

# 中国为什么会有如此多的留守儿童

## ——来自公立学校的非政策性户籍歧视的证据

王智波 李长洪\*

**内容提要** 本文试图解释为什么在政府大力解决随迁子女平等入学的背景下,仍有超过60%的进城务工人员选择让处于义务教育阶段的孩子留守农村。本文提出,“来自公立学校的非政策性户籍歧视”是导致留守儿童大规模出现的一个原因。在实证中,本文以外地户籍学生是否有为入学(且非“择校”)而“找关系”作为非政策性户籍歧视存在与否的代替变量,发现“找关系”显著提高了城镇公立初中招收外地户籍学生比例,即使分别考虑因遗漏变量、双向因果关系和测量误差可能导致的估计偏误,这一结论仍稳健存在。在控制了小学入学年龄和留级等因素之后,无论外地户籍学生是否“找关系”,他们“超龄”的概率都显著高于本地学生。这两个发现表明城镇公立学校存在非政策性户籍歧视,并且对外地户籍学生入学造成了实质性的阻碍。因此,进城务工人员不得不选择让他们的子女留守农村。

**关键词** 留守儿童 “找关系” 户籍歧视 CEPS

### 一 引言

工业化和城市化的本质是劳动力在产业间和空间的重新配置,然而,囿于现有户

\* 王智波,华南师范大学经济与管理学院,电子邮箱:wangzhibo@m.scnu.edu.cn;李长洪(通讯作者),厦门大学王亚南经济研究院,电子邮箱:changhongLeo@126.com。作者感谢广东省自然科学基金(2016A030313429)的资助。本文使用的数据来自中国人民大学中国调查与数据中心设计与实施的中国教育追踪调查(CEPS)项目,该调查的项目负责人是中国人民大学的王卫东博士,作者感谢上述机构及人员提供数据协助,文责自负。

籍制度，与中国工业化和城市化相伴的是大规模的“人户分离”现象。截至2014年，全国“人户分离”人口为2.98亿人，占总人口的21.78%（国家统计局，2015）。同时，处于义务教育阶段的随迁子女和留守儿童规模也不断增长。据教育部统计，2010—2013年全国进城务工人员家庭中处于义务教育阶段的随迁子女数量分别约为1167万、1261万、1394万和1277万；农村留守儿童数量分别约为2272万、2200万、2271万和2127万。也就是说，超过60%的外来人员家庭选择让处于义务教育阶段的子女留守农村。

党和政府历来关心外来务工人员随迁子女的教育问题。针对处于义务教育阶段的随迁子女，中央政府在2001年取消了“限制流出，交费借读”的规定，提出“以流入地政府管理为主，以公办中小学为主”（以下简称“两为主”）政策。财政上，中央政府推行按进城务工人员随迁子女在校人数拨付教育经费。2010—2012年，中央财政安排城市义务教育免学杂费和进城务工人员随迁子女接受义务教育奖励资金224亿元。2012年，全国共有1393.87万以上进城务工人员随迁子女在城市接受义务教育，其中80.2%在公办学校就读。与此同时，政府购买的民办学位数量不断增加，2014年达124.6万个，约占随迁子女人数的10%。

本文关注的问题是，在政府大力解决农民工随迁子女平等接受义务教育的背景下，为什么有超过60%的进城务工人员仍选择让自己的孩子留守农村。已有研究表明，进城务工人员与其子女的地域分离，不仅影响家庭功能的正常发挥，而且不利于家庭成员生理、心理和社会性等方面的健康发展（李强、臧文斌，2011；叶静怡等，2017）。留守儿童由于缺少父母的日常陪伴和关爱，以及有效监护和沟通，容易产生孤独、焦虑、自卑等心理问题（吴霓，2004；谭深，2011；白勤等，2012），从而导致代际间的亲情疏离、价值观对立、甚至家庭内部的严重撕裂（陈雯，2014）。叶敬忠和Murray（2005）报告，49.7%的留守儿童表示想和外出打工的父母在城市共同生活。在这一背景下，研究是什么原因导致了留守儿童大规模出现具有重要的现实意义。

目前，鲜有文献讨论处于义务教育阶段的留守儿童形成原因。在我们找到的文献中，仅有叶静怡等（2015）涉及了这一问题。他们先验地认为居住条件和异地升学限制是影响农民工家庭子女随迁决策的重要影响因素，不过，他们并没有给出相关证据。

本文提出的假说是，“来自公立学校的非政策性户籍歧视”是留守儿童大规模出现的一个原因。所谓“非政策性户籍歧视”是指公立学校在招收符合政策的外地户

籍学生过程中,存在故意刁难或“吃拿卡要”行为。与迁入地政府通过设定高门槛进行“政策性户籍歧视”不同,“非政策性户籍歧视”因没有具体规定和明确标准而无法观测。本文的识别策略是:首先,以外地户籍学生是否有为入学(且非“择校”)而“找关系”为代替变量,通过检验“找关系”能否提高外地户籍学生的入学比例,来推断城镇公立学校是否存在非政策性户籍歧视;其次,通过检验外地户籍学生现就读年级的年龄为“超龄”的概率在统计上是否显著高于本地学生,来推断“非政策性户籍歧视”是否对外地户籍学生入学造成了实质性的负向影响。本文发现,在给定劳动力市场上存在户籍歧视、政策性户籍歧视条件下,“非政策性户籍歧视”在统计上显著存在,并且是阻碍留守儿童来父母身边生活和读书的一个重要原因。

本文余下的结构安排如下:第二部分为分析框架;第三部分为数据来源、模型建立和识别策略;第四部分为实证结果分析;第五部分为结论。

## 二 分析框架

先验地看,导致外来务工人员选择让自己处于义务教育阶段的子女留守农村的因素可能有5个:第一,城市的义务教育资源不足;第二,农村的义务教育资源足够好;第三,外来务工人员缺乏必要的财力和居住条件;第四,政策性的户籍歧视;第五,非政策性的户籍歧视。前四种影响因素在文献上或有涉及,第五种则是本文提出的观点。

现有的文献和媒体报道帮我们排除了第一种和第二种可能。《中国流动儿童数据报告——2014》(以下简称《报告》)根据人口普查数据指出,2000-2010年间,全国处于义务教育阶段的随迁儿童和留守儿童规模从1982万上升到3581万,增长80%;同期城市义务教育阶段在校生人数仅从2852万上升到2879万,增长1.6%;小学学校数从32154所下降至16400所,下降49%;初中学校数从12723所下降至7283所,下降43%。《报告》详细调查了北京、上海和广州的情况,发现这三个城市义务教育阶段在校生人数均大幅低于上世纪90年代在校人数。其中,北京市2012年在校小学生为72万人,比历史最高水平(102万)低29%;上海市2012年在校小学生为76万人,比历史最高水平(114万)低33%;广州市在校小学生为82万人,比历史最高水平(89万)低8%。据此,《报告》的结论是城市具备增加义务教育资源(学位数)供给的能力,在供给水平恢复到历史高点之前,没有充分的理由说学位短缺是城市义务教育资

源不足导致的。数量众多的文献发现外地户籍劳动力在流入地劳动力市场受到“歧视”的证据（Chan, 1996; Knight et al., 1999; Meng, 2000; Meng & Zhang, 2001; Liu, 2005; Sicular et al., 2007; 章莉等, 2014），这暗示流动人员可能缺乏足够的财力让处于义务教育阶段的子女留在自己身边学习和生活。叶静怡等（2015）明确指出，居住条件和异地升学限制是影响农民工家庭子女随迁决策的重要影响因素。有媒体报道，随迁儿童在迁入地接受义务教育时受到“政策性的歧视”，如《半月谈》关注非京籍儿童北京念书需办 28 个证<sup>①</sup>。

基于上述分析，我们认为外来务工人员收入偏低、居住条件不佳和政策性歧视会影响进城务工人员决定子女是否随迁，然而，这些因素似乎还不足以解释为什么大多数外来务工人员选择让子女留守农村。原因如下：第一，最低工资制度在理论上保障了外来务工人员的收入能够达到维持其家庭成员在城市生活的基本要求。第二，《中华人民共和国义务教育法》（2006 年）规定“父母或者其他法定监护人在非户籍所在地工作或者居住的适龄儿童、少年，在其父母或者其他法定监护人工作或者居住地接受义务教育的，当地人民政府应当为其提供平等接受义务教育的条件”，这在一定程度上降低了随迁儿童在迁入地的入学门槛。鉴于此，我们需要寻找其他因素来解释中国为什么有超过 60% 的进城务工人员会选择让自己的孩子留守农村。

我们提出，城镇公立学校的非政策性户籍歧视是外来务工人员选择让子女留守农村的一个重要原因。所谓“非政策性户籍歧视”是指公立学校在招收符合政策的外地户籍学生时，存在故意刁难或“吃拿卡要”等行为。这一歧视与政府限制外来人口流入的政策无关，它产生于学校或老师认定外地户籍学生的学习能力或素质差于本地学生<sup>②</sup>。以“教学质量”为标准的“教育锦标赛”导致公立学校有意减少招收外地户籍学生。

① 来自《流动儿童上学难：非京籍儿童北京念书需办 28 个证》，<http://edu.people.com.cn/n/2015/0615/c1053-27155342.html>。

② 2013-2014 学年中国教育追踪调查数据学校领导问卷报告，在所调查的 83 所城镇公立学校中，分别有 29.9%、26.9%、56.7% 和 27.9% 的学校领导认为“外地户籍学生入学时数学基础/语文基础/英语基础/行为习惯差于本地学生”，仅有 1 位学校领导认为“外地户籍学生优于本地学生”。班主任问卷报告，约 35.8% 的班主任认为班级里有外地户籍学生会使得班级管理的难度加大，仅有约 5.3% 的班主任认为自己可以管理有 1/3 或更多是外地户籍学生的班级。

实证检验上述假说是否成立,面临三个挑战:一是需要排除最低工资水平过低、居住条件不佳和政策性歧视影响;二是非政策性户籍歧视是不可观测的;三是需要证明非政策性户籍歧视对外来户籍学生的入学有实质性影响。我们采取的对应策略是:第一,使用2013-2014学年中国教育追踪调查数据中的城镇公立学校样本数据克服挑战。由于外地户籍学生是已入学的样本,所以此处无需讨论最低工资、居住条件不佳和政策性歧视对外地户籍学生入学的影响。第二,通过检验外地户籍学生家长是否为其子女能在迁入地读书而“找关系”,来推断公立学校在招生过程中是否存在非政策性户籍歧视现象。其中的逻辑是,如果外地户籍学生能够入读迁入地公立学校,则说明他满足迁入地居住条件和政策性门槛要求。但是其家长若为他能够入读现就读的公立学校而“找朋友帮忙或给领导送礼”,则证明学校在招收符合政策的外地户籍学生过程中存在故意刁难或“吃拿卡要”等非政策性户籍歧视行为。值得一提的是,“关系”是中国社会格局的核心模式(Jacobs, 1979, 1980; Gold et al., 2002; 朱光伟等, 2014),部分学生家长可能会为了子女能入读优质初中学校而“找关系”<sup>①</sup>,因此,我们需要排除为“择校”而“找关系”的干扰。第三,通过检验外地户籍学生现就读年级的年龄为“超龄”的概率是否在统计上显著地高于本地学生,来推断公立学校的非政策性户籍歧视是否对外地户籍学生入学造成了实质性的负面影响。

综上,我们将在因果识别框架下,通过检验以下三个命题来证明本文的核心观点:非政策性户籍歧视是外来务工人员选择让子女留守农村的原因之一。

命题1:“找关系”能显著提高外地户籍学生的入读比例。

命题2:在小学入学年龄、留级、休学、转学等因素相同的情况下,外地户籍学生现就读年级的年龄为“超龄”的概率显著高于本地学生。

命题3:通过“找关系”入学的外地户籍学生“超龄”的概率仍显著地高于本地学生。

命题1论证城镇公立学校存在“非政策性户籍歧视”。命题2论证“非政策性户籍歧视”对外地户籍学生入学造成了实质性的阻碍。命题3进一步论证对外地户籍学生家长而言,“找关系”具有较高成本。因为只有当“找关系”的成本足够高,才会使大多数外来务工人员选择让子女留守农村。

<sup>①</sup> 已有研究发现,“关系”会显著影响学生的教育获得(方长春,2005;周群力、陆铭,2009;张东娇,2010;赵延东、洪岩壁,2012)。

### 三 数据来源、模型建立和识别策略

#### （一）数据来源

本文采用的数据来自于2013-2014学年中国教育追踪调查（China Education Panel Survey, CEPS）数据，该调查由中国人民大学中国调查与数据中心设计与实施，是具有全国代表性的大型追踪调查项目。CEPS以2013-2014学年为基线，以初中一年级（7年级）和初中三年级（9年级）两个同期群为调查起点，以人口平均受教育水平和流动人口比例为分层变量，从全国随机抽取了28个县级单位（县、区、市）作为调查点。抽样过程分为四个阶段：首先，在全国县级行政单位中抽取28个县（区、市）；其次，在每个县（区、市）样本抽取4所开设了7年级和/或9年级的学校；再次，在每所学校样本中进一步抽取4个班级，包括2个7年级班和2个9年级班；最后，入选班级的所有学生、家长、班主任、主科目（语数英）任课老师以及学校领导构成最终调查样本。经过上述抽取后，CEPS随机抽取了112所学校，438个班级，共调查了19487名学生（被抽中班级的学生全体入样）。为检验“公立学校的非政策性户籍歧视”，本文仅筛选就读于城镇“公立学校”的学生样本<sup>①</sup>，具体包括：83所公立学校，327个班级，14524名学生。

#### （二）模型建立与识别策略

##### 1. 检验命题1的模型与策略

（1）基准回归。本部分通过实证回归模型（1），研究学生家长“找关系”比例对公立学校招收外地户籍学生比例的影响，推断公立学校是否存在非政策性户籍歧视。模型设置如下：

$$out\_student_{j,k} = \alpha_1 + \beta_1 \times guanxi_{j,k} + \gamma_1 \times control_{j,k} + \varepsilon_{j,k} \quad (1)$$

式（1）中，被解释变量  $out\_student_{j,k}$  表示班级外地户籍学生比例，具体为：在  $j$  学校， $k$  班级内，外地户籍学生人数占班级总人数的比例。其中， $out\_student$  表示学生是否为外地户籍学生，设置为虚拟变量，1 = 外地户籍学生（户籍登记地为外县区），0 = 本地学生（户籍登记地为本县区）。

① 2013-2014学年中国教育追踪调查（CEPS）数据调查了学校所在的地区类型：1. 市县城的中心城区；2. 市县城的边缘城区；3. 市县区的城乡结合部；4. 市县城区以外的镇；5. 农村。本文使用的是城镇，即选项为1/2/3/4的学生样本。

核心解释变量  $guanxi_{j,k}$  表示班级内学生家长“找关系”比例, 具体为: 在  $j$  学校,  $k$  班级内, 学生家长“找关系”人数占班级总人数的比例。  $guanxi$  为学生家长是否为其子女进入现就读学校而“找关系”, 设置为虚拟变量,  $1 =$  “找关系” (学生家长为了让子女进入现就读学校而“找朋友帮忙”或“给领导送礼”),  $0 =$  无“找关系” (学生家长为了让子女进入现就读学校什么事都没做)。

前定控制变量  $control_{j,k}$  包括<sup>①</sup>: ①学生层面特征变量: 班级男生比例、学生平均年龄、学生健康状况为良好比例、“语数外”考试成绩平均分 (取对数形式)。②父母层面特征变量: 父母至少一方受教育程度为大专及以上比例、家庭经济状况为贫穷比例。③校级固定效应。各变量定义如下: a. 性别, 设置虚拟变量,  $1 =$  男生,  $0 =$  女生; b. 年龄, 调查年份与出生年份的差值; c. 身体健康状况, 设置虚拟变量,  $1 =$  身体健康状况为良好 (比较好/很好),  $0 =$  身体健康状况为非良好 (很不好/不太好/一般); d. “语数外”考试成绩, “语数外”原始考试加总, 取自然对数; e. 父母受教育程度, 设置虚拟变量,  $1 =$  父母至少一方受教育程度为大专及以上,  $0 =$  父母双方受教育程度均在大专以下; f. 家庭经济状况, 设置虚拟变量,  $1 =$  家庭经济状况为贫穷 (非常困难/比较困难),  $0 =$  家庭经济状况为非贫穷 (中等/比较富裕/非常富裕)。其中, 学生层面特征变量和父母层面特征变量, 以“班级”为单位计算平均值。校级固定效应设置为哑变量,  $\varepsilon_{j,k}$  为误差项。为了使结论更为稳健, 我们也分别以“学校-年级”、“学校”为计算单位对模型 (1) 中各变量重新定义, 并实证回归。

通过模型 (1) 估计“找关系”对城镇公立学校招收外地户籍学生比例的影响, 推断公立学校是否存在非政策性户籍歧视, 将面临三个内生性挑战: 一是遗漏变量, 能容纳较多外地户籍学生的公立学校, 在招收过程中可能存在“吃拿卡要”行为倾向。二是双向因果关系, 在面临公立学校招收外地户籍学生名额有限 (过低) 的情况下, 外地户籍学生家长可能会主动“找关系”。三是测量误差, 因包含“择校”因素而使得“找关系”不是“非政策性户籍歧视”的有效衡量指标。下面, 我们将分别讨论克服这三个内生性挑战的识别策略。

(2) 降低内生性偏误。①遗漏变量 “一阶差分法”。基准模型 (1) 可能因遗漏

① 影响班级外地户籍学生比例的因素来自两方面: 一是与所处学校本身对外地户籍学生容纳量相关; 二是与外地户籍学生及其家庭禀赋相关。在实证回归中, 我们无法控制所有的遗漏变量, 限于数据获得, 此处仅控制部分外地户籍学生及其家庭禀赋遗漏变量, 后文进一步考虑遗漏变量对基准回归可能产生的影响。

学校层面特征变量（如存在“吃拿卡要”行为倾向）而出现估计偏误。考虑到同一学校招收7年级外地户籍学生比例和9年级外地户籍学生比例受学校层面特征变量因素在短期内是相同的，因此，我们可使用“一阶差分法”，对同一学校的9年级学生群体和7年级学生群体进行差分，以消除学校层面特征变量对班级外地户籍学生比例的影响。模型如下所示：

$$\Delta out\_student_j = \alpha_2 + \beta_2 \times \Delta guanxi_j + \gamma_2 \times \Delta control_j + \Delta \varepsilon_j \quad (2)$$

$$\begin{aligned} out\_student_{j,9} - out\_student_{j,7} &= \alpha_3 + \beta_3 \times (guanxi_{j,9} - guanxi_{j,7}) \\ &+ \gamma_3 \times (control_{j,9} - control_{j,7}) + (\varepsilon_{j,9} - \varepsilon_{j,7}) \end{aligned} \quad (3)$$

式（2）中，被解释变量  $\Delta out\_student_j$  表示在  $j$  学校中，9年级外地户籍学生比例与7年级外地户籍学生比例的变化量。核心解释变量  $\Delta guanxi_j$  表示在  $j$  学校中，9年级“找关系”比例与7年级“找关系”比例的变化量。控制变量  $\Delta control_j$  表示在  $j$  学校中，9年级与7年级相对应控制变量的变化量。控制变量的设置和定义同模型（1）一致，具体包括：年级男生比例、学生平均年龄、身体健康状况为良好比例、“语数外”考试成绩平均分（取对数形式）、父母至少一方受教育程度为大专及以上比例、家庭经济状况为贫穷比例。定义均以“学校-年级”为单位。

②双向因果关系。外地户籍学生家长“找关系”可能产生于以前年度学校招收外地户籍学生比例有限（过低）。鉴于此，我们通过两种方式缓解基准回归可能因外地户籍学生家长主动“找关系”而出现的估计偏误。具体为：其一，对7年级学生群体进行实证回归，并控制同一学校9年级学生群体中外地户籍学生比例，以控制以前年度学校招收外地户籍学生规模。模型设置如下：

$$out\_student_{j,7} = \alpha_4 + \beta_4 \times guanxi_{j,7} + out\_student_{j,9} + \gamma_4 \times control_{j,7} + \varepsilon_{j,7} \quad (4)$$

式（4）中，被解释变量  $out\_student_{j,7}$  表示在  $j$  学校中，7年级外地户籍学生比例，定义如下：以“学校-7年级”为单位，外地户籍学生人数占所处年级抽样学生总人数的比例。核心解释变量： $guanxi_{j,7}$  表示在  $j$  学校中，7年级“找关系”比例，定义如下：以“学校-7年级”为单位，学生家长“找关系”人数占所处年级抽样学生总人数的比例。 $out\_student_{j,9}$  表示在  $j$  学校中，9年级外地户籍学生比例，定义如下：以“学校-9年级”为单位，外地户籍学生人数占所处年级抽样学生总人数的比例。控制变量  $control_{j,7}$  设置同模型（1）一致，具体包括：年级男生比例、学生平均年龄、身体健康状况为良好比例、“语数外”考试成绩平均分（取对数形式）、父母至少一方受教

育程度为大专及以上学历比例、家庭经济状况为贫穷比例。定义均以“学校-7年级”为单位。

其二,我们通过实证回归同一学校里的9年级学生家长“找关系”比例对7年级外地户籍学生比例的影响,以校正外地户籍学生家长主动“找关系”可能导致的估计偏误。模型如下所示:

$$out\_student_{j7} = \alpha_5 + \beta_5 \times guanxi_{j9} + \gamma_5 \times control_{j7} + \varepsilon_{j7} \quad (5)$$

式(5)中,核心解释变量  $guanxi_{j9}$  表示在  $j$  学校中,9年级学生家长“找关系”比例,定义如下:以“学校-9年级”为单位,学生家长“找关系”人数占所处年级抽样学生总人数的比例。其他变量的设置和定义同模型(4)一致。

③测量误差:排除为“择校”而“找关系”干扰。本部分通过实证回归模型(6),并使用 ordered probit (有序 probit) 模型研究“找关系”对外地户籍学生和本地户籍学生入读优质初中的影响,以检验基准回归实证结果是否受测量误差(“择校”因素)影响。其中的逻辑是:如果“找关系”能显著提高外地户籍学生入读优质初中的概率,则表明外地户籍学生家长不仅仅是为子女能在迁入地读书而“找关系”,即基准回归实证结果是有偏的。相反,如果“找关系”无法显著提高外地户籍学生入读优质初中的概率,则表明外地户籍学生家长是为子女能在迁入地读书(非“择校”)而“找关系”。模型设置如下所示:

$$sort\_school_i = \alpha_6 + \beta_6 \times outstudent\_guanxi_i + \beta_7 \times student\_guanxi_i + \beta_8 \times student\_noguanxi_i + \gamma_6 \times control_i + \zeta_i \quad (6)$$

式(6)中,被解释变量  $sort\_school_i$  表示学生现就读学校在本县(区)的排名情况。按照 CEPS 的学校领导问卷,设置为排序变量,1=最差,2=中下,3=中间,4=中上,5=最好。核心解释变量:我们依据“户籍”和有无“找关系”,将学生样本划分为四组:①“找关系”的外地户籍学生  $outstudent\_guanxi_i$ ; ②无“找关系”的外地户籍学生  $outstudent\_noguanxi_i$ ; ③“找关系”的本地学生  $student\_guanxi_i$ ; ④无“找关系”的本地学生  $student\_noguanxi_i$ 。将其设置为哑变量,并以②无“找关系”的外地户籍学生作为基准组。前定控制变量  $control_i$  包括:学生层面特征变量(性别、年龄、身体健康状况、“语数外”考试成绩)、父母层面特征变量(受教育程度、家庭经济状况)和县级固定效应。各控制变量定义同模型(1)一致。县级固定效应设置哑变量。 $\zeta_i$  为误差项。

## 2. 检验命题2的模型与策略

如果公立学校在招收外地户籍学生时存在“非政策性户籍歧视”,外地户籍学生家

长需要“找关系”才能让孩子顺利读书，那么这一歧视必然对外地户籍学生的入学造成了负向影响：假如外地户籍学生家长在孩子适龄入学的当年无法找到“关系”，则外地户籍学生将可能被迫重返家乡，成为（短暂或永久）留守儿童，等待父母第二年（甚至第三年）重新寻找“关系”以进入迁入地公立学校。因此，我们可以通过研究外地户籍学生现就读年级的年龄为“超龄”的概率是否显著高于本地学生，来推断“非政策性户籍歧视”对外地户籍学生入学是否造成了实质性的阻碍。

我们通过分年级样本实证回归模型（7）并使用 logit 模型，研究外地户籍学生现就读年级的年龄是否“超龄”的概率与本地学生是否存在显著差异。模型设置如下：

$$over\_age_{jki} = \alpha_7 + \beta_9 \times outstudent_{jki} + \gamma_7 \times control_{jki} + \psi_{jki} \quad (7)$$

式（7）中，被解释变量  $over\_age_{jki}$  表示在  $j$  学校中， $k$  年级， $i$  学生现就读年级的年龄是否“超龄”。如果学生在年满 6 周岁开始接受义务教育，在就读 7 年级时的年龄应为 12 岁或 13 岁（适龄），在就读 9 年级时的年龄应为 14 岁或 15 岁（适龄）。我们对此分别设置虚拟变量：在 7 年级中，1 = 学生年龄大于 13 岁（“超龄”），0 = 学生年龄不大于 13 岁（“适龄”）；在 9 年级中，1 = 学生年龄大于 15 岁（“超龄”），0 = 学生年龄不大于 15 岁（“适龄”）；作为稳健性检验，我们使用“大于 14 岁”作为 7 年级学生“超龄”衡量指标，使用“大于 16 岁”作为 9 年级学生“超龄”衡量指标。

核心解释变量： $outstudent_{jki}$  表示在  $j$  学校中， $k$  年级， $i$  学生是否为外地户籍学生。控制变量  $control_{jki}$  包括学生层面特征变量（性别、身体健康状况、“语数外”考试成绩）、父母层面特征变量（受教育程度、家庭经济状况）和县级固定效应。 $\psi_{jki}$  为误差项。各变量定义同模型（1）一致。

### 3. 检验命题 3 的模型与策略

如果“找关系”的成本低，则非政策性户籍歧视对外地户籍学生入读迁入地初中的影响不大。对此，我们可以通过检验“找关系”入学的外地户籍学生“超龄”的概率与本地学生是否存在显著差异，来推断外地户籍学生家长“找关系”入学的成本高低。如果“找关系”外地户籍学生“超龄”的概率与本地学生不存在显著差异，则意味着“找关系”的成本低。然而，如果“找关系”外地户籍学生“超龄”的概率显著高于本地学生，则意味着“找关系”的成本高，因为外地户籍学生在当年的“小升初”找不到“关系”，需要等到第二年（甚至第三年）才能找到“关系”入学。模型设置如下：

$$over\_age_{jki} = \alpha_8 + \beta_{10} \times outstudent\_noguanxi_{jki} + \beta_{11} \times student\_guanxi_{jki} + \beta_{12} \times student\_noguanxi_{jki} + \gamma_8 \times control_{jki} + \psi_{jki} \quad (8)$$

式（8）中，被解释变量  $over\_age_{jki}$  表示在  $j$  学校中， $k$  年级， $i$  学生现就读年级

的年龄是否“超龄”。核心解释变量是，我们依据“户籍”和有无“找关系”，将学生样本划分为四组<sup>①</sup>，分别是“找关系”的外地户籍学生 ( $outstudent\_guanxi_{jki}$ )、无“找关系”的外地户籍学生 ( $outstudent\_noguanxi_{jki}$ )、“找关系”的本地学生 ( $student\_guanxi_{jki}$ )、无“找关系”的本地学生 ( $student\_noguanxi_{jki}$ )，设置为哑变量，并以“找关系”的外地户籍学生为基准组。控制变量  $control_{jki}$  的设置和定义同模型 (7) 一致。

### (三) 统计事实

表 1 为模型 (1) 中相关变量的描述性统计。其中，第 (1) ~ (3) 栏为以“班级”为单位，第 (4) ~ (6) 栏为以“学校 - 年级”为单位，第 (7) ~ (9) 栏为以“学校”为单位。由统计结果可知，以“班级”、“学校 - 年级”、“学校”为单位计算的各变量均值和标准差基本一致，其中，“外地户籍学生”占比约为 21.6%，这一数据与 2013 年和 2014 年国家统计局公布的全国“人户分离”人口比例基本一致。通过“找关系”入学的学生比例为 17.4% 左右。模型 (1) 解释变量之间 (不包括校级固定效应) 相关系数绝对值最高不超过 0.7，方差膨胀因子 (VIF) 最高值仅为 1.74，远小于 10，从而不必担心解释变量之间存在严重的多重共线性问题。

表 1 模型 (1) 变量的统计描述

	班级			学校 - 年级			学校		
	均值	标准差	样本量	均值	标准差	样本量	均值	标准差	样本量
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
外地户籍学生比例	0.216	0.206	327	0.215	0.198	164	0.217	0.193	83
“关系户”比例	0.174	0.126	327	0.173	0.107	164	0.171	0.092	83
男生比例	0.514	0.082	327	0.511	0.064	164	0.509	0.052	83
平均年龄	13.097	1.169	327	13.901	1.169	164	13.820	0.610	83
身体健康为良好比例	0.737	0.107	327	0.738	0.094	164	0.740	0.079	83
家境为贫穷比例	0.170	0.158	327	0.169	0.151	164	0.166	0.148	83
父母为大专及以上学历比例	0.241	0.237	327	0.243	0.231	164	0.246	0.228	83
“语数外”自然对数	5.471	0.259	327	5.478	0.238	164	5.486	0.207	83

资料来源：根据 2013 - 2014 学年中国教育追踪调查 (CEPS) 数据计算得到。

<sup>①</sup> Ai & Norton (2003) 指出，使用 probit 模型和 logit 模型等非线性模型对含有交互项的模型进行估计，通过交互项的符号、大小和显著性进一步判断交互影响，可能会得到错误的结论。

表2是模型(5)各变量的描述性统计。第一部分为比较外地户籍学生和本地学生相关特征变量是否存在显著差异。其中,第(1)~(3)栏为外地户籍学生,第(4)~(6)栏为本地学生,第(7)~(9)栏为检验第(1)和第(4)栏之间相关变量是否存在显著差异的t检验结果。第二部分为比较是否“找关系”学生相关特征变量是否存在显著差异。其中,第(1)~(3)栏为“找关系”的学生样本,第(4)~(6)栏为无“找关系”的学生样本,第(7)~(9)栏为检验第(1)和第(4)栏之间相关变量是否存在显著差异的t检验结果。

由表2可知,平均而言,外地户籍学生占比约为18.7%，“找关系”占比约为17.8%。外地户籍学生入读优质初中的比例显著低于本地学生,“找关系”学生入读优质初中的比例显著高于无“找关系”的学生。外地户籍学生“找关系”比例显著高于本地学生。当然,简单的描述性统计仅能揭示户籍身份、“找关系”与所处学校优质与否之间的相关关系,而无法揭示两者之间的因果关系。

表2 模型(5)变量的统计描述

	均值	标准差	样本量	均值	标准差	样本量	均值差	标准误	t 值
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<b>第一部分</b>	外地户籍学生			本地学生			t 检验		
所在学校的县区排名	3.739	0.900	2692	4.101	0.701	11690	0.363	0.016	22.856
“找关系”	0.277	0.448	2133	0.157	0.364	9841	-0.120	0.009	-13.213
男生	0.519	0.500	2717	0.509	0.500	11632	-0.011	0.011	-0.991
年龄	13.895	1.337	2688	13.831	1.309	11546	-0.064	0.028	-2.279
身体健康为良好	0.763	0.425	2736	0.732	0.443	11667	-0.031	0.009	-3.317
“语数外”原始分	243.255	72.591	2656	247.463	75.581	11497	4.208	1.615	2.605
父母为大专及以上学历	0.174	0.379	2684	0.254	0.435	11474	0.081	0.009	8.839
家境为贫穷	0.151	0.358	2584	0.184	0.387	11215	0.033	0.008	3.913
<b>第二部分</b>	“找关系”学生			无“找关系”学生			t 检验		
所在学校的县区排名	4.069	0.786	2117	4.029	0.725	9745	-0.040	0.018	-2.246
外地户籍学生	0.277	0.447	2137	0.157	0.364	9837	-0.120	0.009	-13.213
男生	0.539	0.499	2108	0.491	0.500	9721	-0.048	0.012	-4.016
年龄	14.015	1.379	2088	13.775	1.300	9645	-0.240	0.032	-7.568
身体健康为良好	0.729	0.445	2124	0.743	0.437	9760	0.014	0.010	1.365

续表

	均值	标准差	样本量	均值	标准差	样本量	均值差	标准误	t 值
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
第二部分	“找关系”学生			无“找关系”学生			t 检验		
“语数外”原始分	240.170	73.663	2071	248.850	74.917	9613	8.680	1.810	4.797
父母为大专及以上学历	0.231	0.422	2079	0.229	0.420	9606	-0.003	0.010	-0.271
家境为贫穷	0.191	0.393	2077	0.179	0.383	9715	-0.012	0.009	-1.293

注: t 检验为检验第 (1) 和第 (4) 栏两组样本各变量均值是否存在显著差异, 其原假设为: 外地户籍学生与本地学生相对应变量均值相等 “找关系” 学生与无 “找关系” 学生相对应变量均值相等。

资料来源: 根据 2013 - 2014 学年中国教育追踪调查 (CEPS) 数据计算得到。

## 四 实证结果分析

### (一) “找关系”能提高外地户籍学生的入读比例吗?

#### 1. 基准回归

表 3 报告了模型 (1) 的回归结果。第 (1)、(2) 栏以 “班级” 为单位, 分别研究在有无控制学生及家庭禀赋特征变量 (班级男生比例、学生平均年龄、学生健康状况为良好比例、“语数外” 考试成绩平均分、父母至少一方受教育程度为大专及以上学历、家庭经济状况为贫穷比例) 和校级固定效应之后, “找关系” 比例对公立学校招收外地户籍学生比例的影响。第 (3) ~ (5) 栏为以 “学校 - 年级” 为单位, 其中, 第 (3) 栏为全样本, 第 (4) 栏为 7 年级分样本, 第 (5) 栏为 9 年级分样本<sup>①</sup>。第 (6) 栏为以 “学校” 为单位。在估计方法应用上, 为避免受相同学校因素影响而导致扰动项自相关, 本文报告经由 “学校” 层面聚类调整的稳健标准误。

表 3 的实证结果表明, 无论是以 “班级”、“学校 - 年级”, 还是以 “学校” 为单位, 在控制相关特征变量之后, “找关系” 能显著提高外地户籍学生的入读比例。以第 (2) 栏为例, “找关系” 比例的系数为 0.174, 即在其他条件不变的情况下, “找关系” 比例每上升 1 个点, 公立学校招收外地户籍学生比例相应上升 0.174 个点。

① 从 2013 - 2014 学年 CEPS 数据可知, 学校编号 36、40 仅调查 7 年级, 未调查 9 年级。

表3 基准回归 “找关系”能提高外地户籍学生的入读比例吗

因变量	外地户籍学生比例					
	学校-班级	学校-班级	学校-年级	学校-7 年级	学校-9 年级	学校
回归样本						
模型设定	OLS					
模型编号	(1)	(2)	(3)	(4)	(5)	(6)
“找关系”比例	0.196 (0.194)	0.174* (0.093)	0.300** (0.137)	0.824*** (0.222)	0.454** (0.179)	0.715*** (0.198)
控制变量	—	是	是	是	是	是
学校固定效应	—	是	是	—	—	—
县级固定效应	—	—	—	是	是	是
样本量	327	327	164	83	81	83
R <sup>2</sup>	0.014	0.875	0.949	0.776	0.746	0.781

注：括号内为聚类到学校层面的稳健标准误；\*\*\*、\*\*、\* 分别表示在 1%、5% 和 10% 水平上显著；控制变量具体包括男生比例、平均年龄、身体健康良好比例、“语数外”考试成绩平均分（对数）、父母至少一方的受教育程度为大专及以上比例、家庭经济状况贫穷比例；各控制变量的符号和显著性基本符合预期，限于篇幅，未予列示。

资料来源：根据 2013 - 2014 学年中国教育追踪调查（CEPS）数据计算得到。

## 2. 缓解内生性偏误

(1) 遗漏变量和双向因果关系。本部分我们试图以“学校-年级”为单位，通过使用一阶差分法（模型 2）、代理变量法（模型 4、模型 5）校正基准回归结果可能因遗漏变量和双向因果关系而面临的估计偏误。表 4 第（1）~（3）栏分别汇报模型 2、4、5 估计结果。其中，第（1）栏为利用“一阶差分法”，对同一学校 9 年级学生群体和 7 年级学生群体进行差分，以校正因遗漏学校层面特征变量对基准回归可能造成的估计偏误。结果显示，“找关系”仍在 1% 水平上显著提高公立学校招收外地户籍学生比例。第（2）、第（3）栏分别为在基准回归基础上控制同一学校 9 年级外地户籍学生比例和使用同一学校 9 年级学生家长“找关系”比例作为核心解释变量，以校正 7 年级外地户籍学生家长因面临以前年度学校招收外地户籍学生有限（过低）情况下主动“找关系”而导致的估计偏误。第（2）栏实证结果显示，“找关系”能提高公立学校招收外地户籍学生比例，这一结论仍成立。此外，同一学校 9 年级外地户籍学生比例对 7 年级外地户籍学生比例呈显著正向影响。第（3）栏实证结果显示，结果不变。

表 4 缓解内生性偏误：遗漏变量和双向因果关系

因变量	外地户籍学生比例							
	基准回归			稳健：仅控制县级效应			稳健：9 年级控制变量	
回归类型	校-年级	校-年级	校-年级	校-年级	校-年级	校-年级	校-年级	校-年级
回归样本	全样本	7 年级	7 年级	全样本	7 年级	7 年级	7 年级	7 年级
模型设定	一阶差分	代理变量	滞后项	一阶差分	代理变量	滞后项	代理变量	滞后项
	FD	OLS	OLS	FD	OLS	OLS	OLS	OLS
模型编号	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
“找关系”比例	0.280 <sup>***</sup> (0.103)	0.487 <sup>***</sup> (0.125)		0.304 <sup>***</sup> (0.113)	0.373 <sup>***</sup> (0.112)		0.425 <sup>***</sup> (0.139)	
外地户籍学生比例_9 年级		0.857 <sup>***</sup> (0.135)			0.915 <sup>***</sup> (0.121)		0.907 <sup>***</sup> (0.137)	
“找关系”比例_9 年级			0.476 <sup>**</sup> (0.231)			0.299 (0.255)		0.471 <sup>*</sup> (0.257)
控制变量	是	是	是	—	—	—	是	是
县级固定效应	—	是	是	—	是	是	是	是
样本量	81	81	81	81	81	81	81	81
R <sup>2</sup>	0.213	0.904	0.732	0.099	0.883	0.630	0.900	0.708

注：括号内为聚类到学校层面的稳健标准误；\*\*\*、\*\*、\* 分别表示在 1%、5% 和 10% 水平上显著；控制变量包括男生比例、平均年龄、身体健康良好比例、“语数外”考试成绩平均分（对数）、父母至少一方的受教育程度为大专及以上比例、家庭经济状况贫穷比例；值得说明的是，为缓解各同期控制变量对实证结果可能造成的内生性偏误，第（7）、第（8）栏实证回归中的控制变量，使用 9 年级的控制变量作为 7 年级对应的控制变量；各控制变量的符号和显著性基本符合预期，限于篇幅，未予列示。

资料来源：根据 2013 - 2014 学年中国教育追踪调查（CEPS）数据计算得到。

第（4）~（8）栏做了两个稳健性检验：①仅控制县级固定效应，而不控制其他特征变量。第（4）栏的“一阶差分法”估计结果显示，即使仅控制县级固定效应，“找关系”仍能显著提高外地户籍学生的入读比例。第（5）栏实证结果显示，“找关系”对公立学校招收外地户籍学生比例呈显著正向影响，这一结论不变，9 年级外地户籍学生比例对 7 年级外地户籍学生比例仍呈显著正向影响。第（6）栏实证结果表明，仅控制县级固定效应，9 年级学生“找关系”比例对 7 年级外地户籍学生比例呈不显著正向影响。②使用同一学校 9 年级学生特征变量作为 7 年级学生对应特征变量的代替变量。第（7）栏实证结果显示，即使替换控制变量，“找关系”仍能显著提高外地户籍学生的入读比例，且 9 年级外地户籍学生比例对 7 年级外地户籍学生比例仍呈显著正向影响。第（8）栏实证结果显示，即使替换控制变量，同一学校 9 年级学生“找关系”比例对 7 年级外地户籍学生比例呈显著正向影响。

（2）测量误差，排除为“择校”而“找关系”干扰。表 5 报告了模型（6）的实

证结果。第(1)栏为使用 ordered probit (有序 probit) 模型研究在没有控制特征变量的情况下,外地户籍学生入读优质初中的概率是否与本地学生存在显著差异。第(2)栏为在第(1)栏的基础上控制学生层面特征变量(性别、年龄、身体健康状况、“语数外”考试成绩)、父母层面特征变量(受教育程度、家庭经济状况)和县级固定效应。第(3)栏为在第(2)栏的基础上进一步控制“找关系”变量,以研究“找关系”对学生入读优质初中的影响。第(4)栏为研究“找关系”对本地学生和外地户籍学生入读优质初中的影响。第(5)栏为使用 OLS 方法重新回归第(4)栏。

表5的实证结果表明,外地户籍学生入读优质初中的概率显著低于本地学生。“找关系”能显著提高学生入读优质初中概率,然而,这一正向影响仅体现在本地学生,对外地户籍学生无显著作用<sup>①</sup>。以第(5)栏为分析对象,相比于无“找关系”外地户籍学生,无“找关系”的本地学生入读优质初中概率显著高13.2%,“找关系”的本地学生显著高21.3%,而“找关系”的外地户籍学生则不存在显著差异。因此,外地户籍学生家长“找关系”仅是为其子女能在迁入地读书,而非“择校”。这表明基准回归结果不受“择校”因素测量误差影响。

综上,即使考虑遗漏变量、双向因果关系和测量误差因素,“找关系”能显著提高外地户籍学生的入读比例,即公立学校存在非政策性户籍歧视,这一结论稳健存在。

表5 缓解内生性偏误:测量误差(“择校”因素)

因变量	所在学校在本县区的排名情况				
	有序 probit	有序 probit	有序 probit	有序 probit	OLS
模型编号	(1)	(2)	(3)	(4)	(5)
外地户籍学生	-0.498 *** (0.135)	-0.244 ** (0.104)	-0.273 ** (0.109)		
“找关系”			0.162 ** (0.064)		
“找关系”的外地户籍学生				0.078 (0.116)	0.022 (0.072)
“找关系”的本地户籍学生				0.437 *** (0.113)	0.213 *** (0.062)

① 我们更换基准组,以无“找关系”本地学生作为基准组,有序 probit 模型估计结果发现,在控制了特征变量后,“找关系”外地户籍学生、无“找关系”外地户籍学生和“找关系”本地学生变量系数分别为-0.169、-0.247、0.190,聚类到学校的稳健标准误分别为0.155、0.111和0.066。限于篇幅,该实证结果未列出。

续表

因变量	所在学校在本县区的排名情况				
	有序 probit	有序 probit	有序 probit	有序 probit	OLS
模型编号	(1)	(2)	(3)	(4)	(5)
无“找关系”的本地户籍学生				0.247** (0.111)	0.132** (0.063)
控制变量	—	是	是	是	是
县级固定效应	—	是	是	是	是
样本量	14382	12533	10721	10721	10721
R <sup>2</sup>	0.015	0.228	0.233	0.233	0.385

注：括号内为聚类到学校层面的稳健标准误；\*\*\*、\*\*、\* 分别表示在 1%、5% 和 10% 水平上显著；控制变量包括学生层面特征变量（男生、年龄、身体健康良好、“语数外”考试成绩（对数））和父母层面特征变量（父母至少一方的受教育程度为大专及以上、家庭经济状况贫穷）；各控制变量的符号和显著性基本符合预期，限于篇幅，未予列示。

资料来源：根据 2013 - 2014 学年中国教育追踪调查（CEPS）数据计算得到。

## （二）“非政策性户籍歧视”对外地户籍学生入学有实质性阻碍吗？

### 1. 外地户籍学生“超龄”的概率高于本地学生吗？

本部分我们通过研究外地户籍学生现就读年级的年龄为“超龄”的概率是否显著高于本地学生，以推断“公立学校的非政策性户籍歧视”对外地户籍学生入学是否有实质性的负向影响。表 6 汇报模型（7）的实证结果。第（1）栏为 7 年级学生样本，第（4）栏为 9 年级学生样本。其中，第（1）、第（4）栏分别为使用 logit 模型，研究在控制学生、父母层面特征变量和校级固定效应的情况下，7 年级、9 年级外地户籍学生“超龄”的概率是否显著高于本地学生。实证结果显示外地户籍学生“超龄”的概率在 1% 水平上显著高于本地学生。

从逻辑上讲，除了“非政策性户籍歧视”之外，还有四个因素能够导致外地户籍学生与本地学生在就读 7 年级和 9 年级时出现年龄差异，分别是：小学入学年龄、留级、转学和休学因素<sup>①</sup>。鉴于此，我们在表 6 第（1）、第（4）栏的基础上进一步控制学生的小学入学年龄、留级、休学和转学特征变量。第（2）、第（5）栏的实证结果显示，即使考虑小学入学年龄、留级、转学和休学因素，外地户籍学生“超龄”的概率仍

① 小学入学年龄为学生开始上小学的年龄。留级为虚拟变量，1 = 小学 1 年级到 6 年级留级的次数大于 0（留级学生），0 = 小学 1 年级到 6 年级留级的次数等于 0（非留级学生）。休学、转学的定义同留级一致，1 = “休学学生”，0 = “不休学学生”；1 = “转学学生”，0 = “不转学学生”。

显著高于本地学生。为了结论更为稳健，第（3）、第（6）栏对学生现所处年级的年龄为“超龄”重新定义，具体如下：在7年级中，1 = 学生年龄大于14岁（“超龄”），0 = 学生年龄不大于14岁（“适龄”）；在9年级中，1 = 学生年龄大于16岁（“超龄”），0 = 学生年龄不大于16岁（“适龄”）。实证结果显示，即使重新定义学生“超龄”，外地户籍学生“超龄”的概率仍显著高于本地学生。

表6 外地户籍学生“超龄”的概率高于本地学生吗

因变量	现就读年级的年龄为“超龄”					
	7 年级学生样本			9 年级学生样本		
“超龄”定义	大于 13 岁	大于 14 岁	大于 15 岁	大于 15 岁	大于 16 岁	大于 16 岁
模型设定	logit					
模型编号	(1)	(2)	(3)	(4)	(5)	(6)
外地户籍学生	0.439*** (0.107)	0.296*** (0.103)	0.411** (0.164)	0.705*** (0.123)	0.565*** (0.126)	0.489** (0.216)
其他干扰因素	—	是	是	—	是	是
控制变量	是	是	是	是	是	是
校级固定效应	是	是	是	是	是	是
样本量	6716	6624	5717	5814	5683	4557
Pseudo R <sup>2</sup>	0.324	0.370	0.253	0.336	0.405	0.297

注：括号内为聚类到学校层面的稳健标准误；\*\*\*、\*\*、\* 分别表示在 1%、5% 和 10% 水平上显著；控制变量包括学生特征变量（男生、身体健康良好、“语数外”考试成绩（对数））和父母特征变量（父母至少一方的受教育程度为大专及以上、家庭经济状况贫穷）；其他干扰因素包括小学入学年龄、留级、转学和休学因素；各控制变量的符号和显著性基本符合预期，限于篇幅，未予列示。

资料来源：根据 2013 - 2014 学年中国教育追踪调查（CEPS）数据计算得到。

## 2. “找关系”能降低外地户籍学生“超龄”的概率吗？

表 7 汇报模型（8）的实证结果。第（1）~（3）栏为 7 年级学生样本，第（4）~（6）栏为 9 年级学生样本。其中，第（1）、第（4）栏为在表 6 第（2）、第（5）栏的基础上控制“找关系”变量，使用 logit 模型研究“找关系”对学生“超龄”的概率的影响<sup>①</sup>。结果显示，“找关系”对学生“超龄”概率呈不显著影响。第（2）、第（5）栏为研究“找关系”对外地户籍学生和本地学生“超龄”概率的影响。结果显示，“找关系”外地户籍学生“超龄”的概率仍显著地高于本地学生，这表明

① 研究学生“超龄”对父母“找关系”的影响时，logit 模型估计结果发现，无论是 7 年级，还是 9 年级，在控制特征变量（同表 7 一致 + 外地户籍学生变量）后，学生“超龄”对父母“找关系”均呈不显著的影响。限于篇幅，未予列示。

“找关系”无法降低外地户籍学生“超龄”概率。第(3)、第(6)栏为使用 OLS 方法重新回归第(2)、第(5)栏，结论不变。

综上，“公立学校的非政策性户籍歧视”对外地户籍学生入学造成了实质性阻碍，并且对外地户籍学生家长而言，“找关系”具有较高成本。因此，进城务工人员不得不选择让他们的子女留守农村。

表 7 “找关系”能降低外地户籍学生“超龄”的概率吗

因变量	现就读年级的年龄为“超龄”					
	7 年级学生样本			9 年级学生样本		
“超龄”定义	“超龄”：大于 13 岁			“超龄”：大于 15 岁		
模型设定	logit	logit	OLS	logit	logit	OLS
模型编号	(1)	(2)	(3)	(4)	(5)	(6)
外地户籍学生	0.305*** (0.105)			0.587*** (0.136)		
“找关系”	-0.069 (0.100)			0.008 (0.138)		
无“找关系”的外地户籍学生		0.056 (0.179)	0.008 (0.026)		-0.167 (0.243)	-0.023 (0.034)
“找关系”的本地户籍学生		-0.322* (0.196)	-0.044* (0.025)		-0.754*** (0.271)	-0.097*** (0.037)
无“找关系”的本地户籍学生		-0.245# (0.169)	-0.035# (0.023)		-0.695*** (0.239)	-0.095*** (0.034)
其他干扰因素	是	是	是	是	是	是
控制变量	是	是	是	是	是	是
校级固定效应	是	是	是	是	是	是
样本量	5489	5489	5730	4942	4942	5140
Pseudo R <sup>2</sup> / R <sup>2</sup>	0.376	0.376	0.386	0.417	0.417	0.425

注：括号为聚类到学校层面的稳健标准误；\*\*\*、\*\*、\*、# 分别表示在 1%、5%、10% 和 15% 水平上显著；控制变量包括学生特征变量（男生、身体健康良好、“语数外”考试成绩（对数））和父母特征变量（父母至少一方的受教育程度为大专及以上、家庭经济状况贫穷）；其他干扰因素包括小学入学年龄、留级、转学和休学因素；各控制变量的符号和显著性基本符合预期，限于篇幅，未予列示。

资料来源：根据 2013 - 2014 学年中国教育追踪调查（CEPS）数据计算得到。

## 五 结论

本文试图解释为什么在政府大力解决随迁子女平等入学的背景下，仍有超过 60% 的进城务工人员选择让处于义务教育阶段的孩子留守农村。与已有文献探讨的最低工资、居住条件不佳和政策性歧视影响因素不同，本文提出，“来自公立学校的非政策性

户籍歧视”是导致留守儿童大规模出现的一个原因。在实证中，我们以外地户籍学生家长是否为其子女能在迁入地读书（非“择校”）而“找关系”作为学校是否存在非政策性户籍歧视的代替变量，并利用2013-2014学年中国教育追踪调查（CEPS）数据中的城镇公立学校样本实证检验了这个观点。研究发现：

第一，公立学校存在非政策性户籍歧视。从“班级”、“学校-年级”和“学校”视角，我们均发现，“找关系”能显著提高外地户籍学生入读迁入地初中的概率。即使使用“一阶差分法”处理遗漏变量，使用核心解释变量滞后项缓解双向因果关系，考虑“择校”因素测量误差，以及进行一系列稳健性检验，这一结论仍稳健成立。第二，非政策性户籍歧视对外地户籍学生入学造成了实质性阻碍。在控制了小学入学年龄、留级、休学、转学等因素之后，无论外地户籍学生是否“找关系”，他们“超龄”的概率都显著高于本地学生，这意味着“找关系”的高成本使得外地户籍学生不得不推迟入学初中。

综上，本文的结论是，城镇公立学校存在非政策性户籍歧视，而且它对外地户籍学生入学造成了实质性阻碍，因此，进城务工人员不得不选择让处于义务教育阶段的子女留守农村。这一结论的政策含义是，解决农民工子女的入学难问题，除了提高进城务工人员的工资、居住条件，以及降低随迁儿童在迁入地的入学门槛之外，同时需要降低迁入地公立学校的入学“门槛”，发挥公立学校主渠道作用，严禁公立学校的非政策性户籍歧视，促进教育公平。

## 参考文献：

- 白勤、林泽炎、谭凯鸣（2012），《中国农村留守儿童培养模式实验研究——基于现场干预后心理健康状况前后变化的数量分析》，《管理世界》第2期，第62-72页。
- 陈雯（2014），《青年民工家庭“弱化”与婚育“催化”研究》，《中国青年政治学院学报》第1期，第89-96页。
- 方长春（2005），《家庭背景与教育分流——教育分流过程中的非学业性因素分析》，《社会》第4期，第105-118页。
- 国家统计局（2015），《中华人民共和国2014年国民经济和社会发展统计公报》，[http://www.stats.gov.cn/tjsj/zxfb/201502/t20150226\\_685799.html](http://www.stats.gov.cn/tjsj/zxfb/201502/t20150226_685799.html)。
- 李强、臧文斌（2010），《父母外出对留守儿童健康的影响》，《经济学（季刊）》第10

卷第1期,第341-360页。

谭深(2011),《中国农村留守儿童研究述评》,《中国社会科学》第1期,第138-150页。

吴霓(2004),《农村留守儿童问题调研报告》,《教育研究》第10期,第15-18页。

叶静怡、张睿、王琼(2017),《农民进城务工与子女教育期望——基于2010年中国家庭追踪调查数据的实证分析》,《经济科学》第1期,第90-105页。

叶静怡、张睿、杨洋(2015),《户籍制度、工作流动性与进城务工人员子女就学——基于2008年CHIP数据的研究》,《学习与探索》第5期,第109-114页。

叶敬忠、James Murray(2005),《关注留守儿童:中国中西部农村地区劳动力外出务工对留守儿童的影响》,北京:社会科学文献出版社。

张东娇(2010),《义务教育阶段择校行为分析:社会资本结构的视角》,《教育发展研究》第2期,第12-17页。

章莉、李实、William Darity Jr.、Rhonda Vonshay Sharpe(2014),《中国劳动力市场上工资收入的户籍歧视》,《管理世界》第11期,第35-46页。

赵延东、洪岩璧(2012),《社会资本与教育获得——网络资源与社会闭合的视角》,《社会学研究》第5期,第47-68页。

周群力、陆铭(2009),《拜年与择校》,《世界经济文汇》第6期,第19-34页。

朱光伟、杜在超、张林(2014),《关系、股市参与和股市回报》,《经济研究》第11期,第87-101页。

Ai, Chunrong & Edward Norton (2003). Interaction Terms in Logit and Probit Models. *Economics Letters*, 80 (1), 123-129.

Chan, Kam Wing (1996). Post-Mao China: A Two-Class Urban Society in the Making. *International Journal of Urban and Regional Research*, 20 (1), 134-150.

Gold, Thomas, Doug Guthrie & David Wank (2002). *Social Connections in China: Institutions, Culture, and the Changing Nature of Guanxi*. New York: Cambridge University Press.

Jacobs, Bruce (1979). A Preliminary Model of Particularistic Ties in Chinese Political Alliances: Kan-ch'ing and Kuan-hsi in a Rural Taiwanese Township. *The China Quarterly*, 78, 237-273.

Jacobs, Bruce (1980). *Local Politics in a Rural Chinese Cultural Setting: A Field Study of Mazu Township, Taiwan*. Canberra: Australian National University Press.

- Knight , John , Lina Song & Huaibin Jia ( 1999 ) . Chinese Rural Migrants in Urban Enterprises: Three Perspectives. *The Journal of Development Studies* ,35 ( 3 ) ,73 – 104.
- Liu , Zhiqiang ( 2005 ) . Institution and Inequality: the Hukou System in China. *Journal of Comparative Economics* ,33 ( 1 ) ,133 – 157.
- Meng , Xin ( 2000 ) . *Labour Market Reform in China*. New York: Cambridge University Press.
- Meng , Xin & Junsen Zhang ( 2001 ) . The Two-Tier Labor Market in Urban China: Occupational Segregation and Wage Differentials between Urban Residents and Rural Migrants in Shanghai. *Journal of Comparative Economics* ,29 ( 3 ) ,485 – 504.
- Sicular , Terry , Ximing Yue , Björn Gustafsson & Shi Li ( 2007 ) . The Urban-Rural Income Gap and Inequality in China. *Review of Income and Wealth* ,53 ( 1 ) ,93 – 126.

## Why Are There So Many Left-behind Children in China?

Wang Zhibo<sup>1</sup> & Li Changhong<sup>2</sup>

( School of Economics & Management , South China Normal University<sup>1</sup>;

The Wang Yanan Institute for Studies in Economics , Xiamen University<sup>2</sup>)

**Abstract:** This paper is intended to explain why over 60 percent of rural migrant workers choose to leave their school-age children in the rural areas , even though the government is doing their best to provide equal enrollment for migrant children. Here we propose that non-policy household registration discrimination from public schools is a major reason that creates the large-scale of left-behind children. Empirically , we use the behaviour of parents finding “guanxi” for their children to study in public schools ( not choosing schools ) as a proxy variable of non-policy household registration discrimination from public schools. The results show that , firstly , finding “guanxi” could significantly improve the proportion of rural migrant students from public schools in urban areas. Secondly , after controlling for enrollment age and repeating grades in primary school , the probability of a rural migrant student being “over-aged” is significantly higher than that of a local student , regardless the rural student has used “guanxi” or not.

**Keywords:** left-behind children , “guanxi” , household registration discrimination , CEPS

**JEL Classification:** J13 , J16 , J18

( 责任编辑: 封永刚)