

农村学生“高分低录”的现象及原因

——基于宁夏高考数据的分析

陈 思, 丁延庆, 刘 霄, 伍银多, 叶晓阳

[摘 要] 农村学生“高分低录”是指高考分数相同的情况下, 农村学生比城市学生录取到分数线更低的学校。同其他国家类似, 这个现象在中国普遍存在。基于宁夏回族自治区高考数据, 本文对农村学生“高分低录”及其原因进行了实证分析。研究发现, 农村学生相对城市学生普遍“高分低录”; 这一现象在高分考生中更为明显。农村学生志愿填报存在“高分低报”现象, 这是导致“高分低录”的直接原因。农村学生缺乏对院校和专业信息的真实观察, 仅能根据院校和专业的名称等基本信息, 获知院校和专业的水平。信息的缺失造成了农村学生“高分低报”以致“高分低录”现象的出现。

[关键词] 教育决策; 高等教育入学机会; 高考志愿; 高分低报

一、引言

高考志愿是人生中最重要选择之一。对很多人来说, 这又是一个非常

[收稿日期] 2017-11-10

[基金项目] 国家自然科学基金 2016 年面上项目“‘教育致贫’与‘教育脱贫’”: 现状及相关政策评估”(71613013), 教育部人文社会科学重点研究基地北京大学教育经济研究所 2016 年度重大项目“经济新常态下的教育扶贫与教育公平研究”(16JJD880001)。

[作者简介] 陈思, 北京服装学院发展规划处; 丁延庆, 北京大学教育学院/教育经济研究所, 电子邮箱地址: yqding@gse.pku.edu.cn; 刘霄, 北京大学教育学院; 伍银多, 福建省宁德市蕉城区人民政府; 叶晓阳, 美国密歇根大学教育学院。

[致 谢] 感谢戴冰清、李伟、耿未、张燕以及宁夏教育厅、宁夏教育考试院、宁夏大学等诸多项目合作人员的努力使得本研究成为可能; 杨晋、孙冰玉、康乐、陆风鸣、李虹辉、王琼琼等项目成员的辛勤工作; 陈晓宇、丁小浩、哈巍、卢晓东、马丽萍、杨朴、岳昌君等老师的宝贵意见。

困难的问题：选择什么学校，什么专业，什么地区？学生需要充分考虑高考分数、个人和家庭偏好（如学费、职业发展等），以及所获取的目标学校和专业信息。在过去几十年里，随着个体偏好和各类信息的变化，以及高等教育机会和异质性的增加，高考志愿选择变得日益复杂。

公共政策的评估和最优设计需要去理解学生的这些复杂的高考志愿选择。一系列美国的大学选择研究揭示，大学学费、质量、距离以及生活设施是影响学生选择的重要因素（Manski and Wise, 1983; Long, 2004; DesJardins et al., 2006; Jacob et al., 2017）。然而，还有一些学生选择行为不能被上面的标准模型所解释。Hoxby 和 Avery(2013)发现美国的低收入高分考生并不报考那些包含奖学金后实际花费更低的精英私立大学，而选择花费更高的公立大学，因而产生了“高分低就”现象（Bowen et al., 2009; Smith et al., 2013; Dillon and Smith, 2016）。这种现象在中国同样存在，农村学生等相对弱势的学生，相对于其他学生而言，在高考分数相同的情况下，更容易被平均质量较差的学校和专业录取（Loyalka et al., 2016）。

高考制度虽然有亟待改进的空间，但仍然是一种相对公平的大学选拔机制。这种公平体现在高校录取学生中依据高考分数的一视同仁——“分数面前人人平等”。与此同时，严格的考试制度看似提供了一种机会公平竞争的途径，但实际上竞争的过程并非是绝对公平的。在同一院校、同一专业，农村学生的分数往往高于城市学生，即使城乡学生的高考分数相同，农村学生也比城市学生报考学校的录取分数线更低，造成了农村学生“高分低报”和“高分低录”的问题。农村贫困考生由于经济、文化、社会资源上的缺乏，往往倾向于报考一所考取机会大的学校，因此常常“牺牲”掉一些分数，也就是他们主动选择了“高分低报”或者“高分低录”。

在高考、志愿填报及录取的整个过程中，拥有较多社会资本的高、中级管理和专业人员家庭出身的学生，不仅没有浪费任何分数，反而能以远低于平均分的成绩进入热门专业。而农民家庭子女能享受的“优惠分”为零，甚至只有考出比城市家庭学生高 20 分的分数，才有机会进入大学（张杨波，2002）。农村学生上大学的难题也引起了教育主管部门的高度重视。为了帮助更多农村学生进入大学接受优质教育，我国政府采取了一系列向贫困地区的农村考生倾斜的专项招生政策。在国家、地方和高校不同层次设立了专项计划（余秀兰，2016），大范围地针对农村学生实施专项的优惠。但这些针对农村学生高考志愿填报的扶助政策的执行情况并不乐观。2015 年江苏省 11 所省属高校共投放了 550 个计划用于招收“地方农村专项”考生。但是在进行征求志愿后，所有 550 个计划最终只招收了 384 人，浪费达三成之多。甚至还

有部分学生认为通过“农村专项计划”进入高校后，自己会被贴上“农村学生”的标签，因此拒绝报考该计划。这些面向贫困地区定向招生专项计划虽然一定程度上促进了城乡学生在重点高校的比例均衡，实际上只是惠及贫困地区成绩优秀的学生，忽略了其他大多数考生的高考公平权益，极易造成贫困地区高考招生的两极分化现象(别敦荣，2015)。

通过调整全国招生计划增加农村学生上重点高校人数，确实收到了“立竿见影”的效果。但是政策的另两个预期目标，即提高贫困地区基础教育水平和吸引专项生回乡就业或创业，效果却不明显，或者说力有未逮。教育公平关系千家万户，也关系到国家的未来和民族的振兴，仅仅靠“补偿性”的专项政策远远不够也不尽合理。而农村学生的高考志愿填报，直接关涉到不同社会阶层接受高等教育升学机会的公平，是值得深入研究的问题。

在理论方面，国内关于农村学生填报高考志愿、获得高等教育机会及教育决策的研究，较多的关注城乡或不同阶层家庭的子女被录取的结果，即他们在接受何种层次和质量的高等教育。但这是以所有学生都在同等条件下填报高考志愿，并同样被录取为前提的。而对其分数是否都“用到刀刃上”，以及同样的分数对于填报志愿及最终录取结果的城乡差异则关注不足。同时忽视了影响城乡学生志愿填报差异的因素。本文力图全面考察农村学生“高分低录”的情况，并根据理性选择理论，比较影响高分农村考生志愿填报的因素。丰富高等教育入学机会公平、农村学生教育决策的案例和研究，对于制定教育改革政策与措施，为我国高等教育未来发展的方向提供参考。

在实践方面，高分农村学生“高分低录”引起的后果是极其深远的。如果高等教育学历能够作为通往较好职业的敲门砖，那么更好的学校或专业会让学生将来在事业方面更有优势。“高分低录”不仅限制了他们所接受高等教育的水平，也给他们个人及家庭带来极大的损失。这也使得高考这个看似公平的选才方式在当代中国社会转型中产生出了新的社会分层，带来不公平。因此，我们需要关注农村学生在高考志愿填报过程中的现实情况并了解其选择动机，为今后高等教育领域方针政策的制定提供理论与数据基础。这有助于学生未来在更为公平的制度中接受教育。

一个理想的高考志愿选择模型应该是，分数相同和偏好相同的学生应该被相同或者相似的学校、专业所录取。然而上述差异的出现，如果是农村、少数民族学生不会报志愿，而非报不同志愿所致，那么我们应当为这些弱势学生提供政策、项目辅助，增进他们的大学入学机会。在讨论什么的政策、项目可以对这些学生提供帮助之前，我们需要更加准确地理解他们进行高考志愿选择的过程。

二、相关文献

欧美国家的“高考”普遍采用申请制，学校在招生录取方面拥有较大的自主权，并且对学生采取更为多元化的评价方式：大学录取委员会不仅考察申请者的高中学业成绩，同时要综合考虑学生个人的特长、课外或社区活动、SAT 或 ACT 得分、AP(由高中提供的大学课程，可转成大学学分)等情况，来决定是否录取。一般来说，一名高中生可同时向多所大学提出申请，因此某个学生同时被数所大学录取的情形非常普遍。

尽管国外学生可以在升学时自由申请学校，高校拥有较大的招生裁量权，并且原则上允许学生被多个学校重复录取——相较于我国的制度设计来说，参与各方都有较大的灵活性——但是人们发现，在院校选择及录取上，依然会普遍存在类似中国学生“高分低录”或“高分低报”的“低匹配”(under-match)的现象，特别是在低社会经济背景人群中表现得十分突出(Smith et al., 2013)。

那么为什么高分、低收入学生不报考更好的大学呢？对这一问题的探索始于学界利用人力资本投资模型进行的解释(Manski and Wise, 1983; Ellwood and Kane, 2000; Avery and Hoxby, 2004; Long, 2004;)。人力资本投资模型对这个问题的贡献点在于，在学生院校选择的阶段，考虑到了“经济”对教育决策的影响，发现家庭收入、学费和经济资助都是导致“低匹配”偏差出现的重要原因(Terenzini et al., 2001)。然而 Ellwood 和 Kane(2000)对这一传统的视角提出了质疑，尽管人力资本投资模型能在一定程度上解释家庭收入和学习能力等因素对院校决定的影响，却无法回答为什么不同群体之间的选择会表现出明显差异。因此，利用人力资本投资模型，简单分析高等教育投资的经济成本和回报，对于高分、低收入学生为何不报考更好大学这一问题来说，给出的答案并不能令人十分满意。

学者们就此展开了更多的探索，研究目光也开始投向欧美以外的其他国家。概括来说，在个体层面上，导致学生院校选择行为出现偏差的原因主要包括学生学业成绩、父母受教育程度和家庭收入等；在制度环境层面，研究多指向学校质量(以排名衡量)、学生可能获得的经济资助(以及学费)等因素，它们也会在一定程度上影响到学生的院校选择(Manski and Wise, 1983; Niu and Tienda, 2008)。

在新近的研究中，学者们还考虑到了其他因素的作用。Hoxby 和 Avery (2013)验证了同等成绩的高收入和低收入家庭的学生在院校选择上存在显著

差异,进而提出并验证了“农村学生偏好学费低的学校”、“农村学生偏好省内学校”(home bias)以及“农村学生更加风险规避”这三个假设。他们的研究还发现,高分、低收入家庭的学生不报考更好大学,还可能是因为所在地区太小,因此无法进入更好的高中,与典型的高分学生群体有实质上的差异,他们得到毕业于优秀院校的教师指导的机率也更低。在对这些高分低收入家庭的学生的院校选择干预中,院校录取率、参观大学的校园访问及学校的培养机制和项目可能是无效的,因此建议有效的政策干预不应往高分学生过多的地区集中,在不同地区应有区分度地干预。

近期一系列研究对中国农村学生的高等教育决策进行了讨论。多数学者关注到了家庭背景对我国农村学生高等教育决策的影响。第一,由于家庭出身的不利,农村学生在进行教育选择时,无法获得较高的教育预期和信息。加之父母缺乏相关经历和知识,农村学生也无法从朋友、同学、学长等同伴中获得相应的支持,结果导致他们的升学选择主要基于其个体的认知,在升学准备、志愿信息收集和报考抉择中的不足(鲍威,2013)。第二,农村学生也并不是从高考那一刻才开始处于不利地位。大量研究表明,由于农村贫困学生在经济、社会资源上的缺乏,从幼儿教育开始,他们就面临着更多的学习困难和压力,在学业成就上相对城市学生表现出显著的弱势(Lucas,2001;刘精明,2014),他们更可能在层层考试选拔过程中被淘汰。再者,农村学生能够参与的重点高中竞争力就很低,而高中的教育水平又很可能影响是否能继续接受高等教育以及接受高等教育的类型。所以对于很多初三的农村学生,要么选择进入普通高中,要么初中就辍学。“筛选一回家”是农村学生教育选择的一种常见模式。同时,即使通过一层一层筛选的农村高中生,由于信息相对封闭,他们在高考填报志愿时经常表现出盲目性和随意性。第三,进行院校选择的主体也往往不是学生本人,而是他们的家庭。现实中客观的教育投资风险并不直接对家长们产生影响。农村家长获得外界信息的途径较少,对于子女专业的选择、院校的选择比较盲目,因而风险因素对家庭高等教育投资决策的影响都是通过家长的感知风险完成的(张学军,2008)。除了家庭的经济、文化资本的影响外,还有其他因素对农村学生上大学有较为显著的影响,比如子女的能力及学习态度、周边邻里对子女读大学的态度、拟报考高校的资助政策、读大学间接成本、高等教育投资风险、健康状况、医疗保险、智力、理解能力、外貌等(乐志强等,2014)。

上述各种因素都是造成“低匹配”现象的间接因素,高考志愿填报本身很可能才是导致“低匹配”的直接原因。从高考制度设计本身来看,导致“高分低录”现象出现的原因包括:(1)高考统一录取机制使得志愿填报成了考生的一

个显示偏好博弈,而这个博弈可能存在多重均衡或者没有纯策略均衡。国内一些学者讨论了高考志愿的填报制度的效率,以及学生对于高校的偏好问题(钟笑寒,2002;聂海峰,2006);(2)高考志愿填报的不同机制对录取结果的影响:考前报、估分报和知分报(钟笑寒、程娜,2004),“顺序志愿”和“平行志愿”(康乐、哈巍,2016)。

除了对于填报高考志愿这一制度的讨论外,也有学者讨论了高考志愿填报的影响因素。学者一般通过运用问卷、访谈等调查研究方法对考生进行数据收集,最后通过数据处理得出影响考生志愿填报的影响因素,主要有家长建议、个人爱好、学校名气、学校地域环境、学费、高考成绩、就业前景、家庭文化资本等(赵文波,1998;谢珊珊,2012;陈永华,2016)。特别地,居住地为农村的考生会更多考虑学费的影响。而且,社会和其他人因素对志愿选择的作用超过个人因素。农村考生比城镇考生更为自主,而对办学条件、学校知名度等考虑得较少(赵文波,1998)。

综合上述分析,学界对农村学生的志愿报考及录取过程中“高分低录”现象,能够给予的解释是:(1)不务实,没有把高考分数作为填报志愿的首要条件;(2)由于信息环境闭塞,与外界沟通不便,无法在网上浏览大量信息,缺乏专家指点等,导致对所报考院校的了解不够全面;(3)不能充分理解高考成绩的高低直接影响志愿的填报,不能以正确科学的态度面对报考志愿。

但是,由于缺乏高考学生的总体数据,学界尚未对农村学生“高分低录”的问题做较为全面的实证分析。而且,由于学生志愿填报真实数据的匮乏,现有研究并没有从高考志愿填报的“高分低报”这一可能导致“高分低录”现象的原因本身去解释“高分低录”的问题。在此基础上,本文将弥补学界由于数据缺乏所产生的研究缺陷,对农村学生的“高分低报”做重点分析,并尝试从该视角解释农村学生“高分低录”的现象,以求补充和完善对该问题的解释。

三、实证策略

(一)理论假设

本文尝试从学生个体、家庭、制度等不同层面,尽可能囊括现有文献涉及的相关因素,探析其对于学生志愿报考和录取的影响。参考 Hossler 等(1999)的观点,针对高等教育决策的问题,我们可以从以下三个视角来展开分析:首先,从经济层面的角度考虑,出于教育经济成本的考虑,农村学生(特别是收入更低的家庭子女)会更为注重接受教育所面临的经济开支,更倾向于选择离家更近的学校。其次,在社会文化层面,出于“以经济安全为第一

原则”，农村学生在志愿报考中更倾向于保守策略，力求报考和录取的稳妥；相应地，在志愿选择上农村学生倾向于选择与自己分数更为接近的学校。再者，目前的高考统一录取机制使得志愿填报成了考生的一个现实偏好博弈，学生在多数情境下所获得的信息是不完全的，因此农村学生得到的信息多寡和获取信息的途径，对于其院校选择有着重要的影响。本文提出以下假设：

(1)在院校录取结果上，农村学生“高分低录”的情形比城市学生更为严重。也即进入分数线相同或相近的院校，农村学生的分数会更高。

(2)在志愿填报环节，农村学生更倾向于以分数换取更大的录取把握，“高分低报”的程度更高。

(3)出于教育经济成本的考虑，农村学生偏好省内或者位于西北地区(离家更近)的高校。

(4)农村学生对大学信息掌握更少，因此更倾向于依据院校和专业名称等简单的信号来填报志愿。

基于以上假设，本文之后的分析主要包括三部分：一是基于最终录取结果，比较农村和城市学生在相同分数下进入大学的差异，估计并展示农村学生“高分低录”的现实状况。二是按照同样的思路，分析并展示农村学生“高分低报”的现象，这是导致“高分低录”的直接原因。三是就“高分低报”的现象，分析经济因素以及信息因素对城乡学生志愿填报差异的作用。

在分析的各个部分，我们都以农村学生与城市学生作对照，对以上假设进行检验。特别地，针对农村学生，我们也将分析其内部的异质性。这是考虑到：对于特定的农村学生来说，其志愿报考和录取结果也有一定的区别。那么哪些农村学生报考的学校会“更差”？他们有什么样的特点？对这些问题的回答可以进一步深化和检验之前的假设。本文将从民族、复读生、高考成绩等角度进行分析。特别是关于高考成绩，我们认为在高分农村学生群体中，“高分低录”和“高分低报”的问题可能更严重；而实际上，如果说高考改变农村学生命运的话，那也是针对高分农村学生而言，因此结合高考成绩的同质性分析无疑强化了本文的研究意义。

(二)数据

本研究所使用的数据来自宁夏回族自治区教育主管部门2016年的学生高考数据，这是一份非常难得和罕见的总体性行政数据，囊括了自治区全体高考学生样本。该数据包括五个部分：(1)考生报名库，包括考生考试号、考试类别、考生特征(例如高校专项计划资格、川区考生等，这些决定了考生是否可以报考某些专项计划，或者在不同程度上享受到高考加分优惠)、毕业高中、户口所在地等高考申请表的信息；(2)考生成绩库，包括考生各科目高考

成绩、总成绩,以及基于考生特征计算出的特征成绩;(3)考生志愿库,记录了考生所有填报的志愿信息。包括不同的批次、科类、不同的招生计划,以及相应的志愿。2016年的录取库数据中,共有8个招生计划、6个科类、6个批次;每个批次可以填报4个院校,每个院校可以填报6个专业;(4)考生录取库,记录了学生的志愿录取情况。包括实际录取的院校和专业的代号及名称,以及考生的志愿录取号(标识考生是被第几个志愿所录取)。同时,该数据库提供了学生最终录取的学校的投档成绩;(5)高校计划库,包括了所有在宁夏招生的高校当年的招生计划。其中每条计划按照批次、院校、考试科类、招生计划性质、专业来记录。共计有19378条招生计划。

2016年,全宁夏回族自治区共有69119名考生,农村学生的比例(58.33%)要高于城镇学生;少数民族学生(30.72%),特别是回族学生(29.81%)的比例非常高;在所有考生群体中,享受了高校专项计划资格(27.45%)、国家专项计划全国资格(36.93%)和地方专项计划资格(38.69%)的比例都非常高^①。这都反映出宁夏回族自治区的特殊性,也说明以宁夏回族自治区的学生高考数据为例,来进行“高分低录”和“高分低报”的研究,是非常恰当和有针对性的。

(三)概念界定

1. 城乡

本文实际关心的问题是,弱势群体或者说弱势阶层子女,在教育选择上是否也表现出一定的弱势?研究选取农村学生作为研究对象,考察高考志愿填报行为和录取结果上的城乡差别,从一个比较小的突破口探索了这一问题。作如此考虑的原因在于,城乡差别在中国是最主要的阶层差异之一。城乡二元分割的体制由来已久,对于中国社会分化和阶层流动有着深刻而重大的影响,对这一问题的讨论,有助于我们在教育仍然作为社会流动的主要途径的背景下,廓清弱势群体的现实处境,探求改善现状的政策进路。本文对学生城乡的划分是根据其户籍身份确定的,在高考申请表中对此有专门的记录。^②

^① “专项计划”是面向贫困地区定向招生专项计划,由普通高校招生计划中专门安排适量计划,面向集中连片特殊困难地区。

^② 需要说明的是,本文对城乡差别的分析和讨论,仅仅是建立于学生群体的某一基本特点,也可以移植于对其他群体类型差异的分析。例如,我们可以进一步关注贫困群体与非贫困群体、少数民族群体与汉族群体,以及更广泛的其它弱势群体与非弱势群体在该问题上表现出的差别。

2. “高分低录”和“高分低报”

本文所讨论的农村学生报考的学校更差，与一般文献中所说的“高分低报”密切相关。在院校录取结果上，与“高分低报”对应的即为“高分低录”。“高分低报”或者“高分低录”主要有四种不同的情况。本文所讨论的农村学生不报考更好学校的情况，是针对最后一种情形，即考生虽然考取了比较高的分数，但是因为某些原因牺牲部分分数，主动填报录取分数线更低的院校或专业。其他三种情况不在本文分析之列。^①事实上，由于任何学生最终投档的学校都有固定的录取分数线，所有该校招录的学生分数都不低于该分数线；从这个意义上来说，几乎任何考生都会存在不同程度的“高分低报”或者“高分低录”问题（除非“踩着线”被录取）。但是本文着重强调的是，农村学生在这个问题上会表现得比城镇学生更为明显和突出。也就是说，即便进入同一个大学或者专业，农村学生的分数整体上也要比城镇学生更高。

3. “更好的”大学：大学质量的测度

本文的讨论，涉及对农村和城市学生所报考学校“好坏”的比较。然而如何判断大学质量，是学界和实务界至今仍然争论不休的问题。由于大学的产出十分多元，因此并没有严格的单一的标准来对大学质量进行判断。本文选取“大学录取分数线”作为大学质量的测度。这有两方面的原因：一是，尽管大学产出多元，但是在学生报考的环节，很多学生并不知道大学的实际产出（例如科研成果、学生毕业起薪等），或者说最主要考虑的仍然是大学的录取分数线，这也是为什么多数高校都表现出对高分学生强烈兴趣的原因。二是，大学质量难以衡量，从学生的选择结果来估计学校质量不失为可行的办法，由于中国高校在录取时最重要的（甚至是唯一的）标尺仍然是高考分数，因此考生选取的大学质量能够在很大程度上通过其自身的考试分数自我揭露（self-reveal）出来。故而本文选择大学和专业的录取分数线，作为大学和专业质量的测度。

据此，我们可以将大学按照分数线作为质量进行排序。并且，学生分数

^① 其他三种情况为：（1）各省提档比例原则上超过了100%，上分数线的考生数总是高于被录取人数，以确保学校在录取学生时有一定余地，因此存在即便上线，也有很少的一部分学生落选的情况。这被称为“自然退档”。（2）对于是否“服从分配”一栏，大多数考生因怕不能被选报院校录取而分配到不愿去的院校或专业，没有勾选这一栏。在平行志愿规则下，考生一旦被提档而未被录取，将直接进入下一批次再次录取。这就是通常所谓的“滑档”到更差的学校。（3）考生对全省的招生情况不够了解，错误估计自己的考分水平或理想院校当年的录取分数线，导致从第一志愿落选，且如果此时第二、三志愿的院校又招满，则会出现高分落选的情况。

一定的情况下,所报考的大学录取分数线更高,我们就认为其报考质量更高,反之则报考质量更低^①。同时为了保证该测度的稳健性,本文也会使用特定高校特定专业在全自治区录取考生的平均分来进行分析。

各高校的分数线根据高考录取结果的总体数据计算得到。由于本文所使用的数据囊括了全自治区所有的考生,凡是被直接录取(不包括征集志愿)的考生都会被记录,并且会被记录其投档到该校的成绩^②。此外,该数据还提供了各高校在宁夏的所有招生计划。据此,我们可以比较方便地计算出各个高校最终在宁夏的录取分数线。按同样的办法,本文亦计算得到各高校不同专业的最低录取分数线、平均分。

(四) 计量模型

1. 录取结果:农村学生“高分低录”的证据

此部分先是采用描述统计的办法,从录取结果的角度,比较农村学生与城市学生表现的异同。更进一步的,本文采用 OLS 回归分析,考察农村学生和城市学生在录取结果上有何差异,以及此种差异在统计上表现是否显著。设定基本模型为:

$$y_i = \alpha + \beta rural_i + \gamma repeat_i + \theta rural_i \times repeat_i + \delta tdcj_i + \phi rural_i \times tdcj_i + \zeta controls_i + \epsilon_i \quad (1)$$

(1)式中,因变量 y_i 为考生“高分低录”情况的衡量。我们考虑四个不同的指标,包括投档成绩减去录取院校的最低录取分数、投档成绩减去录取院校的平均分数,投档成绩减去录取专业的最低录取分数、投档成绩减去录取专业的平均分数。这四个指标数值越大,表明学生花费的额外的分数越多,即“高分低录”表现得更为严重。分别以四个指标作为因变量进行回归。

自变量中, $rural_i$ 为是否农村学生的虚拟变量(以城镇学生为基底), $repeat_i$ 为是否复读的虚拟变量(以应届生为基底)。我们还加入了是否农村学生与是否复读的交互项,显然,该交互项刻画的是农村复读学生。 $tdcj_i$ 为学生的投档成绩(注意哪怕是同一个学生,如果投档到不同学校,其投档成绩也是不同的,这与实际考分有一定差异),在控制该变量之后,我们可以认为因变量的差异是来自于考试成绩以外的其它因素。

^① 这和“学校录取分数线一定的情况下,学生自身分数线越高,报考质量越差”本质上是一样的。

^② 由于不同学校招生计划、优惠政策不同,因此即便同一个学生,投档到不同学校的投档成绩和他本身的高考原始分以及特征成绩都不同。后文对此有进一步的说明。

上式中还加入了农村学生与投档成绩的交互项。这里暗含的假定是,农村学生“高分低录”的情况,可能存在与其分数的交互。前面分析中是将所有不同批次、不同成绩水平的学生都纳入进行讨论。但实际上,随着考生成绩的变化,“高分低录”的情况可能是有所差别的。这是由于,随着分数水平的提升,由于分数间隔增大,“高分低录”的空间也就随之变大;因此对于高分农村学生群体,“高分低录”的现象可能更为严重^①。

2. 志愿填报:农村学生“高分低报”的现象和特点

考生录取结果与志愿填报有直接关系,一个自然的解释是,“高分低录”来自于“高分低报”。在指出农村学生“高分低录”证据的基础上,此部分进一步分析农村学生在志愿填报上的特点。

本文首先通过描述统计和计量回归的办法,检验农村学生是否有“高分低报”问题的存在,以及这一表现在统计上是否显著。其中,亦要重点关注,是否存在随着考生成绩上升、“高分低报”问题更趋严重的现象。以此呈现农村学生“高分低报”的证据。

在此基础之上,进一步讨论城乡学生不同顺序志愿上的差异表现。结合录取结果,本文使用描述统计的办法,展示学生顺位录取的分布特点。进而采用 OLS 和 logit 回归,考察学生被不同顺序志愿学校、不同顺序志愿专业录取的情况。重点分析是否是由于农村学生在靠前的志愿上填报了分数线更低的学校和专业,才导致“高分低录”现象的出现。

3. 农村学生报考质量低下的原因

参考 Hoxby 和 Avery (2013) 的做法,此部分采用条件 logit 模型 (conditional logit model) 来具体讨论农村学生相比较于城镇学生报考质量更低的影响因素。此方法也广泛适用于大学选择文献中,如 Long (2004), Loyalka 等 (2016), Jacob 等 (2017)。

与 logit model 相比,条件 logit model 针对一个人的多种备选方案,根据每一种方案的特征变量,计算选择每一种方案的相对概率。以一批次本科志愿为例:一个高考分数在一本线以上的学生,理论上可以报考所有第一批次本科学校当中的其中任何一所大学,我们使用每一所大学对应的信息(如预测分数线、招生人数、学费、所在地区等)估计学生是否向该大学提交高考志愿

^① 虚拟变量,包括学生性别、民族、考试科类、批次、计划性质以及学生所获得的专项优惠等等。宁夏回族自治区 2016 年高考中,共有 9 个不同的批次、7 个考试科类;同时,招生计划包含了 14 个不同的性质类别。我们以虚拟变量的形式将其进行区分。另外,全国各地高考均有一定优惠政策,宁夏为省级民族自治地方,共有 17 种加分政策名目,不同加分政策可以在一定情形下累加,为此我们同样加入了各个优惠政策的虚拟变量。

的概率。每一个学生针对每一个大学都面临不同的选择条件,因而采用条件logit模型是合适的。模型基本形式为:

$$P(y_i = j) = \frac{e^{x_i' \beta_j}}{\sum_{i=1}^I e^{x_i' \beta_i}} (j = 2, \dots, J) \quad (2)$$

(2)式中, j 代表不同的方案, y_i 等于1或者0,1意味着最终选择了该方案。 x_i' 为方案影响因素的选择向量集合。 β_i 即为第*i*种院校(或专业)特征对学生填报该院校(或专业)的概率影响。其对数似然函数形式为:

$$\text{Ln} L_i(\beta_1, \dots, \beta_I) = \sum_{i=1}^n \sum_{j=1}^J d_{ij} \ln P(y_i = j | X) \quad (3)$$

为比较农村学生和城市学生在志愿填报上的差异,我们在模型中加入所有院校(或专业)特征与农村学生虚拟变量的交互项。我们主要报告原始系数(机会比对数, log odds),如果系数为正,说明该学生更倾向于填报具有该特征的院校(或专业);如果与农村学生虚拟变量的交互项系数为正,说明农村学生相对于城市学生更偏好于该特征。

四、农村学生“高分低录”的现象

如何衡量一个学生“高分低录”的程度?本文将根据宁夏回族自治区2016年所有学生的高考成绩及其录取院校结果,对每个院校在宁夏的录取学生分数进行排序,排序最后一位的学生分数即为该院校在宁夏的最低录取分数线,同理可计算得出各高校的各专业在宁夏录取的最低录取分数线。本文将考生分数与其被录取院校及专业的最低分数的差值作为测度,该差值越高,表明考生在高考志愿被录取时牺牲的分数越多,即表示该考生“高分低录”的情况越严重。

(一)“高分低录”的描述性分析

我们对所有学生按照考生城乡户籍和是否应届进行分类,包括城镇应届、农村应届、城镇往届和农村往届四种学生类型。分别计算这四个类型的考生分数与其被录取院校及专业的最低分数的差值,得到每个群体的差值的最大值、上四分位数、中位数、下四分位数和最小值的箱线图,图1报告了各类考生其分数与所录取院校的最低录取分数线的差值的基本情况。

结果显示,(1)农村学生高考分数与其被录取院校分数线的差值,要明显高于城镇学生;(2)往届学生,即复读生的分数与录取院校分数线的差值,要低于应届学生;(3)考生分数与被录取院校分数线差值最大的情况出现在农村复读生中,但是农村复读生整体被录取情况(中位数)优于农村应届生;(4)考

生分数与被录取院校分数线差值最小(即填报情况最好),以及整体被录取情况最优的情况均出现在城镇往届生。

根据本文的理论框架,考生分数与录取院校最低分的差值,反映了考生志愿填报结果的“好坏”,即有多少分数被“牺牲”掉。因此根据图中结果能大致推测出,农村学生志愿填报要劣于城镇学生,复读生则比应届生有所改善。

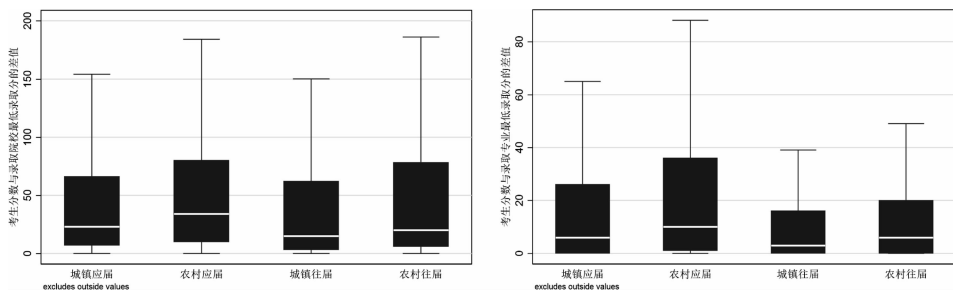


图1 考生分数与录取院校、录取专业最低录取分的差值

采用同样的办法,我们计算得出各高校不同专业在宁夏的最低分数线。因为同一院校的不同专业录取分数线参差不齐,因此相对于分数与院校录取最低分数的差值,考生分数与被录取专业的分数线差值就没有那么巨大。但结论与院校的分数差基本一致,且城镇学生优于农村学生、往届学生优于应届的趋势更为突出。

为了深入了解每类考生填报志愿时“牺牲”分数的情况,我们进一步讨论考生分数与被录取院校分数线之间相差的分数差值,并比较城乡学生的表现差异。图2绘制了考生分数与院校录取最低分相差分数的分布情况。我们将

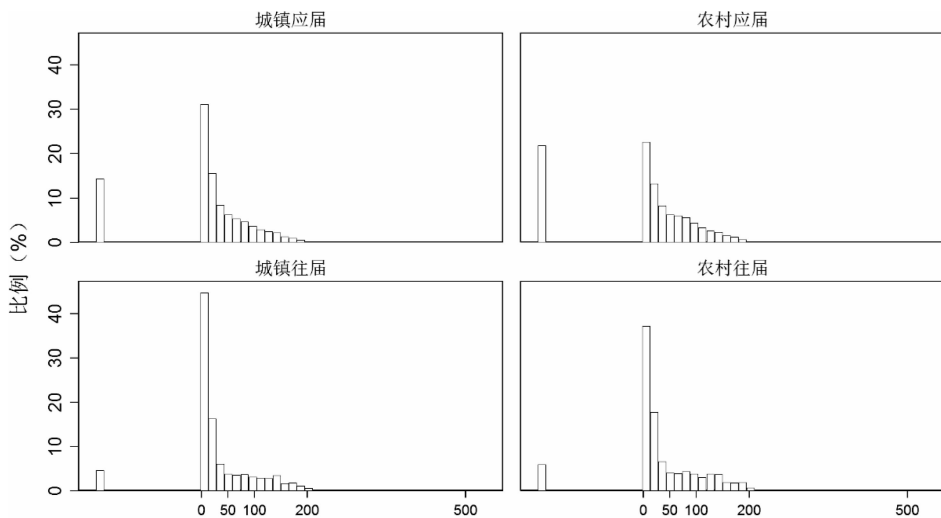


图2 考生分数与录取院校最低录取分的差值的分布

每个学生(即一个观测值)权重设定为1,并进行累计。纵坐标反映的即为差值为特定取值的学生比例。

值得注意的是,在2016年宁夏回族自治区所有参加高考的学生数据中,有一部分学生没有录取数据。排除行政统计的遗漏,笔者推测可能出于两个方面的原因:一是这些学生由于滑档、退档等原因最终没有被录取;二是部分学生参加了征集志愿,因此他们的录取结果不会反映在此数据中。这两类学生的分数相对较低,可以被认为是由于填报志愿过高,导致最终未能被录取。所以将这一部分学生群体单独绘制在图中最左边。

对图2的四个分数差值分布图比较分析,可以得到与前面分析类似的结论:(1)城镇学生在0点附近的分布更为集中,这意味着城镇学生被录取时普遍“牺牲”的分数较少,而农村学生相对更为分散,说明农村学生更容易出现“高分低录”的情况;(2)从图中最左边的最终没能录取的学生来看,这种未被录取的情况在农村学生中出现比例高于城镇学生;(3)复读生的表现要明显好于应届生,无论城乡学生皆是如此,均高度集中在0点附近。

(二)“高分低录”的实证检验

本节采用多元线性回归的方法,在控制其他因素不变的前提下,对描述性统计的发现进行实证检验。表1报告了回归结果,为避免冗长,略去了一系列控制变量的系数,保留了农村、复读、投档成绩等变量以及相应交互项。下表中,第(1)至(4)列没有加入农村与投档成绩的交互项,其因变量依次为投档成绩减去录取院校的最低录取分数、投档成绩减去录取院校的平均分数、投档成绩减去录取专业的最低录取分数、投档成绩减去录取专业的平均分数,以GAP1至GAP4作为指代。第(5)、(7)列因变量为投档成绩减去录取院校的最低录取分数;第(6)、(8)列为投档成绩减去录取专业的最低录取分数。第(5)和(6)列中控制了学生的班级固定效应;第(7)、(8)列则加入了投档成绩与农村学生的交互项。

首先,前四列的结果显示,农村学生系数为正,说明农村学生相比城镇学生,进入同样学校,要花费掉更多的分数(1.7—3.7分);在专业上,这一差值有所缩小,但也在1.6—2.3分之间。这证明了“高分低录”在农村学生身上表现显著。但是往届学生的系数却基本为正,意味着复读生需要比应届生考更高的分数进入同样的学校,即便考虑其与农村学生交互项的系数后仍是如此;特别是在第(7)和(8)列中其系数已不显著,这表明复读生报考得更好这一假设没有得到验证。

表 1 OLS 回归结果：城乡学生投档成绩与录取院校和专业的分数的差值

变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	GAP1	GAP2	GAP3	GAP4	GAP1	GAP3	GAP1	GAP3
农村学生	3.705*** (0.490)	2.263*** (0.287)	1.754*** (0.379)	1.619*** (0.254)	1.804*** (0.606)	0.826* (0.475)	-38.793*** (4.539)	-12.581*** (2.297)
往届学生	1.502** (0.698)	2.420*** (0.344)	-1.043** (0.423)	0.860*** (0.297)	-0.212 (1.247)	-1.584* (0.849)	-0.252 (0.505)	-0.314 (0.275)
农村×往届	-0.030 (0.854)	-1.859*** (0.424)	-0.951* (0.543)	-0.857** (0.370)	0.766 (0.975)	-0.044 (0.641)	0.472 (0.663)	0.330 (0.381)
投档成绩	0.256*** (0.005)	0.323*** (0.004)	0.064*** (0.004)	0.204*** (0.003)	0.268*** (0.005)	0.074*** (0.004)	0.074*** (0.005)	0.013*** (0.003)
农村× 投档成绩							0.077*** (0.009)	0.026*** (0.004)
少数民族	-0.981 (1.539)	-1.512 (0.950)	0.171 (1.005)	-0.793 (0.736)	1.306 (1.534)	0.738 (1.014)	0.098 (0.993)	-0.313 (0.469)
回族	6.000*** (1.561)	2.920*** (0.962)	2.335** (1.025)	2.284*** (0.749)	3.316** (1.564)	1.455 (1.040)	1.238 (1.016)	1.346*** (0.486)
N	57,935	57,935	57,216	57,216	57,935	57,216	10,540	10,540
R-squared	0.396	0.329	0.219	0.208	0.424	0.250	0.104	0.064

注：1. 括号中为标准误，下同；
2. *** p<0.01, ** p<0.05, * p<0.1, 下同。

为了增强结果的稳健性，第(5)和(6)列加入了考生的班级固定效应。这是考虑到，同一个班级的学生具有较强的同质性，所接触和接受到的高考志愿填报和录取的信息以及辅导都比较类似^①。回归结果表明，同一个班级里城乡学生“高分低录”的差距有所缩小，但是仍然显著存在。

从系数绝对值(0.064—0.323)来看，“高分低录”的现象不是很严重。但如前所述，随着分数水平的提升，农村学生“高分低录”可能表现得更为明显。为此，第(7)和(8)列加入了农村学生与投档成绩的交互项。需要注意的是，在总体数据中，我们提取出的是第一批次本科的子样本进行分析。回归结果显示，农村学生的系数为负，但是与投档成绩的交互项系数为正。这证实了上文的猜测。根据这一结果，本文绘制了图 3 所示的拟合的情况。

^① 需要说明的是，本文所使用的高考行政数据中，没有完善的考生班级记录，实际回归中是以“考号减去最末两位”作为班级的代理变量。这是因为现实中，考生申报参加高考一般是以班级为单位进行，也即同一个班级的学生考号往往编在一起。

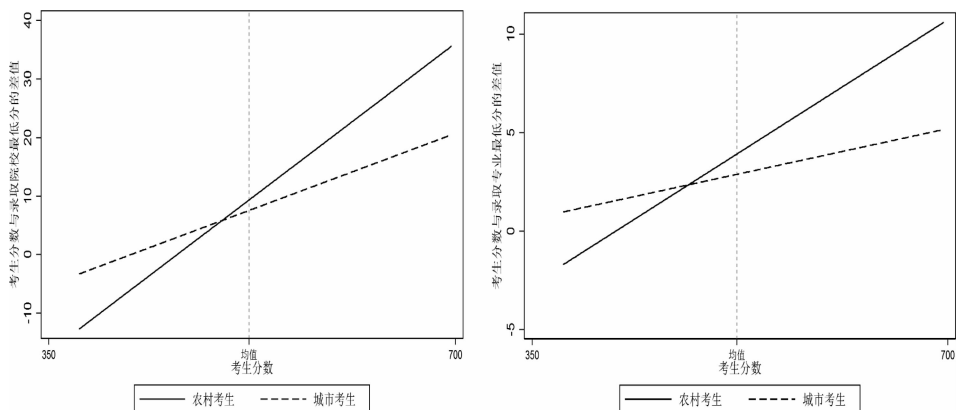


图3 考生分数与录取院校和录取专业最低分的差值随考生分数的变化

从图3可以看出,随着考生分数的增高,农村考生分数与录取院校最低分的差值越来越大,也就意味着出现越来越严重的“高分低录”现象。图中两条线的交汇点位于考生分数均值的左侧,但距离均值不远。因此不难理解,为什么之前的模型所得到的农村这一变量的系数较小——因为估计系数反映的是均值点的平均效应。即对于分数较低的考生,其分数差值不大,但随着考生分数的提升,农村学生的“牺牲分数”这一问题就变得更为严重。当然,即便被同一所大学录取,不同专业之间的录取分数线有时也相去甚远。为了增强前述结果的稳健性,我们采用专业的最低录取分数进行了相同的分析。所得到的结论基本一致,尽管系数和差值有所缩小,但是农村学生“高分低录”的情形依然突出,且随着农村学生分数的增加,其分数与其录取专业最低分的差值也越来越大。

五、“高分低报”——城乡学生志愿填报的差异

(一)农村学生“高分低报”的证据

前文的分析中,我们发现了农村学生“高分低录”的证据。但是,是什么样的填报行为导致了“高分低录”现象的出现?我们采用上文中同样的方法,把城乡学生考试分数与该学生填报的志愿院校及专业的最低分数线的差值作为衡量学生“高分低报”的测度,对其进行回归分析。针对上文的模型,我们将其因变量做了调整,分别替换为:考生总成绩减去志愿院校的最低录取分数、总成绩减去志愿院校的平均分数、总成绩减去志愿专业的最低录取分数,以及总成绩减去志愿专业的平均分数。

需要说明的是,不同于上文使用的学生被录取时的“投档成绩”,此处使

用的总成绩是指学生的“特征成绩”。所谓“特征成绩”，是指考生原始分数加上各类考生特征优惠(例如山区考生)之后的成绩。在宁夏回族自治区，有些特征政策仅限区内有效，因此又有“区内特征成绩”的说法。为了确保结果的稳健性，我们不考虑那些仅在自治区范围内有效的特征政策，也就是说把“区外特征成绩”作为学生的总成绩。注意这可能与实际的投档成绩是不一致的，当然这对于结果并无重大影响。

按照宁夏回族自治区高考规则，学生有不同的科类，每个学生理论上可以报考多个不同批次(如果分数允许的话)；这些批次包含在十来种招生计划内；在每个批次里，又可以报考多个志愿院校。2016 年的宁夏自治区录取库数据中，共有 8 个招生计划、6 个科类、6 个批次；每个批次可以填报 4 个院校，每个院校可以填报 6 个专业。因此实际组合情况达八千余种。为了降低数据分析的运算量，我们仅保留其中一批次本科的数据，并将学生填报的每一个志愿院校的每一个专业视为一个个案。共计得到 286642 条观测值。

表 2 中，GAP1 至 GAP4 为不同的因变量，分别为考生特征成绩减去志愿院校的最低录取分数、特征成绩减去志愿院校的平均分数、特征成绩减去志愿专业的最低录取分数，以及特征成绩减去志愿专业的平均分数。其中第(5)和(6)列加入了考生的班级固定效应；第(7)和(8)列加入了特征成绩与农村学生的交互项。

回归系数符号的方向以及显著性情况与上文中关于录取结果的估计十分接近。城乡学生“高分低报”的问题同样存在；同一个班级里城乡学生“高分低报”的差距有所缩小。但与上文情况不同的是，往届学生的系数始终保持显著为正，这意味着往届学生不仅在录取上表现出一定的“高分低录”的情形(或者至少没有表现出比应届学生更好的录取结果)；在志愿填报上，这一问题表现得格外突出。这验证了我们前文的发现，复读生因为填报志愿时可能出于更为保守的策略，导致他们更倾向于“高分低报”，进而导致了在录取环节的“高分低录”。

表 2 OLS 回归结果：考生特征成绩与志愿院校和专业分数的差值

变量	(1) GAP1	(2) GAP2	(3) GAP3	(4) GAP4	(5) GAP1	(6) GAP3	(7) GAP1	(8) GAP3
农村学生	4.601*** (0.222)	3.437*** (0.214)	3.147*** (0.223)	2.713*** (0.220)	2.250*** (0.254)	1.207*** (0.254)	-63.281*** (3.001)	-62.299*** (2.997)
往届学生	8.319*** (0.257)	7.565*** (0.252)	7.630*** (0.261)	7.313*** (0.258)	7.763*** (0.537)	7.777*** (0.545)	6.434*** (0.292)	5.816*** (0.296)
农村×往届	-3.142*** (0.297)	-2.739*** (0.291)	-2.667*** (0.302)	-2.582*** (0.299)	-0.825** (0.330)	-0.893*** (0.337)	-0.156 (0.341)	0.204 (0.346)

续表

变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	GAP1	GAP2	GAP3	GAP4	GAP1	GAP3	GAP1	GAP3
特征成绩	0.626*** (0.003)	0.630*** (0.003)	0.621*** (0.003)	0.629*** (0.003)	0.622*** (0.003)	0.615*** (0.003)	0.573*** (0.005)	0.570*** (0.005)
农村× 特征成绩							0.132*** (0.006)	0.127*** (0.006)
少数民族	-4.654*** (0.520)	-4.722*** (0.494)	-4.787*** (0.501)	-4.522*** (0.497)	-3.848*** (0.481)	-4.149*** (0.459)	-4.341*** (0.515)	-4.493*** (0.496)
回族	0.199 (0.527)	-0.590 (0.501)	-0.685 (0.509)	-1.346*** (0.504)	-0.416 (0.491)	-1.003** (0.469)	-0.071 (0.522)	-0.939* (0.503)
N	286,457	286,457	285,648	285,648	286,457	285,648	286,457	285,648
R-squared	0.557	0.574	0.544	0.557	0.630	0.618	0.563	0.550

关于特征成绩以及其与农村学生的交互项，我们得到了类似的回归结果。根据回归系数，绘制出了考生分数与志愿学校和专业最低分的差值随考生分数的变化情况，如图 4 所示。在志愿填报上，农村学生亦存在显著的“高分低报”情形，并且随着考生分数的增加，这一问题更趋明显。但是从斜率来看，城乡学生的情况比较接近，特别是在考生分数的均值附近，城乡差别不是很明显。从二者相差的绝对值来看，系数差异不大，“高分低报”的平均效应不严重，但随着特征成绩提升，二者差别更趋明显，即分数越高，农村考生的分数与其填报志愿的最低录取线的分差不仅越大，并且比城市学生增加的录取分差要更多。总而言之，本部分的实证分析提供了农村学生“高分低报”的证据。

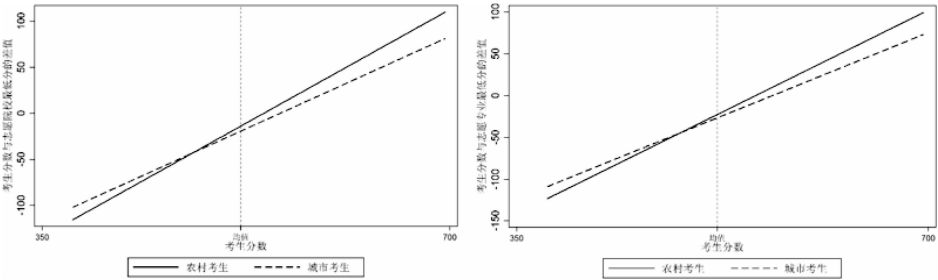


图 4 考生分数与志愿学校和专业最低分的差值随考生分数的变化

(二)城乡学生在不同顺序志愿上的差异表现

宁夏回族自治区 2016 年高考采取的是“平行志愿”的填报方式。根据志愿填报规则，在每个批次中，考生可以填报 4 所学校，每个学校可以填报 6 个不同的专业。高等院校按照 120%的比例进行提档。在 4 所学校层次，如果

学生被前面的学校提档,那么他将不会被其他学校提档;一旦最终未能被该校录取,那么他就会直接进入下一批次重新进行录取,这也就是通常所说的“滑档”。而在6个不同专业的层次,考生一旦被提档,就可以从前到后按照分数进行每个专业的录取,这一步与传统的顺序志愿录取方式一致。学生还可以选择是否服从专业的调剂。

本文首先考察考生被不同顺序的志愿学校录取的情况。图5绘制了城乡、应往届学生被4个不同顺序的志愿学校最终录取情况的分布。纵轴为被各顺序的志愿学校录取的比例。仅就应届学生来说,农村应届学生被第1个和第2个志愿录取的比例分别为34.65%和27.54%,而城镇学生大约为28.05%和25.88%。显然,在前两个志愿上被录取的比例,城镇学生比农村学生更低。

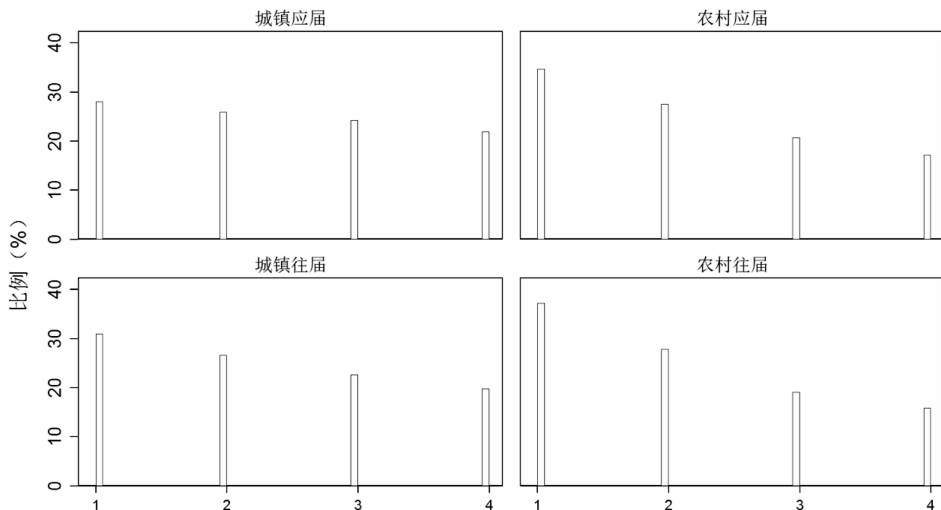


图5 考生被不同顺序志愿学校录取的情况

往届学生当中,城乡不同学生群体的志愿分布也呈现出类似的特点。这个有趣的结果反映出城乡学生填报志愿的策略差异。城市学生被靠后的志愿学校所录取,说明他们在前面的志愿上填报了分数线更高的大学;但是他们也在靠后的志愿上采取了较为积极稳妥的策略,填报了分数线相对较低的学校。这种“冲”和“保”的策略组合是高考志愿填报中非常常见的做法。但是显然,城市学生运用得更为普遍。因为对农村学生来说,他们在靠前的志愿位上就已经选择了更为“保险”的学校,故而反映在志愿学校录取的分布上,他们被靠前的学校录取的比例更高。

本文仅保留被第一顺序志愿学校录取的情况,考察城乡学生被不同志愿专业录取的情况。图6中横坐标依次为各个专业的顺序,纵坐标为被各个顺

序志愿专业录取所占的比例。比较城镇和农村学生的情况,亦可以发现城镇学生被靠前的专业录取的比例更低。以应届为例:城镇学生被第1专业录取的比例为23.46%,而农村学生被第1专业录取的比例为27.68%;但城镇学生被第6专业录取的比例为11.50%,农村学生仅为9.65%。专业填报上的“冲”与“保”策略的运用在城镇学生身上得到了更为明显的体现。

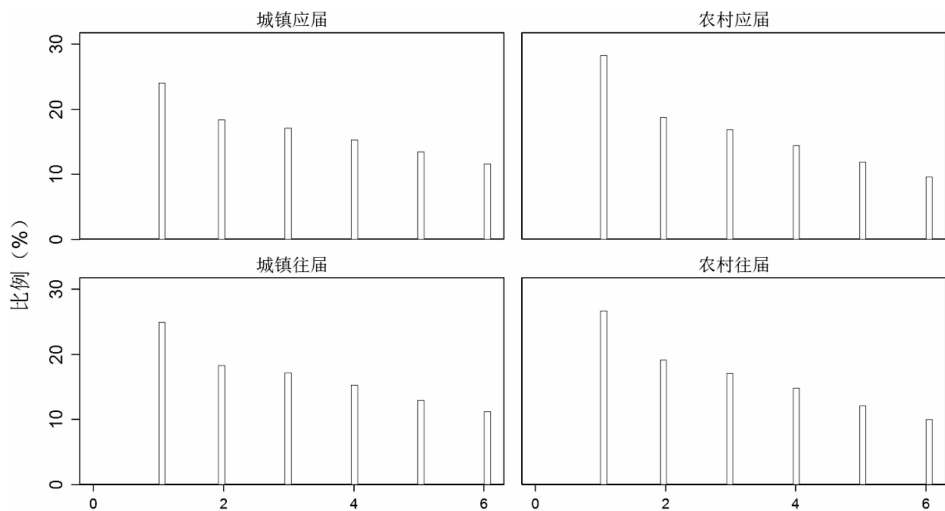


图6 被第一志愿录取的考生被不同顺序志愿专业录取的情况

为定量地说明城乡学生在不同顺序志愿填报上的行为差异。我们采取计量回归手段,控制住其他变量,对此种行为差异进行判断。模型设定基本如模型(1),具体形式如下:

$$y_i = \alpha + \beta rural_i + \gamma repeat_i + \theta rural_i \times repeat_i + \delta tdcj_i + \zeta controls_i + \epsilon_i \quad (4)$$

式中,因变量为考生被录取的学校和专业志愿的顺序。我们采取两种处理:一是,考生被录取的学校或者专业的志愿序号,处理为连续变量的形式。二是,将靠前的志愿序号编码为0,靠后的编码为1,也即对于学校志愿来说,第3、4位为靠后的志愿;对于专业志愿来说,第4、5、6为靠后的志愿。各个自变量的含义与模型(1)一致。

本文利用第一批次本科的子样本进行分析。表3报告了回归的结果。其中,第1和3列的因变量为考生被录取的志愿学校的顺序,第2和4列为被录取的专业的顺序。第1和2列因变量为顺序号码的原始数值,3和4列为重新编码的多元变量。相应地,对于前两列,我们使用采取稳健标准误的OLS估计,后两列使用logit模型估计。

表3 回归结果：考生被录取的志愿学校和专业顺序

变量	(1) 学校顺序	(2) 专业顺序	(3) 学校顺序(靠前=0)	(4) 专业顺序(靠前=0)
农村学生	-0.041*** (0.008)	-0.104*** (0.011)	-0.062*** (0.015)	-0.117*** (0.015)
往届学生	-0.027*** (0.009)	-0.035*** (0.013)	-0.038** (0.017)	-0.037** (0.017)
农村×往届	0.002 (0.011)	0.010 (0.016)	-0.005 (0.021)	0.007 (0.021)
特征成绩	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)
少数民族	-0.007 (0.017)	-0.028 (0.025)	-0.012 (0.032)	-0.019 (0.034)
回族	-0.028 (0.017)	-0.052** (0.025)	-0.046 (0.033)	-0.066* (0.035)
N	286,632	286,632	264,868	286,624
R-squared	0.074	0.087		

回归结果验证了之前的判断，农村学生的系数皆显著为负。这意味着，农村学生相比城镇学生，更多地是被录取到靠前的志愿学校和专业上。也就是说，在志愿填报上，农村学生在靠前的志愿位，选择了录取分数更低的学校和专业。由于实际的录取存在一定的顺序，故而，农村学生更可能被前面的录取分数线更低的学校和专业直接录取，错失了其他本可能“冲”的学校和专业。除此之外，回归结果也显示往届学生的系数都显著为负。复读生在志愿填报上与农村学生一样，都倾向于更为保守和稳妥的选择。这也就在一定程度上解释了，农村学生和复读生“高分低录”的结果。

六、城乡学生志愿填报差异的影响因素分析

志愿填报的差异，是导致“高分低录”问题出现的原因。那么，农村学生为何倾向于在一开始就填报分数线更低的学校呢？或者说，基于学生在填报志愿过程中并不知道同批次内各学校及专业录取分数线的前提下，是什么样的报考特点和动因，导致了同样的分数情况下，农村学生比城市学生报考的学校及专业的录取分数线更低的结果？对这个问题的回答，是理解城乡学生志愿报考行为差异、以至结果差异的关键。

(一)城乡学生在院校选择上的差异

我们首先分析在院校选择上,农村学生和城镇学生报考行为上的差异。之前的分析已呈现出的是,农村学生更倾向于将分数线更低的院校或专业作为自己的志愿首选。自然而然地我们会思考:对于农村学生来说,他们“牺牲”掉一定的分数,是为了得到什么?换句话说,这些学校必然有某种“特质”,能够满足农村学生一定的需求,才足以使他们选择这些学校,而不是分数线更高的其他学校。

根据理性选择理论(Rational Choice Theory),我们假设填报志愿的学生是一个理性的经济人,希望以最小的付出获得最大的回报,其院校选择的基础是建立在衡量高等教育给他们带来的教育成本、教育收益、升学失败的风险和社会地位提升这四个因素的前提之上。我们同时假定,由于学生,特别是农村学生事先缺乏对于高等院校和专业的充分了解,因此更倾向于从院校的名称来作出教育决策。在信息不充分的情况下,学校的名称一定程度上直接反映出学校的层次和水平,这一信号可以强烈地影响到学生的志愿填报。更具体地说,即便学校水平都差不多(分数线接近或处于同一批次),但是名称能够暗示着其办学层次和办学水平,或者名称暗示其能帮助学生获得更多的长期回报,则更有可能获得农村学生的青睐。

我们采用条件 logit 模型(conditional logit model)来分析学生的志愿填报情况。我们考察的院校指标包括如下几个方面:(1)院校名称是否包含“大学”二字,将诸如“学院”在内的学校作为对照组。相关研究发现,学校名称改为“大学”往往对学生更有吸引力。我们试图分析这对于农村学生是否作用更为强烈。(2)院校名称是否包含“国字头”的关键词,例如“中国××大学”、“中央××大学”,与前述逻辑类似,这样的学校可能同样吸引相当农村学生报考。(3)院校是否位于离学生更近的位置,主要考察的是是否包含“宁夏”、“银川”以及“西北”等关键词,这是考虑到我们所分析的学生群体为宁夏回族自治区考生,学生可能更倾向于报考位于区内或者西北地区的院校。(4)院校是否为 985、211 院校。毫无疑问,985、211 是能够一定程度上表征院校水平的强烈信号;即便非 985、211 院校也有相同的录取分数线,但我们猜测农村学生可能依然更倾向于选择 985、211 院校。

我们使用所有第一批次本科的志愿数据,并与学生的报名数据进行匹配。并去除掉其中定向类的志愿,剩余 52725 条记录。为了分析的便利,我们只考虑学生的第一志愿,这样得到 14412 条记录。学生填报的所有学校共计有 220 所,由于理论上学生可以报考这其中任何一所一本院校,因此除了学生实际报考的记录外,我们还为每个学生匹配了其他学校,即潜在的可能报考的院校,这样得到 3170640 条观测值。

我们以考生是否报考某个学校作为因变量,实际报考的为 1,没有报考的为 0。自变量中,除了考生特征成绩减去该校录取分数线的差值(这可以控制住学生成绩与院校分数线的差距)以外,我们依次考虑了前述院校特征,以及其与学生是否来自农村的虚拟变量的交互项。由于条件 logit 模型是以学生作为组(group)来分析学生的志愿选择,因此学生的其他变量和学校变量都不存在组内变化,在分析时这些变量亦会被自动舍去。故而我们没有加入其他的指标。

表 4 回归结果:院校特征与城乡学生的志愿填报

院校特征	系数	系数×农村学生
学校名称包含“大学”	-0.355*** (0.088)	0.990*** (0.171)
学校位于区内或西北	1.993*** (0.030)	1.020*** (0.039)
学校名称为“国字头”	-0.016 (0.041)	0.245*** (0.063)
学校为 985 院校	0.513*** (0.033)	0.595*** (0.044)
学校为 211 院校	-0.151*** (0.029)	0.218*** (0.042)

注:表中系数为回归估计得到的原始系数,没有计算出机会比。但是从符号的方向足以判断出城乡学生的志愿填报的显著差异。下同。

表 4 报告的回归结果显示,在所有特征上,城乡学生呈现出了非常显著的志愿选择的差异。具体来说,(1)城镇学生并不关心学校名称是否为“大学”,相反选择了一些名称为“学院”的学校;而农村学生对于这个特征高度敏感,更愿意报考“大学”而非“学院”。(2)无论是农村还是城镇学生,都倾向于报考区内或者西北地区的学校,且该名称条件在所有名称要素里的影响程度中遥遥领先;但是农村学生对此意愿更为明显。(3)城镇学生对于学校是否为“国字头”不敏感,但是农村学生更愿意选择带有“中国”、“中央”这类字号的学校。(4)985 院校得到了学生普遍的认可,农村学生表现尤甚;这可能是因为 985 院校分数超出一般本科院校更多。(5)学校是否为 211 院校,在城镇学生群体中表现不明显,这可能是由于 211 院校与许多普通一本院校分数线其实十分接近;但是农村学生仍然更倾向于选择此类院校。

回归结果验证了之前的猜测。农村学生的确更倾向于通过名字判断院校层次和办学水平,且他们比城市学生更倾向于选择具有以下“特质”的院校:

名称包括大学和“国字头”的、位于区内或距离家乡更近的、985 或 211 的院校。固然,这是学校实际水平的表征和直观的信号,但具备这些特质的学校的办学水平和层次不一定比不具备这些特质的同批次其他高校更好。从农村学生的报考结果来看,这些学校的录取分数线甚至还更低,农村学生为此付出了额外的分数。某种意义上来说,这是农村学生教育决策过程中付出的代价。

(二)城乡学生在专业选择上的差异

本文采取同样的思路,分析城市学生与农村学生在专业选择上的行为差异。并同样假定,农村学生之所以选择那些分数线相对更低的专业,一定程度上可能是因为这专业所代表的特点,或者说在专业名称上暗示的信号特点。但与上文分析不同的是,由于高考志愿填报中,学生可选择的院校和专业的数量都相当可观:在宁夏 2016 年所有志愿填报记录中,共计出现 571 所院校,每个学校都可能数十种不同专业,如果将所有专业都与学生进行匹配,数据量将会十分庞大。为此,本文仅考虑学生报考的院校在宁夏计划招生的专业。

由于着重关注的是农村高分考生,因此本文仅保留一批次本科志愿中非定向招生计划的志愿。对于每个学生,根据其报考的院校,将考生与该学校所有他可能报考的专业进行匹配,共得到 908728 条观测值。如此我们得到一个专业选择的集合,采用条件 logit 模型可以就专业特征和学生的志愿决策关系进行分析。

本文考察的专业特征包括如下几个方面:专业名称是否显示为经管类,是否为金融类,是否包含“国际”、“全球”等关键词,是否为计算机相关专业,是否为人文专业,以及院校名称是否包含“民族”、“农村”。在回归分析中,以考生是否报考某个专业作为因变量,实际报考的为 1,没有报考的为 0。自变量中,除了考生特征成绩减去该专业录取分数线的差值(这可以控制住学生成绩与专业分数线的差距)以外,本文依次考虑了前述专业特征,以及其与学生是否来自农村的虚拟变量的交互项。

表 5 回归结果:专业特征与城乡学生的志愿填报

专业特征	系数	系数×农村学生
经管类专业	-0.000	0.033**
	(0.010)	(0.014)
计算机类专业	0.464***	-0.264***
	(0.011)	(0.016)
金融类专业	0.641***	0.019
	(0.019)	(0.026)

续表

专业特征	系数	系数×农村学生
人文类专业	0.488*** (0.032)	0.389*** (0.045)
专业名称包含“民族”	-0.943*** (0.240)	-0.150 (0.349)
专业名称包含“国际”、“全球”	0.196*** (0.027)	0.130*** (0.041)
专业名称包含“农村”、“乡村”	-0.729*** (0.224)	0.320 (0.347)

表 5 报告的回归结果显示,在许多专业特征上,城乡学生的估计系数都有一定的不同。(1)城镇学生对于经管类专业不敏感,但是农村学生对此意愿更强烈。(2)城镇学生对于计算机类专业有较大的兴趣,但是农村学生兴致缺乏。(3)城镇学生相当青睐金融类专业,但是农村学生的交互项系数不显著,表明城乡之间没有明显差别。(4)城乡学生对于人文类专业都有比较强烈的兴趣,特别是农村学生,对此兴趣更为明显。(5)对于与“民族”相关的专业,无论城乡学生都不感兴趣,城乡之间差别亦不显著。(6)城乡学生都比较青睐名称包含“国际”、“全球”的专业,而农村学生更甚。(7)无论是城镇学生还是农村学生,都不倾向选择涉及“农村”、“乡村”的专业,城乡之间没有显著差别。

回归结果在一定程度上验证了之前的猜测。相较于城镇学生,农村学生更倾向于选择经管类专业和涉及“国际”或者“全球”的专业。这在一定程度上解释了他们与城镇学生志愿填报以及录取结果的差别。此外,计算机专业得到了城镇学生较强烈的关注,但是农村学生对此意愿较低,这可能是由于农村学生对此接触过少的缘故。在其他方面,城乡学生没有表现出明显的区别。

七、结论与建议

(一)研究结论

本文结合宁夏回族自治区 2016 年高考总体行政数据,就农村学生高考中出现的“高分低录”和“高分低报”问题,进行了分析和讨论。研究结论概括如下:

第一,“高分低录”现象在城乡学生、应届生和复读生中普遍存在。但是农村学生“高分低录”的情况比城市学生更为严重。农村学生平均要比城市学生多花费 1.7—3.7 分的分数,才能进入相同的学校;若要进入同一个学校的

同一专业,要多花费1.6—2.3分。而且学生的投档分数越高,其“高分低录”的态势越明显。此外,随分数的增加,农村考生的分数差值比城市考生的差值增加得更多。

第二,“高分低录”与城乡学生志愿填报行为的差异有关。无论是志愿院校还是专业,农村学生相比城市学生都存在突出的“高分低报”现象,且越是高分农村考生,越会倾向于“高分低报”,牺牲分数越多。另外,考虑被录取志愿的顺序,农村学生比城市学生更多地被录取到靠前的志愿院校和专业上。城市学生在志愿填报上有比较明显的“冲”与“保”策略组合的运用,因此更多地被靠后的志愿院校和专业录取。农村学生在志愿填报中趋于保守,因此被靠前志愿录取的比例更大。

第三,农村学生选择更为稳妥和保守的志愿填报策略,有经济和信息两方面的原因。我们发现,农村学生的确更倾向于通过名字判断院校层次和办学水平,更倾向于选择名称包括大学和“国字头”的、位于区内或距离家乡更近的、985或211的院校。也就是说,因为这个院校选择的倾向,导致农村学生报考的志愿不如城市学生,牺牲了他们的分数。在专业选择方面,农村学生更倾向于选择经管类专业和涉及“国际”或者“全球”的专业;人文类专业和计算机专业得到了城镇学生较强烈的关注,但是农村学生对此意愿较低。不难看出,无论是经管类专业,还是涉及“国际”或者“全球”的专业,都是当下的热门专业。农村考生填报志愿时,与其说是喜欢这些专业,不如说是为了将来的就业考虑。但是冷门专业,如人文类专业,或名称包含“农村”、“乡村”的专业(通常有经济优惠政策),或者农村学生较为陌生的计算机类专业,对于农村考生的吸引力则不大。但是这些农村学生眼中的热门专业并不一定就是更好的专业。相反,他们额外付出了高考分数,以期能稳妥地进入所谓的热门专业。

(二)政策建议

本文的发现表明,经济和信息两方面的因素造成了“高分低报”以致“高分低录”现象的出现,特别是农村学生关于大学和专业信息的缺失,是造成该问题的重要原因。和农村学生相比较而言,城市学生有丰富的信息资源和社会资源,经济负担没有农村考生那样明显,这样扩大了他们的选择余地。为此,本文提出如下政策建议:

首先,农村学生“高分低录”和“高分低报”的问题,已成为近年来学界关注的热点,并直接影响到城乡不同群体高等教育入学机会的均等。这在一定程度上源于现行高考招生制度下,志愿录取的平行志愿制度会让高分考生面临着相当的滑档、退档的风险。倘若改革目前的高考平行志愿录取模式,为学生提供更为充分的考虑空间,以及更为安全的志愿填报保障的话,那么也

许农村学生可以更好地作出选择,填报与自身分数更为匹配的学校和专业。

其次,结合本文的研究发现,农村学生高分低报,相当程度上来源于对大学和专业信息的匮乏。正因如此,他们更为依赖院校、专业名称等十分简单的信号,以之作为院校和专业水平的判断。相比之下,城镇学生对大学的信息掌握得更为专业。因此,教育主管部门有必要为农村学生提供更充分的关于大学、专业的信息,包括院校和专业的质量、办学情况、就业情况等等。

课题组在田野工作中,发现很多农村考生接收到的来自老师、家长或者同学朋友的建议,很多也只是象征性的,比如“女娃应该留在我们(家长)身边上大学,以后在银川工作”,“父母希望我报个稳定的、出来能进国家单位的学校”,“父母希望我能够找到一所好的大学上学,选一个热门专业,因为农村很看中一份好的工作”,但实质性内容并不多。除非家长对子女报考院校及专业有明确的目标和倾向,其实最后往往还是由学生自己拿主意。面对互联网上鱼龙混杂的院校信息,掺杂着纷繁的高考志愿咨询广告,缺乏判断能力和理性报考指导的农村考生,也只能通过院校名字、离家近不近、是不是985或211、专业是否流行等直观的信息来做出判断。

最后,在现行的平行志愿录取制度之下,考生“冲”与“保”的策略组合及运用,可以在不明显扩大滑档、退档风险的前提下,有效提高其分数与院校、专业的匹配度。但是农村学生对此运用得相对有限。因此,有必要对农村考生志愿填报给予一定的信息干预和辅导。

[参考文献]

- 鲍威,2013:《第一代农村大学生的升学选择》,《教育学术月刊》第1期。
- 陈永华,2016:《高考改革背景下高校提高生源质量的对策——基于影响考生志愿填报的关键因素调查分析》,《辽宁教育行政学院学报》第6期。
- 别敦荣、韦莉娜、唐汉琦,2015:《高等教育治理体系和治理能力现代化的基本原则》,《复旦教育论坛》第3期。
- 丁小浩、梁彦,2010:《中国高等教育入学机会均等化程度的变化》,《高等教育研究》第2期。
- 康乐、哈巍,2016:《高考志愿填报改革对录取匹配质量的影响》,《北京大学教育评论》第1期。
- 乐志强、卢曼萍、周杨平,2014:《农村家庭高等教育投资决策影响因素实证研究——以江西省为例》,《高等农业教育》第3期。
- 李春玲,2003:《社会政治变迁与教育机会不平等——家庭背景及制度因素对教育获得的影响(1940—2001)》,《中国社会科学》第3期。
- 李煜,2006:《制度变迁与教育不平等的产生机制——中国城市子女的教育获得(1966—2003)》,《中国社会科学》第4期。

- 刘精明, 2014:《能力与出身:高等教育入学机会分配的机制分析》,《中国社会科学》第8期。
- 聂海峰, 2007:《高考录取机制的博弈分析》,《经济学(季刊)》第3期。
- 文东茅, 2005:《家庭背景对我国高等教育机会及毕业生就业的影响》,《北京大学教育评论》第3期。
- 谢珊珊, 2012:《场所依恋对高考志愿填报的影响》,曲阜师范大学硕士学位论文。
- 杨东平, 2006:《高等教育入学机会:扩大之中的阶层差距》,《清华大学教育研究》第1期。
- 余秀兰、白雪, 2016:《向农村倾斜的高校专项招生政策:争论、反思与改革》,《高等教育研究》第1期。
- 张杨波, 2002:《社会分层与农村学生受教育机会不平等——家庭经济、社会背景对农村考生高考填报志愿的影响》,《青年研究》第11期。
- 张学军, 2008:《农村家庭高等教育投资决策研究》,西北农林科技大学硕士学位论文。
- 钟笑寒、程娜、何云帆,《花落谁家——高考志愿填报机制的博弈模型》,《经济学(季刊)》第2期。
- 赵文波、傅志刚, 1998:《影响高考志愿填报的因素及成因分析》,《浙江师范大学学报(自然科学版)》第2期。
- Avery, C. N., M. E. Glickman, C. M. Hoxby and A. Metrick, 1998, “A Revealed Preference Ranking of U. S. Colleges and Universities”, *Quarterly Journal of Economics*, 128(1): 425—467.
- Becker, R., 2003, “Educational Expansion and Persistent Inequalities of Education: Utilizing Subjective Expected Utility Theory to Explain Increasing Participation Rates in Upper Secondary School in the Federal Republic of Germany”, *European Sociological Review*, 19(1): 1—24.
- Hoxby, C. and C. Avery, 2013, “The Missing ‘One-Offs’: The Hidden Supply of High-Achieving, Low-Income Students”, *Brookings Papers on Economic Activity*, Spring (1): 1—50.
- Desjardins, S. L., D. A. Ahlburg, and B. P. McCall, 2006, “An Integrated Model of Application, Admission, Enrollment, and Financial Aid”, *Journal of Higher Education*, 77(3): 381—429.
- Ellwood, D. and T. J. Kane, 2000, “Who is Getting a College Education? Family Background and The Growing Gaps in Enrollment”, Danziger, S., & Waldfogel, J. (Eds.) *Securing The Future: Investing in Children From Birth To College*, 283—324. Russell Sage Foundation, New York, NY.
- Hossler, D., 1999, “Using the Internet in College Admission: Strategic Choices”, *Journal of College Admission*, N/A.
- Jacob, B. A., B. P. McCall and K. Stange, 2017, “College as Country Club: Do Colleges cater to Students’ Preferences for Consumption?” *Journal of Labor Economics*, forthcoming.

- Long, B. T. , 2004, “How Have College Decisions Changed over Time? An Application of the Conditional Logistic Choice Model”, *Journal of Econometrics*, 121 (1—2): 271—296.
- Loyalka, P. , Y. Qu, B. Wu and X. Ye, 2016, “How Disadvantaged Students Make Inefficient College Choices under a Centralized Admissions System?”, *Working paper*.
- Manski, C. F and D. A. Wise, 1983, *College Choice in America*, Harvard University Press.
- Niu, S. X. and M. Tienda, 2008, “Choosing Colleges: Identifying and Modeling Choice Sets”, *Social Science Research*, 37(2): 416—433.
- Raftery, A. E. and M. Hout, 1993, “Maximally Maintained Inequality: Expansion, Reform, and Opportunity in Irish Education”, *Sociology of Education*, 66(1): 41—62.
- Lucas, S. R. , 2001, “Effectively Maintained Inequality: Education Transitions, Track Mobility, and Social Background Effects”, *American Journal of Sociology*, 106 (6): 1642—1690.
- Smith, J. , M. Pender and J. Howell, 2013, “The Full Extent of Student-college Academic Undermatch”, *Economics of Education Review*, 32, 247—261.
- Terenzini, P. T. , A. F. Cabrera and E. M. Bernal, 2001, “Swimming against the Tide: The Poor in American Higher Education. Research Report No. 2001-1”, *College Entrance Examination Board*, 1—62.

College Undermatch of Rural Students: Evidence from Ningxia Gaokao Data

CHEN Si, DING Yan-qing, LIU Xiao, WU Yin-duo, YE Xiao-yang

Abstract: College undermatch describes that rural students tend to be admitted by colleges with lower admissions scores than urban students do, even conditional on having same *Gaokao* scores. Like in other countries, undermatch exists widely in China. Using population *Gaokao* data from Ningxia Hui Autonomy District, this paper conducts detailed analyses on the causes and consequences of rural students' undermatched college choices. We show that there is a remarkable rural-urban gap in college undermatch admissions, particularly among high achieving students. The direct reason is that rural students are more likely to apply for colleges with lower admissions scores, as they lack information about colleges and majors.

Key words: Educational choice; college access; *Gaokao*; undermatch

(责任编辑: 杨 娟 责任校对: 杨 娟 胡咏梅)