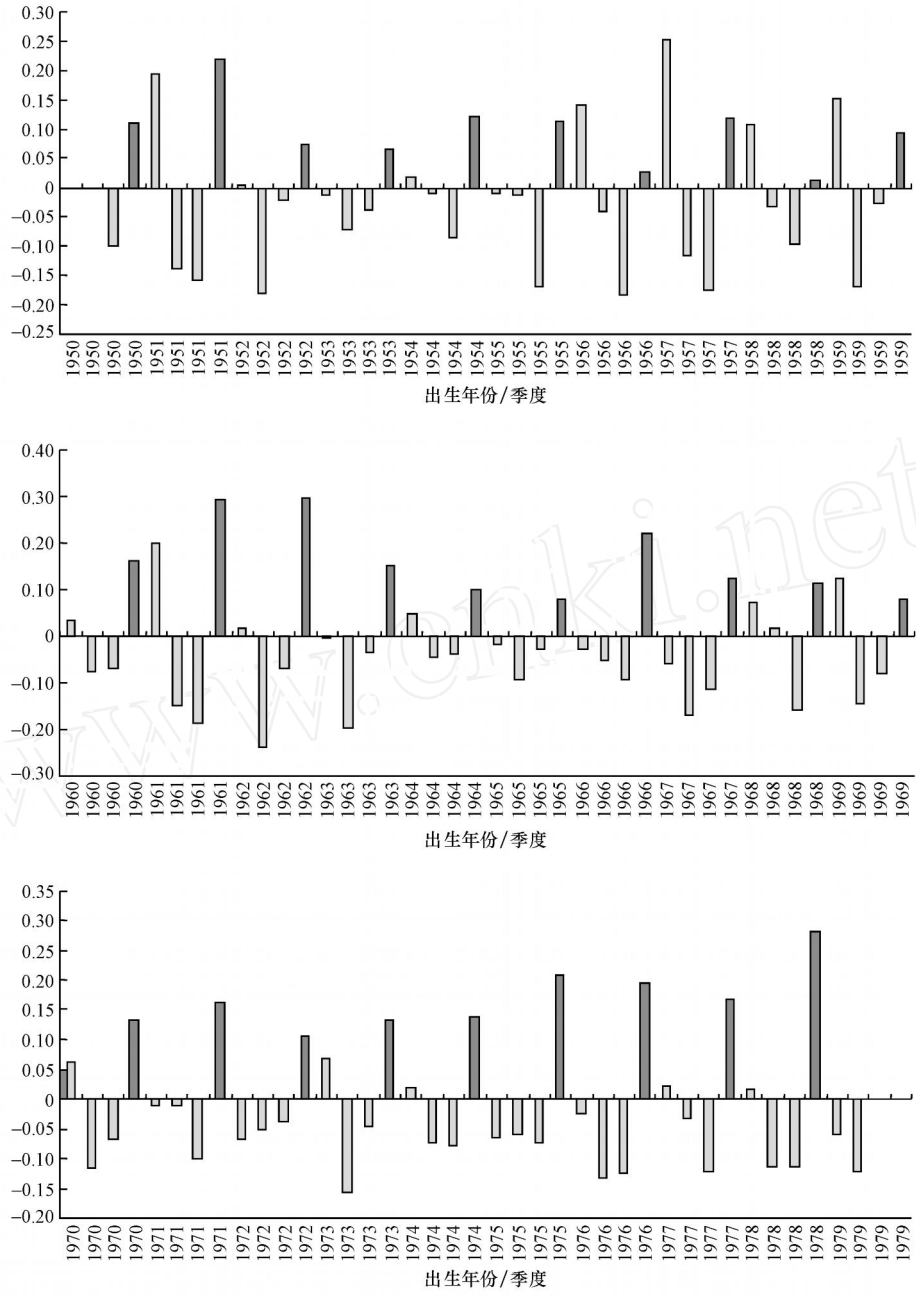


图 2 出生季度与从移动平均值 (+2, -2) 中得到的受教育年限偏差 (纵轴单位: 年)

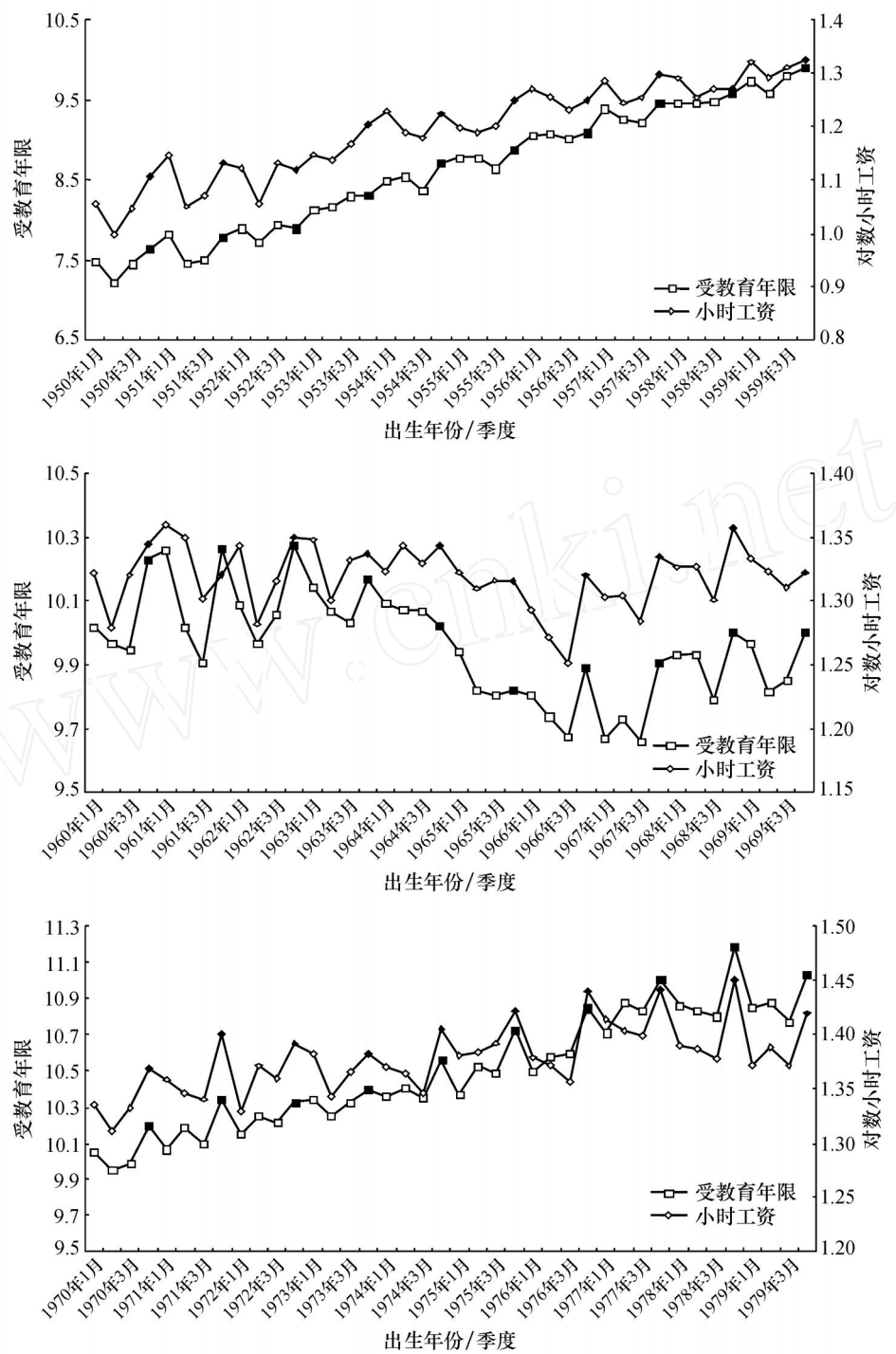


图 3 受教育水平和小时工资的关系识别 (实心黑点为第四季度观测值)

五、2SLS 估计结果及解释

由于 A K91 使用的经验方程可能存在误设, 本文对其建议进行必要的调整: A K91 没有直接使用出生季度作为工具变量, 而是用出生季度乘以出生年份形成交互项。在扩展模型里, 又用 50 个州与 3 个出生季度相乘产生交互项, 排除性工具数量最多达到 180 个。但约简方程中的工具变量并不是越多越好: 一方面, 一些交互项很可能是无效工具 (如冗余工具等); 另一方面, 过多的交互项会减少受教育年限在不同出生季度间的变异, 导致内生解释变量与工具变量之间的相关性下降, 产生弱工具问题。Donald and Newey (2001) 使用 A K91 的数据 (只观察 1930—1939 年出生队列样本), 分析了在存在众多有效工具的前提下, 如何选择工具变量的数量问题。在基础模型里, 本文把出生季度直接作为工具变量; 然后增加交互项, 比较基础模型与扩展模型间的估计系数和检验统计量。

本文的经验方程如下⁸:

$$E_i = X_i + \sum_c Y_{ic} + \sum_j Q_{ij} + \mu_i, \quad (1)$$

$$\ln W_i = X_i + \sum_c Y_{ic} + E_i + \mu_i, \quad (2)$$

其中 E_i 为第 i 个观测者的受教育年限; X_i 是影响受教育年限的一组向量; Q_{ij} 为一组虚拟变量, 指出生在第 j 季度 ($j = 1, 2, 3$), Y_{ic} 为一组虚拟变量, 代表是否出生在队列 c 中 ($c = 1, \dots, 10$), W_i 在此为小时工资 (A K91 采用周工资), μ_i 为结构性工资方程的残差。在扩展模型里, 可在方程 (1) 中增加 $\sum_c \sum_j Y_{ic} Q_{ij}$ 项。A K91 把所在州与出生季度和出生年份形成交互项, 因为不同州的义务教育法可能有不同的内容; 中国的立法权仅集中在中央政府, 任何省市都无权规定自己的“法定劳动年龄”, 使用这个交互项没有经验依据。

按照 A K91 建议的研究路线, 作者同时报告 OLS 估计和 2SLS 估计的结果并进行对比。控制性解释变量分别为“出生队列”、“劳动者性别”、“婚姻状况”及“所在省份”; 受教育年限为 (被怀疑的) 内生解释变量, 工具变量为“出生季度”, 以及“出生季度”与“出生年份”的交互项。由于出生季度为虚拟变量, 排除性工具的数量超过了内生解释变量, 存在过度识别问题。与 A K91 不同的是, 作者把男女都纳入考察范围, 而不是仅限定为男性。没有理论或者经验上的依据, 认为出生季度只影响某个性别而不影响另一个的教育变异。

⁸ 本文采用更简洁的 0—1 变量, 第四季度出生者为 1, 其他季度为 0; 将这个 0—1 变量乘以出生年份, 得到交互项。其经验依据是中国的升学时间 (9 月份), 更可能影响第四季度出生者。

由于“年龄-工资剖面”呈抛物线状，把“年龄”和“年龄平方”作为控制变量，可以控制工资变化的年龄效应，但出生队列与年龄之间高度共线性，本文的经验方程中先不加入年龄及年龄平方项，然后再加入这些变量，比较教育回报率是否发生变化。从附表3可看出，增加了“年龄”、“年龄平方”后并不改变教育回报率的估计值。三组人口队列中，都满足阶条件和秩条件，没有欠识别问题，在回归分析结果中不再报告LM检验统计量。在每个年龄组的回归分析中，都按照国家统计局提供的权数进行加权回归。在存在弱工具的条件下，2SLS可能存在估计偏差，一个必要的矫正是调整估计方法，本文采用LIML、FULLER和B2SLS估计。

表5报告了1950—1959年出生人口队列的回归分析结果。第(1)列为OLS估计结果，受教育年限的系数为0.108，有非常小的标准差。⁹第(2)列为2SLS估计，工具变量为0—1变量，第四季度定义为1，前三个季度为0。模型恰好识别，得到的受教育年限系数为0.153，标准差为0.034，显著高于OLS估计值。第(3)列将每个出生季度处理为虚拟变量，这时，排除性工具为3个，受教育年限的系数提高到0.159，标准差与(2)一样。第(4)列采用出生年份与出生季度(0—1)形成交互项，这时，排除性工具的数量为10个。受教育年限的系数为0.156，介于(2)和(3)之间，标准差略降到0.031。第(5)列继续增加工具变量的数量，将出生季度(3个)加上出生季度与出生年份的交互项(27个)，共30个排除性工具，这时，教育回报系数提高到0.161。但应该指出，在(3)的基础上增加交互项时，无法拒绝第二季度、第三季度分别与出生年份形成的交互项为冗余工具。

当怀疑有弱工具问题存在时，改进估计方法是一个可行的经验策略(Baum *et al.*, 2007; Hansen *et al.*, 2006)。第(6)列保留30个排除性工具，使用有限信息最大似然估计方法(LIML)，这时，受教育年限系数提高到0.186，比(5)高出约2.5个百分点，标准差也提高到0.037；第(7)采用FULLER估计¹⁰，受教育年限系数为0.184，标准差与(6)相同；第(8)列采用调整偏差的2SLS估计(B2SLS)，受教育年限的系数为0.185，标准差为0.037。第(6) — (8)列的结果几乎没有差异。从这个队列的不同估计结果看出，无论是采用广义矩估计还是似然估计，受教育年限的系数都显著高于OLS的估计值。

⁹ AK91在作2SLS估计之前还对比了OLS和Wald估计的结果，发现两个人口组中，估计系数之间的差异在统计上不显著。而中国三个人口组的Wald估计系数，分别为0.180、0.210和0.208，都显著高于OLS估计值，更接近于后面的2SLS和LIML的估计结果。

¹⁰ 当采用点估计时，Fuller估计的表现相当优越。Fuller估计一般采用 $\alpha = 1$ 或 $\alpha = 4$ 。前者的估计结果是接近无偏差的；后者的估计结果MSE最小。本文报告的结果为 $\alpha = 1$ 时的结果。如果使用“区间估计”，可以采用条件似然比方法——CLR，这个方法为Moreira(2003)以及Yogo(2004)所采用。CLR估计的结果与LIML、FULLER方法几乎完全一致，在此未报告。Donald and Newey(2001)还建议用偏差调整型两阶段最小二乘估计(B2SLS)，这是K级估计的一种特殊形式，其公式为 $k = T / (T - K_2 + 2)$ ，其中 K_2 为工具变量的数量。

表 5 回归分析结果(1) (1950—1959 年出生人口)

	OLS (1)	2SLS (2)	2SLS (3)	2SLS (4)	2SLS (5)	LIML (6)	FULL (7)	B2SLS (8)
受教育年限	0.108 (0.001)	0.153 (0.034)	0.159 (0.034)	0.156 (0.031)	0.161 (0.025)	0.186 (0.037)	0.184 (0.037)	0.185 (0.037)
男性(=1)	0.217 (0.005)	0.144 (0.055)	0.133 (0.055)	0.138 (0.050)	0.130 (0.040)	0.090 (0.061)	0.092 (0.060)	0.091 (0.061)
有配偶(=1)	0.062 (0.011)	0.021 (0.033)	0.015 (0.033)	0.017 (0.030)	0.013 (0.025)	- 0.009 (0.036)	- 0.008 (0.035)	- 0.009 (0.036)
排除性工具变量:		1	3	10	30	30	30	30
观测值		103 036	103 036	103 036	103 036	103 036	103 036	103 036
K-P 秩 Wald F 统计量		36.801	12.59	4.51	2.401	2.401	2.401	2.401
S-Y 临界值: 5%最大相对偏差		16.38 *	13.91	20.74	21.42	3.88	2.26	—
Hansen J 统计量			3.362	10.236	23.354	21.671	21.762	21.705
卡方 P 值			0.1862	0.3317	0.7601	0.8335	0.8299	0.8322
MSE		0.4816	0.4893	0.4858	0.4914	0.5287	0.5265	0.5278

注: * 此处为 Wald 检验目标参数最大偏差小于 0.1 时, 第一阶段 F 值的临界值。
未列示省份虚拟和出生年份虚拟的值, 括号内为异方差稳健标准误, 以下各表同。

由于弱工具是争论的焦点, 本文对此作进一步讨论。在 1950—1959 年出生人口数据中, 是否存在弱工具呢? 根据 Staiger and Stock (1997) 建议的经验法则, 只有 1 个内生变量时, 第一阶段回归的 F 值大于 10 是个经验切割点。表 5 第 (2) 列显示的 F 统计量为 36.8, 远远超过了 10; 当排除性工具数量为 3 时, 第一阶段的 F 统计量为 12.6, 仍超过了 10。可以初步判断, 出生季度不是“弱工具”。当经验方程中工具变量的数量不多时, 这个经验法则更适用, 随着工具数目的增加, F 值大于 10 这个经验法则就会出现较大的偏差。后来, Stock-Yogo (2004) 建议了更加准确的识别办法, 衡量 2SLS 估计出现最大相对偏差的临界值: 当只有 1 个内生解释变量和 3 个排除性工具时, 2SLS 估计系数的偏差不超过 OLS 估计偏差的 5% 时的 (F) 临界值为 13.91, 不超过 10% 时的临界值为 9.08, 这样一来, 第 (3) 列的 F 值可以拒绝工具变量偏差超过 10%, 但不能拒绝偏差超过 5%。可以说, 这个偏差已经很小了。

第 (4) 列的排除性工具变量增加到 10 个, 第一阶段的 F 统计量则下降到 4.51, Stock-Yogo 建议的最大偏差不超过 20% 的 F 值为 6.61, 这时, 弱工具的风险随之增高了。随着工具变量的数量增加到 30 个, 第 (5) 列的 F 统计量继续下降到 2.4, 按照 Stock and Watson (2006) 建议的测量 2SLS 估计偏差的方法, F 值接近 2 时与 OLS 估计的偏差会趋同, 由于 OLS 估计有更小的标准差, 这时, 2SLS 估计已没有任何优势。既然用 30 个排除性工具会导致弱工具问题, 第 (6) 列改用 LIML 方法来估计, 这时, 尽管 F 统计量不变, 但 Stock-Yogo 建议的临界值却大幅度下降, 所以, LIML 的估计偏差大为减少。Hansen *et al.* (2006) 发现, 在使用众多工具的条件下, 利用 Fuller

(1977) 建议的估计方法也能实现类似的改进，第 (7) 列报告了 FULLER 改进后的结果，其 5 % 临界值只有 2.26，这意味着，所估计系数的最大相对偏差不超过 5 %。

在第 (3) — (8) 列中，工具变量数量超过了内生解释变量，但过度识别检验的 J 统计量都不显著，不存在过度识别问题，即无法拒绝方程中的工具变量都是有效的。不难发现，在 2SLS 估计 (第 2—5 列) 中，随着排除性工具的增加，第一阶段的 F 统计量不断下降。由此推断，AK91 之所以出现弱工具问题，除了数据本身缺少出生季度间的教育变异外，另一个原因是使用了过多的工具变量 (模型误设)，其他经济学家的研究结论中也证实了这一点 (Imbens, 2006; Hahn and Hausman, 2003)。在四个工具数量组合下，哪个 (2SLS) 方程最适宜呢？Donald and Newey (2001) 建议选择结构性方程中 MSE 最小的那个，可以看出，第 (2) 列的 MSE 最小，这时，教育回报系数为 0.153，而且可以非常确信地拒绝弱工具。

表 6 报告了 1960—1969 年出生队列中各模型的回归结果，每列的解释变量和工具变量组合与表 5 一致。第 (1) 列的 OLS 估计结果显示，受教育年限的系数为 0.12，统计上显著。第 (2) 列采用 2SLS 估计方法，排除性工具为 1 个，受教育年限的系数为 0.159，这时，第一阶段的 F 统计量为 49，显然为强工具；当使用的排除性工具数量为 3 个时 (第 3 列)，第一阶段的 F 统计量下降到 19.03，大于经验切割点 10，也大于 Stock-Yogo 提出的 IV 最大相对偏差 5 % 的临界值 13.91，所以，仍可以非常确信地拒绝弱工具。

表 6 回归分析结果 (2) (1960—1969 年出生人口)

	OLS (1)	2SLS (2)	2SLS (3)	2SLS (4)	2SLS (5)	LIML (6)	FULL (7)	B2SLS (8)
受教育年限	0.120 (0.001)	0.159 (0.030)	0.168 (0.028)	0.158 (0.024)	0.155 (0.021)	0.187 (0.042)	0.186 (0.041)	0.187 (0.042)
男性 (= 1)	0.266 (0.003)	0.238 (0.022)	0.232 (0.020)	0.239 (0.018)	0.241 (0.016)	0.218 (0.030)	0.219 (0.030)	0.218 (0.030)
有配偶 (= 1)	0.033 (0.008)	0.020 (0.013)	0.017 (0.013)	0.020 (0.012)	0.021 (0.011)	0.011 (0.016)	0.011 (0.016)	0.0105 (0.016)
工具变量数:		1	3	10	30	30	30	30
观测值	179 520	179 520	179 520	179 520	179 520	179 520	179 520	179 520
K-P 秩 Wald F 统计量		49.04	19.03	7.46	3.29	3.29	3.29	3.29
S-Y 临界值: 5 % 最大相对偏差		16.38 [*]	13.91	20.74	21.42	3.88	2.26	—
Hansen J 统计量			0.634	15.148	46.809	43.436	43.562	43.407
卡方 P 值			0.7285	0.087	0.0195	0.0415	0.0404	0.0417
MSE		0.4233	0.4306	0.4225	0.4212	0.4523	0.4509	0.4525

注：* 此处为 Wald 检验目标参数最大偏差小于 0.1 时，第一阶段 F 值的临界值。

第 (4) 列增加工具变量的数量达到 10 个，受教育年限的系数与 (2) 基本一致，标准差略有下降，但第一阶段的 F 统计量迅速下降到 7.46，小于经验切割点 10，而且，小于 Stock-Yogo 提出的 IV 最大相对偏差 20 % 的临界值

11.49,不能拒绝弱工具。但我们注意到,这时 J 检验的 p 值为 0.087,意味着模型可能存在过度识别。

第(5)列的结果显示,排除性工具变量增加到 30 个时,教育年限的系数与(2) — (4)几乎没有区别,但第一阶段的 F 统计量下降到 3.29,弱工具问题非常明显,但模型此时面临的过度识别问题也更严重,即一些工具变量可能不是合适的工具,应该考虑使用较少工具时的回归结果。

在模型中使用 30 个排除性工具的条件下,采用 LIML 得到的系数为 0.187, FULLER 估计的系数为 0.186, B2SLS 估计的系数为 0.187,估计值和标准差几乎完全一致。在 1960—1969 年出生队列中,第(2) — (5)列的教育系数都很接近,因为第(4) — (5)列存在过度识别问题,那么,在剩下的模型里,第(2)列的 MSE 值最小,受教育年限的系数为 0.159。

在 1960—1969 年出生队列中,2SLS、LIML、FULLER 和 B2SLS 估计得到的受教育年限系数,都显著高于 OLS 估计值,改进估计方法后的系数都高于 2SLS 估计值。

表 7 报告了同样模型设定形式下的 1970—1979 年出生队列的回归结果。OLS 估计的教育系数为 0.122,与前面的出生队列接近。从第(2) — (5)列的结果看出,依次增加工具变量的数量时,受教育年限系数几乎没有差异,都在 0.14 左右。虽然大于 OLS 的估计值 0.122,但却略小于前面出生队列的估计值。第一阶段的 F 统计量同样随着工具数量的增加而下降,但当排除性工具增加到 10 时, F 值为 18.77,仍然高于经验切割点 10,低于 Stock-Yogo 建议的最大相对偏差 5%临界值 20.74。当排除性工具达到 30 个时,2SLS 估计的第一阶段 F 值下降到 7.138,弱工具风险升高。采用 LIML、FULLER 和

表 7 回归分析结果(3) (1970—1979 年出生人口)

	OLS (1)	2SLS (2)	2SLS (3)	2SLS (4)	2SLS (5)	LIML (6)	FULL (7)	B2SLS (8)
受教育年限	0.122 (0.001)	0.142 (0.016)	0.139 (0.015)	0.143 (0.015)	0.140 (0.014)	0.142 (0.016)	0.142 (0.016)	0.142 (0.016)
男性(=1)	0.244 (0.003)	0.241 (0.004)	0.241 (0.004)	0.241 (0.004)	0.241 (0.004)	0.241 (0.004)	0.241 (0.004)	0.241 (0.004)
有配偶(=1)	-0.014 (0.005)	0.001 (0.012)	-0.001 (0.012)	0.002 (0.012)	0.000 (0.011)	0.001 (0.013)	0.001 (0.013)	0.001 (0.013)
排除性工具变量:		1	3	10	30	30	30	30
观测值		179 885	179 885	179 885	179 885	179 885	179 885	179 885
K-P 秩 Wald F 统计量		168.076	57.308	18.77	7.138	7.138	7.138	7.138
Stock-Yogo 检验临界值: 5%最大相对偏差		16.38 [*]	13.91	20.74	21.42	3.88	2.26	—
Hansen J 统计量			2.979	2.879	23.059	23.007	23.009	23.037
卡方 P 值			0.2255	0.9689	0.7738	0.7762	0.7761	0.7749
MSE		0.3665	0.3655	0.3670	0.3660	0.3668	0.3668	0.3668

注: * 此处为 Wald 检验目标参数最大偏差小于 0.1 时,第一阶段 F 值的临界值。

B2SLS 方法得到的系数都是 0.142，与 2SLS 方法得到的结果几乎没有差异。在所有的 2SLS 估计方程中， J 检验结果显示，不存在过度识别问题。通过比较，发现第（3）列的 MSE 数值最小，这时，排除性工具数量为 3 个，受教育年限的系数为 0.139。

把表 5—表 7 的结果联系起来，证实了本文的假说：数据质量改善会降低弱工具风险。在每一个对应模型中，1970—1979 年人口组的第一阶段 F 统计量都大于 1960—1969 年人口组的值，而 1960—1969 年人口组的 F 统计量又大于 1950—1959 年人口组的值。这验证了本文的推论：当高中毕业生在各个人口队列中的比例渐次提高并向 50% 逼近时，出生季度间的教育变异会逐渐增大从而减轻弱工具风险。¹¹ 模型中增加了“年龄”和“年龄平方”以及“城市规模”等控制性变量时，各列的第一阶段回归的 F 统计量有所下降，但受教育年限系数几乎没有变化，而且，第一阶段回归的 F 统计量也同样随着人口队列的演进而渐次提高（见附表 3）。回顾一下 Staiger and Stock（1997）在表 2 中报告的美国 1940—1949 年出生队列的回归结果，更印证了本文的论断：美国的数据质量越来越差，即使改进估计方法也无法得到稳健一致的系数。与美国 1940—1949 年出生队列的估计值剧烈波动相比，中国这三组队列的估计结果都是稳健、一致且与经验相吻合的。因此，本文可以确信地断言：中国的数据质量高于美国，“出生季度”在中国是个强工具变量，尤其是在 1970—1979 年队列中。

本文另一目的是弄清中国城镇劳动力市场上的教育回报率。如果工具变量足够弱，那么，2SLS 估计结果的偏差就会与 OLS 偏差趋同，前文看到，在 1950—1969 年队列中，即使在工具变量达到 30 个时，2SLS 估计结果似乎更接近 LIML 等估计方式的结果，而没有与 OLS 结果趋同。2SLS 和其他估计方式的结果存在明显的差异。当工具变量都有效时，无论是采用 2SLS 估计还是其他估计形式，都需要评估多少工具数量才是最优的经验方程，因此，本文把不同估计形式下、不同工具数量组合的 MSE 数值，报告在表 8 中。然后根据最小化的 MSE 选择最适宜的经验方程，报告了三个人口队列组中的受教育系数，见各个人口队列组窗格下面的“受教育年限系数”行。显然，在同一人口队列组中，不同方法得出的估计系数都是非常接近的；在三个人口队列组之间，教育收益率稳定在 13.9%—15.9%。

¹¹ 依照一位审稿人的建议，作者还分别比较了“三个入口队列和不同工具数量”下的“A-R Wald 检验 F 值”、“A-R Wald 检验卡方值”和“Stock-Wright LM S 统计量卡方值”，发现在每种工具变量组合下，随着人口队列演进，这三个指标都在持续增大。

表 8 不同工具数量时的 MSE 及最优化工具时的教育收益率

工具变量组合	工具数量	MSE _{2SLS}	MSE _{LIML}	MSE _{B2SLS}	MSE _{FULL}
1950—1959					
Q1	1	0.4816	0.4816	0.4816	0.4807
Q3	3	0.4893	0.4956	0.4953	0.4939
Q1 * Y	10	0.4858	0.5052	0.5048	0.5031
Q3 * Y + Q3	30	0.4911	0.5287	0.5278	0.5265
受教育年限系数	最优 \hat{K}	0.1527	0.1527	0.1527	0.1517
		(0.0338)	(0.0338)	(0.0338)	(0.0331)
1960—1969					
Q1	1	0.4233	0.4233	0.4233	0.4229
Q3	3	0.4306	0.4311	0.4306	0.4305
Q1 * Y	10	0.4225	0.4305	0.4306	0.4299
Q3 * Y + Q3	30	0.4209	0.4523	0.4525	0.4509
受教育年限系数	最优 \hat{K}	0.1555	0.1588	0.1588	0.1582
		(0.0213)	(0.0300)	(0.0300)	(0.0295)
1970—1979					
Q1	1	0.3665	0.3665	0.3665	0.3665
Q3	3	0.3655	0.3657	0.3657	0.3657
Q1 * Y	10	0.3670	0.3671	0.3671	0.3671
Q3 * Y + Q3	30	0.3660	0.3668	0.3668	0.3668
受教育年限系数	最优 \hat{K}	0.1387	0.1389	0.1389	0.1389
		(0.0153)	(0.0156)	(0.0156)	(0.0155)

六、结 论

“给我一个立足之处和一根足够长的杠杆，我就可以撬动地球”。阿基米德在自然界永远也不可能找到的杠杆，在经济学家手中成为现实：在分析观测性数据时，工具变量成为强有力的杠杆。但经济学家也必须清楚：他既有立足之处——有效工具，其杠杆又足够长——强工具。今天，经济计量学家在弱工具领域的不懈探索和所取得的最新成就，使我们对驱散笼罩在工具变量分析头顶的不确定性阴云逐渐增强了信心。

本文的研究显示，虽然中美之间有着不同的社会文化背景，但“出生季度不同 受教育年限差异 工资差异”这个传导机制在中国劳动力市场上同样清晰存在。在美国，由于普及了高中阶段的教育，不同出生季度者之间很少有受教育年限的变异，出生季度因而成为一个弱工具；中国因尚未普及高中教育，不同的出生季度带来了更大的受教育年限变异，因而成为一个强工具。从数据角度看，是否为弱工具直接取决于高中升学率高低，而背后反映的却是一个国家所处的发展阶段或国民受教育水平。作者推测，其他处于相同发展阶段的国家，出生季度可能同样是个强工具。相反，在那些普及了高

中阶段教育的其他发达国家，出生季度可能也同样是弱工具。

A K91 使用 2SLS 得到的估计系数与 OLS 没有显著差异，但样本更少的中国数据却得出了完全不同的结论：（1）在所有人口队列中，2SLS 的估计结果都显著高于 OLS 估计值，改进估计方法后差距似乎更大，这意味着，OLS 估计可能严重低估了劳动力市场上的教育收益率。（2）由于中国所处的发展阶段和特殊的入学考试制度，高中入学率徘徊在 20 %—50 % 之间，在此范围内，出生季度对升入高中的决策有显著影响，从本文的分析结果看，可以排除弱工具，至少与美国相比是强工具。（3）随着由远而近出生队列高中入学率日益提高（趋于 50 %），出生季度作工具变量时第一阶段回归中的 *F* 统计量也持续提高，可以在非常高的置信水平上拒绝弱工具假设。

对经验分析来说，两个最重要的条件是“数据质量”和“模型设定”（Angrist *et al.*，2006）。由于美国 1940 年以后出生队列的受教育水平进一步提高，“数据质量”变得更差，即使改进模型设定和改善估计方法，数据质量问题仍无法解决，后来文献显示，经济计量学家不约而同都抛弃了 1940 年以后出生队列的数据。本文使用中国人口数据并得出了不同的研究结论，使得本研究区别于已有文献：“出生季度”本身并不是一个弱工具，模型设定和数据质量才使得“出生季度”在美国变成了弱工具。

为什么要格外关注“数据质量”？相信本文给出了一个回答：“数据质量”是经验分析中无法替代的关键因素，从事经验分析的经济学家都应关注这个问题，只有对经验数据不断强调、重视和应用，经验分析才可能在中国经济科学研究中真正成为主流，才可能阻止更多的学者跌入“李嘉图恶习”的陷阱中——用复杂的数学推导来遮掩“澡盆中没有婴儿”¹²的尴尬。这对推动中国经济学研究来说，尤具现实意义。

附 表

附表 1 在校率差异 (Diff- In-Diff 结果)			单位：%
出生年份	出生季度		差异 (1) - (2)
	前三季度 (1)	第四季度 (2)	
1990 年			
1973	0.582	0.648	- 0.066
(1)	(0.003)	(0.004)	(0.005)
1974	0.718	0.770	- 0.051
(2)	(0.003)	(0.004)	(0.005)
差异	- 0.136	- 0.122	- 0.014
(1) - (2)	(0.004)	(0.006)	(0.007)

¹² 无论经济学 (Economics) 还是经济计量学 (Econometrics)，最终关注的永远都是经济事实或经济现象 (Econ)，这才是“澡盆中的婴儿”。

(续表)

出生年份	出生季度		差异 (1) - (2)
	前三季度 (1)	第四季度 (2)	
	2005 年		
1989	0.727	0.777	- 0.050
(1)	(0.002)	(0.003)	(0.004)
1990	0.860	0.891	- 0.031
(2)	(0.002)	(0.003)	(0.003)
差异	- 0.133	- 0.114	- 0.019
(1) - (2)	(0.003)	(0.004)	(0.0054)

资料来源:1990 年人口普查数据和 2005 年 1 %人口普查数据。

附表 2 高中毕业及以上者在不同出生队列中的比例

单位: %

出生年份	前三季度	第四季度	差异	出生年份	前三季度	第四季度	差异
1950	0.191	0.197	- 0.006	1965	0.349	0.351	- 0.002
	(0.004)	(0.007)	(0.008)		(0.004)	(0.006)	(0.007)
1951	0.199	0.212	- 0.013	1966	0.329	0.353	- 0.025
	(0.004)	(0.007)	(0.008)		(0.004)	(0.006)	(0.007)
1952	0.196	0.213	- 0.017	1967	0.319	0.352	- 0.033
	(0.004)	(0.006)	(0.007)		(0.004)	(0.006)	(0.007)
1953	0.225	0.247	- 0.023	1968	0.353	0.369	- 0.016
	(0.004)	(0.006)	(0.007)		(0.003)	(0.006)	(0.006)
1954	0.264	0.294	- 0.030	1969	0.358	0.370	- 0.012
	(0.004)	(0.006)	(0.007)		(0.004)	(0.006)	(0.007)
1955	0.302	0.340	- 0.039	1970	0.368	0.394	- 0.026
	(0.004)	(0.007)	(0.008)		(0.003)	(0.006)	(0.007)
1956	0.345	0.360	- 0.014	1971	0.379	0.402	- 0.023
	(0.004)	(0.007)	(0.008)		(0.004)	(0.006)	(0.007)
1957	0.381	0.407	- 0.026	1972	0.389	0.414	- 0.025
	(0.004)	(0.007)	(0.008)		(0.004)	(0.006)	(0.007)
1958	0.409	0.414	- 0.005	1973	0.405	0.424	- 0.019
	(0.004)	(0.007)	(0.008)		(0.004)	(0.006)	(0.007)
1959	0.442	0.471	- 0.029	1974	0.403	0.424	- 0.021
	(0.005)	(0.008)	(0.009)		(0.004)	(0.006)	(0.007)
1960	0.465	0.508	- 0.043	1975	0.421	0.457	- 0.036
	(0.004)	(0.008)	(0.010)		(0.004)	(0.006)	(0.007)
1961	0.483	0.495	- 0.013	1976	0.436	0.482	- 0.045
	(0.005)	(0.008)	(0.010)		(0.004)	(0.007)	(0.008)
1962	0.443	0.463	- 0.020	1977	0.471	0.498	- 0.027
	(0.004)	(0.006)	(0.007)		(0.004)	(0.007)	(0.008)
1963	0.418	0.427	- 0.009	1978	0.483	0.525	- 0.042
	(0.003)	(0.006)	(0.007)		(0.004)	(0.007)	(0.008)
1964	0.395	0.386	0.009	1979	0.478	0.529	- 0.050
	(0.004)	(0.006)	(0.007)		(0.004)	(0.007)	(0.008)

注:括号内为标准差。

附表 3 不同出生队列的回归结果和相关信息

	OLS	2SLS	2SLS	2SLS	2SLS	LIML	FULL	B2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1950—1959	0.101	0.149	0.156	0.149	0.153	0.186	0.184	0.186
受教育年限	(0.001)	(0.044)	(0.043)	(0.039)	(0.029)	(0.050)	(0.049)	(0.051)
K-P F 统计量		22.345	7.789	2.886	1.75	1.75	1.75	1.75
MSE	0.4451	0.4707	0.4782	0.4709	0.4743	0.5249	0.5210	0.5264
1960—1969	0.113	0.139	0.145	0.131	0.126	0.159	0.157	0.160
受教育年限	(0.001)	(0.042)	(0.040)	(0.031)	(0.026)	(0.092)	(0.088)	(0.094)
K-P F 统计量		24.333	9.39	4.809	2.261	2.261	2.261	2.39
MSE	0.3975	0.4032	0.4063	0.4002	0.3989	0.4165	0.4150	0.4574
1970—1979	0.116	0.150	0.144	0.151	0.143	0.149	0.149	0.149
受教育年限	(0.001)	(0.024)	(0.023)	(0.022)	(0.019)	(0.024)	(0.024)	(0.024)
K-P F 统计量		70.683	25.838	8.296	3.763	3.763	3.763	3.76
MSE	0.3514	0.3614	0.3580	0.3619	0.3576	0.3608	0.3607	0.3608

参 考 文 献

[1] Angrist, J. , and A. Krueger, “Does Compulsory School Attendance Affect Schooling and Earnings?” *Quarterly Journal of Economics*, 1991, 106(4), 979—1014.

[2] Angrist, J. , and A. Krueger, “Empirical Strategies in Labor Economics”, Chapter 23 in Ashenfelter, O. , and D. Card (ed.) , *Handbook of Labor Economics*, volume 3. Elsevier, 1999, 1277—1366.

[3] Angrist, J. , and A. Krueger, “Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments”, *Journal of Economic Perspectives*, 2001, 15(4), 69—85.

[4] Angrist, J. , D. Blau, A. Falk, J. Robin, and C. Taber, “How to Do Empirical Economics”, *Investigaciones Económicas*, 2006, 30(2), 179—206.

[5] Baum, C. , M. Schaffer, and S. Stillman, “Enhanced Routines for Instrumental Variables/ GMM Estimation and Testing”, Boston College Economics Working Paper No. 667, 2007.

[6] Bound, J. , D. Jaeger, and R. Baker, “Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak”, *Journal of the American Statistical Association*, 1995, 90, 443—450.

[7] Donald, S. , and W. Newey, “Choosing the Number of Instruments”, *Econometrica*, 2001, 69(5), 1161—1191.

[8] Fuller, W. , “Some Properties of a Modification of the Limited Information Estimator”, *Econometrica*, 1977, 45(4), 939—953.

[9] Hahn, J. , and J. Hausman, “A New Specification Test for the Validity of Instrumental Variables”, *Econometrica*, 2002, 70(1), 163—189.

[10] Hahn, J. , and J. Hausman, “Weak Instruments: Diagnosis and Cures in Empirical Econometrics”, *American Economics Review*, 2003, 93(2), 118—125.

[11] Hansen, C. , J. Hausman, and W. Newey, “Estimation with Many Instrumental Variables”, Memo, 2006.

- [12] Imbens, G., "Endogeneity 2: Two Stage Least Squares, Control Function, and Limited Information Maximum Estimation", Lecture 18, ARE213 Spring '06, UC Berkeley, 2006.
- [13] Kleibergen, F., and R. Paapm, "Generalized Reduced Rank Tests Using the Singular Value Decomposition", *Journal of Econometrics*, 2006, 127(1), 97—126.
- [14] 李学武, "成长於新世界诞生之初——1950 至 1970 年代少儿读本中‘成长’模式考察", 《二十一世纪》网络版, 2003 年 4 月。
- [15] Lleras-Muney, A., "Were Compulsory Attendance and Child Labor Laws Effective? An Analysis from 1915 to 1939", *Journal of Law and Economics*, 2002, 45(2), 401—435.
- [16] Moreira, M., "A Conditional Likelihood Ratio Test for Structural Models", *Econometrica*, 2003, 71(4), 1027—1048.
- [17] Murry, M., "Avoiding Invalid Instruments and Coping with Weak Instruments", *Journal of Economic Perspectives*, 2006, 20(4), 111—132.
- [18] Staiger, D., and J. Stock, "Instrumental Variables Regression with Weak Instruments", *Econometrica*, 1997, 65(3), 557—586.
- [19] Stock, J., and M. Yogo, "Testing for Weak Instruments in Linear IV Regression", NBER Technical Working Paper 284, 2001. (revised version: January 2004)
- [20] Stock, J., and M. Watson, *Introduction to Econometrics*. Addison-Wesley, 2006.
- [21] Wooldridge, J., *Introductory Econometrics: A Modern Approach*. South-Western College Pub, 2005.
- [22] Yogo, M., "Estimating the Elasticity of Intertemporal Substitution when Instruments Are Weak", *Review of Economics and Statistics*, 2004, 86(3), 797—810.

Searching for the Archimedes' Lever : Is Quarter-of-Birth Really A Weak Instrumental Variable ?

YAOWU WU

(Chinese Academy of Social Sciences)

Abstract Using the 1 % population sample survey data newly collected in 2005, this paper answers a theoretical question that attracts labor economists and micro-econometricians: is quarter-of-birth really a weak instrumental variable? In developed countries where very few individuals do not finish high school, the effect of quarter-of-birth on the variation of education is very small. Thus, the problem of quarter-of-birth being a weak instrument appears. However, in developing countries, more than half of the population do not enter high school. As a result, quarter-of-birth does significantly affect the variation in educational attainments and serves as a strong instrumental variable.

Key Words Instrumental Variable, Weak Instruments, Return to Education

JEL Classification J01, J24, C15, C31