

数字鸿沟与家庭教育投资不平等

杨 钊 徐 颖

(北京大学 教育学院, 北京 100871)

摘 要 本研究旨在探究在互联网加速普及的背景下,家庭信息资本向人力资本投资的转化。本文利用“中国家庭追踪调查”2014 年追踪数据,建立配套亲子样本,应用泛精确匹配和加权回归方法来识别母亲互联网使用与家庭教育投资的关系。研究发现母亲的互联网使用显著地提高了家庭教育投资,但其效果具有异质性:城镇家庭是受益者,母亲的互联网使用对进城务工家庭和农村家庭教育投资的影响不显著。在影响机制上,互联网使用可通过降低交易成本和解决信息不对称来促进教育投资。我国城市信息优势阶层已经成功地将信息资本转化为人力资本投资,但是缩小数字鸿沟未必能弥合不断扩大的城乡教育鸿沟。

关键词 数字鸿沟;家庭教育投资;互联网使用;不平等

中图分类号:

文献标识码:

文章编号:

DOI:

一、问题的提出

互联网为个人创造了更多的就业机会、提高了劳动生产率,并为消费者带来了福利(世界银行,2017)。然而,社会各群体使用互联网的机会并不均等。欧美国家 90% 以上学生可以在家使用计算机完成作业或在家上网,仅有 52% 的墨西哥学生可以在家上网学习(Bulman & Fairlie, 2016)。互联网接入、技能和使用方面的差距被研究者称为“数字鸿沟”(Dimaggio & Hargittai, 2001; Golder & Macy, 2014; Norris, 2001; 卡斯特, 2001; 闫慧, 孙立立, 2012)。2000 年以来,中国政府利用“数字机遇”成功地缩小了本国与发达国家、本国城乡之间的两大数字鸿沟(胡鞍钢, 王蔚, 周绍杰, 鲁钰锋, 2016)。截止 2016 年 12 月,中国网民规模达到 7.31 亿人,互联网普及率为 53.2^①。

收稿日期:2017-09-05

作者简介:杨钊,女,北京大学教育学院副教授,博士。

徐颖,女,北京大学教育学院硕士研究生。

^① 中国互联网信息中心:《第 39 次中国互联网络发展状况统计报告》,2017 年 1 月 22 日, http://www.cac.gov.cn/2017-01/22/c_1120352022.htm。

整体数字鸿沟的弥合并并不意味着不同社会群体能同等程度地从中受益。根据2014年“中国家庭追踪调查”(以下简称CFPS)数据计算,收入最低的20%家庭中母亲的互联网使用比例为13.3%,收入最高的20%家庭中该比例是前者的3.6倍。母亲与子女之间的互联网使用密切相关。75.6%的使用互联网母亲的子女也上网;不使用互联网母亲的子女的上网比例仅为39.5%。可见我国互联网普及中的家庭差距仍然存在,且代际不平等初现端倪。

更值得忧虑的是,数字鸿沟有可能转化为家庭教育投资的鸿沟,从而强化教育的代际不平等。我国使用互联网家庭在教育投资方面呈现显著的优势。CFPS2014数据中使用和未使用互联网家庭的课外辅导班参与比例分别为35.8%和11.3%;这两组家庭的每周课外课程辅导时间分别为2.47小时和1.14小时。在投资规模方面,两组家庭的年教育支出分别为5045元和2175元;课外补习年支出分别为2113元和382元。比较CFPS2010和2014年数据,还可以发现使用互联网家庭在教育投资方面的优势并没有随着数字技术的普及而缩小。

学术界尚未系统分析互联网普及对家庭教育投资的影响。一方面,互联网有可能弥合家庭间的教育投资差异、缩小个体在学历文凭方面的差距,促进社会流动;另一方面,接入和使用互联网方面的差异也有可能扩大家庭教育投资差距,从而造成教育机会和成就方面的差异、加剧社会分层。数字鸿沟是否加剧教育投资的鸿沟?信息资本不平等是否会强化社会分层?本文借助CFPS2014数据分析以下三个问题:第一,互联网使用是否影响家庭教育投资?第二,互联网对不同群体的影响是否存在差异?第三,互联网通过哪些机制影响家庭教育投资?

二、数字鸿沟与教育不平等

(一) 知识鸿沟与不平等的再造

国内外经济社会学研究特别关注数字化技术普及过程中的“知识鸿沟”效应,即不同社会经济地位群体接触新技术的时机与受益程度的差异(Blau, 1977; Tichenor, Donohue, & Olien, 1970)。对新技术普及过程的研究指出,社会经济背景优势阶层在新信息的探索方面也具有优势。这种优势体现在接触信息、使用信息和利用信息获利三个方面。由于优势阶层独特的社会地位,他们可以首先发现新信息。又由于该阶层具备的收入优势,他们有能力优先购买新信息。学校教育为优势群体提供了认知方面的优势,使得他们能够更有效地处理新信息,该群体对新信息投资的回报也更高(Donohue, Tichenor, & Olien, 1975)。高社会经济地位群体不仅习得更多知识,而且由于他们在新信息接入和使用方面的优势,不同群体间的知识差距会不断扩大。

互联网普及产生的知识鸿沟可能会加剧社会分层(Dimaggio et al., 2001)。卡斯特率先提出了信息鸿沟的概念,指出网络拥有者和未拥有者之间的差异可能扩大不平等、增强社会排斥(Castells, 1996)。作为早期研究的集大成者,迪马乔等(Dimaggio & Hargittai, 2001; Dimaggio, Hargittai, Celeste, & Shafer, 2001)将互联网技术对社会的影响分为两大类:一是对经济福利的影响,包括教育机会与成就、求职和劳动力市场表现、消费者福利;二是对社会参与的影响,包括政治参与、社会活动参与、政府服务和公共产品的接受等。从结构主义的视角进行分析,不同社会群体接触和利用互联网技术的差距在很大程度上与他们的社会阶层相关;由于各群体的信息差距,他们将信息资本转化为其他资本的能力呈现很大差异,转化能力差异可能导致社会各阶层之间的社会经济地位差距进一步扩大(李升, 2006)。

从互联网接入与使用差距到社会分层扩大是一个连续的过程(参见图1)。首先,社会经济背景决定了个体接触和使用信息技术的可能性,决定了信息资本积累方面的差距,导致了“信息富有者”和“信息贫困者”的出现(Haywood, 1995),即出现数字鸿沟。其后,凭借其信息优势,社会优势阶层可以通过多种渠道强化自己的社会经济地位,强化社会分层。这表现在三个方面:一、优势阶层利用其信息优势,获得社会经济地位较高的职业和较高的收入,从而强化其经济资本优势,即实现“经济资本—信息资本—经济资本”的再生产;二、基于信息优势,优势阶层率先投资于人力资本的生产和再生产,提高本人和子女的教育成就,即实现“人力资本—信息资本—人力资本”的再生产;三、优势阶层利用其处理新信息的认知优势,实现信息资本与社会资本等的转化,即实现“社会资本—信息资本—社会资本”的转化。无论是通过哪种途径,信息资本的差距都有可能转化为社会经济地位的差距,强化了原有的社会分层。

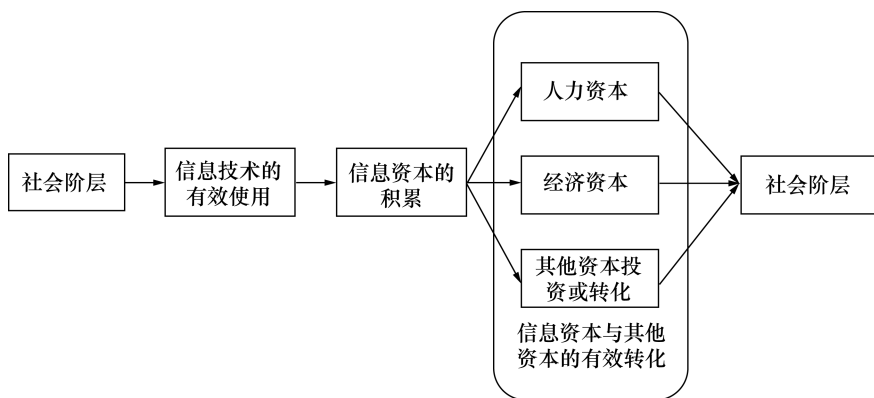


图1 数字鸿沟与不平等再造过程

以往对数字鸿沟的研究聚焦于数字鸿沟的内涵、表现维度和影响因素(闫慧, 孙立立, 2012)。有鉴于信息技术与社会分层的密切关系,大量实证研究也

讨论了信息资本与人力资本、经济资本和社会资本的转化。尤其是随着信息技术对教育领域的渗透,有关信息资本与人力资本转化的研究也不断涌现,聚焦于在学校和在家使用计算机与互联网对教育成就的影响。

例如,研究者利用教育生产函数模型估计了计算机网络对个体成绩的影响,发现其对教育产出的影响较为复杂。一方面对计算机技术的经费投入可能挤出其他更有效或者无效的教育投入(如教师培训),另一方面用于计算机互联网教学的时间可能替代有效或者无效的传统教学时间或者家庭教育活动。因此,难以判断计算机网络对学生成绩的净影响(Bulman & Fairlie, 2016)。我国研究也表明互联网使用对学生成绩产生了混合影响(陈纯槿, 邳庭瑾, 2016; 程千里, 张炜华, 2016; 陈纯槿, 顾小青, 2017)。综上所述,互联网使用未必能直接提高以标准化考试成绩来测量的人力资本积累的数量和质量,且互联网普及率的提高未必缩小小学生群体中的学业差距,甚至有可能加剧教育不平等。

(二) 家庭教育投资差距

优势阶层不仅垄断教育资源,而且可以通过持续的高额教育投入,不断巩固其教育优势,即将经济资本和社会资本转化为人力资本,再将人力资本转化为未来的经济和社会资本,实现社会阶层的固化。因此在发展中国家,教育不平等不仅表现为教育机会和学业成就的差距,更体现在家庭教育投资的群体差异。家庭教育投资包括家庭在校内和校外所进行的投资,其中既包括必要性投资(如交通费),也包括选择性投资(如择校费)。研究者一般用家庭教育支出来衡量家庭的教育投资。

义务教育阶段的家庭教育竞争集中在校外教育投资方面。对东亚国家的研究发现,各国家庭教育投资数量庞大,家长的教育投资意愿比较强烈。当免费的公共教育质量不能让人满意,公立学校的师资水平、硬件设备条件比较差的情况下,来自家庭的额外教育支出会不断增加,择校、择教、校外补习的现象会持续出现(Bray, 2014; Bray, Ding, & Huang, 2004; Bray & Kobakhidze, 2014; Bray & Kwo, 2013; Mcvey, 2012)。中国家庭在教育投资方面的分化尤为明显。利用国家统计局2007年和2011年的城镇居民教育情况调查数据,迟巍等的系列研究发现,家庭教育投资的增长很大程度上取决于校外支出增长。教育投资随着家庭收入的增加而不成比例的变化,与高收入家庭相比,低收入家庭的教育投资规模较低,但是教育支出占家庭收入的更大比例(Chi & Qian, 2016; 迟巍, 2013; 迟巍, 钱晓焯, 吴斌珍, 2011)。丁小浩和薛海平(2005)利用城镇居民教育情况调查数据的分析发现,城镇居民家庭的义务教育负担随着家庭收入的上升而下降。

值得注意的是,农村家庭的教育投资意愿和行为与城镇家庭有较大的差别。殷红霞(2010)和张锦华等(2014)发现,农村家庭根据预期的收益和成本来进行投资决策,由于家庭收入和投资收益的不确定性、收入水平过低或者投资

成本过高,农村家庭往往教育投资能力不足、投资水平较低。龚继红和钟涨宝(2015)的分析表明,随着农村家庭收入提升,教育投资相应扩大,可见收入是投资的重要约束。农村家庭经历城乡迁移后,教育观念可能变化,并相应改变教育投资行为(杨振宇和张程,2016)。最新研究还发现农村义务教育政策——如撤点并校等会显著增加教育支出(潘光辉,2017)。由此可见,成本、收益、家庭收入、教育观念、政策等是农村家庭教育投资的影响因素。

家庭教育投资显著影响子女的教育选择和学业成就,投资差距可能导致教育不平等的加剧。李湘萍(2008)基于北京师范大学教育学院2006—2007年“义务教育阶段家庭教育支出调查研究”的调查数据,分析了义务教育阶段择校行为与家庭背景的关系。相对于不择校家庭,择校家庭的社会经济地位与子女就读学校水平之间的正相关关系更强。梁琦(2012)的分析表明家庭教育支出的资金投入(特别是校外资金投入)、时间投入显著正向影响子女考试成绩的变化。由此可见,我国家庭教育投资的差距在扩大,这种教育投资方面的差距可能转化为子女人力资本数量和质量方面的差异。

(三) 已有研究评析

互联网使用与人力资本转化的研究,以及家庭教育投资不平等研究尚存在一些问题。第一,已有研究关注计算机网络接触和使用对教育成就的影响,忽视了其对教育投资的影响。第二,以往研究过于关注互联网使用对学生成绩的直接影响,忽略了互联网可能通过家庭教育投资而产生的间接影响,尤其是通过择校或者课外补习来改变学生成绩。第三,在中国特定的社会背景下,父母是家庭教育投资的主体(Bray et al., 2004; Chi & Qian, 2016; 丁小浩 & 翁秋怡, 2015; 李湘萍, 2008)。若要考察互联网的人力资本投资效应,应该关注父母亲的互联网使用对子女教育投资的影响。第四,以往对中国家庭教育投资的研究较为狭隘的聚焦于社会阶层、儿童特征和能力、地域、城乡等对总教育投资规模的影响,较少关注信息技术使用对家庭教育投资的影响。信息资本差异已经逐步成为社会分层的新来源,对家庭教育投资差距的讨论应纳入对父母互联网使用情况的考察。第五,我国政府致力于弥合城乡之间的数字鸿沟。这种做法是否能缩小城乡家庭教育投资差距尚未得到系统性的分析。

有鉴于此,本文拟填补上述研究空白。本文讨论了母亲互联网使用情况对家庭教育投资规模和行为的影响,并考察了信息资本转化为人力资本投资过程中是否存在城乡、户籍等维度的差异。本文采用的泛精确匹配方法在一定程度上控制了母亲互联网使用的内生性。本文还进一步讨论了互联网投资拉动效果的异质性与稳健性,并检验了互联网使用影响家庭教育投资的不同机制。

三、研究假设与实证策略

(一) 研究假设

参考迟巍等(2016)提出的中国家庭教育投资模型,本文构建了一个包含家庭是否使用互联网在内的截面模型来检验信息技术普及对家庭教育投资的影响。本研究以家庭教育支出来衡量当期的家庭教育投资。考虑到母亲在家庭教育投资和子女教育发展中的重要作用(Crook, 1995; Donkoh & Amikuzuno, 2011; Reeder & Conger, 1984; 郭秋菊 & 靳小怡, 2016),本文采用母亲是否使用互联网来衡量家庭互联网使用情况。模型如下:

$$\ln \text{Exp}_{ij} = \alpha + \beta \text{Internet}_{ij} + \gamma F_{ij} + \lambda X_{ij} + \delta P_j + \mu_{ij} \quad (1)$$

公式(1)中,因变量 $\ln \text{Exp}_{ij}$ 为第 j 个省份的第 i 个少儿教育总支出的对数形式; Internet_{ij} 是衡量母亲互联网使用情况的虚拟变量; F_{ij} 代表一系列家庭和母亲特征,包括母亲学历、母亲职业、家庭人均年收入、家庭子女个数等; X_{ij} 为一系列子女特征变量,包括子女性别、年龄、民族、教育层级、是否就读重点校、是否就读重点班、是否择校、是否寄宿、是否属于流动儿童、户籍、父母自评语文和数学成绩等; P_j 表示省份虚拟变量,用来控制家庭教育投资环境的差异以及其他省份层面不可观测特征的影响; μ_{ij} 为随机扰动项。

有鉴于知识鸿沟假说提出的信息资本与人力资本转化的关系,本文假设母亲的互联网使用可能增加家庭教育投资。简言之,优势家庭拥有的信息资本可以帮助他们率先投资于人力资本的生产和再生产,提高本人和子女的教育成就,即实现“信息资本—人力资本”的再生产。在具体影响机制方面,周广肃(2016)的研究指出,互联网使用可能通过三个途径来促进家庭风险金融资产投资,即降低交易成本、削弱有限参与机会限制和增强社会互动。若此类机制同样作用于家庭教育投资,可以预期互联网使用会促进投资。本文在第六部分中对部分影响机制进行了检验。

假设 1: 在控制其他因素的条件下,母亲使用互联网能显著增加家庭的教育投资。

我国城乡家庭的互联网使用情况有很大的差异(中国互联网信息中心, 2017),因此互联网对农村和城市家庭教育投资的影响也可能存在差异。根据知识鸿沟假说,基于其社会经济地位和教育优势,城市家庭更容易获取信息资本。城市优势阶层利用其处理新信息的认知优势,可以实现信息资本与人力资本的转化。因此,接入互联网会加大城市家庭的教育投资。然而,互联网未必能增加农村家庭的教育投资。一方面,文献综述表明互联网技术尚未成为农村家庭教育投资的主要影响因素。另一方面,大比例的农村家庭中存在“亲子分离”现象,这也可能影响使用互联网家长对子女的投资。亲子分离会提高留守

儿童的自闭症倾向、降低幸福感和学业成绩(唐有财和符平,2011),也会导致青少年偏差行为的出现(杨建,2014),甚至影响留守或寄宿儿童的人际交往、亲子沟通、父母的学历成就期望等(王静美,2015;金小红和王静美,2015)。即使农村父母接入互联网,长期亲子分离也可能遏制父母的教育投资。

为了分析这种异质性的影响,本文在总样本讨论的基础上进行分样本回归。首先,根据受调查少儿户籍,将样本分为非农村户籍少儿和农村户籍少儿。其次,按照接受调查时少儿所在地,将样本分为非农村户籍少儿、农村户籍流动少儿、农村户籍未迁移少儿三类,分样本讨论母亲互联网使用对家庭教育投资的影响。

假设 2.1: 母亲互联网使用会显著增加非农村户籍少儿获得的家庭教育投资。

假设 2.2: 母亲互联网使用对农村户籍流动少儿和农村户籍未迁移少儿获得的家庭教育投资无显著影响。

不同社会群体在互联网的使用程度和依赖性方面具有显著差异,且二者可能影响个体的劳动力市场表现和家庭投资行为。陈玉宇和吴玉立(2008)的早期研究注意到个人使用计算机工作的时间有很大差异。因此在估计互联网工资溢价时,他们在模型中加入电脑使用时间及其二次项来控制信息技术的使用强度。庄家焯等人(2016)的研究发现男性互联网使用比例高于女性,且在互联网用途方面,男性在“发展型因子”方面高于女性,且发展型因子与互联网工资溢价正相关。有鉴于此,本文认为母亲的互联网使用强度和依赖性同样会影响家庭的教育投资,其使用强度越大、依赖性越高,则家庭教育投资越大。

假设 3: 母亲的互联网使用时长和对互联网作为信息渠道的看法会显著影响家庭教育投资。

(二) 泛精确匹配

公式(1)控制了大量个人和家庭特征以及地区特征,该模型中仍然可能存在互联网使用的自我选择问题及其带来的估计偏差。母亲的互联网使用并非在人群中随机分配,使用与不使用互联网的母亲可能具有研究者未能观测的异质性特征,从而导致控制组和对照组在可观测的控制变量方面也存在显著的不平衡。

为了解决与互联网使用相关的自选择问题,常采用的方式是倾向得分匹配法(Propensity Score Matching Method)。罗森鲍姆(Rosenbaum)和鲁宾(Rubin)(1983)、陈玉宇和吴玉立(2008)和胡安宁(2017)等提出,该方法通过估计可观测变量的倾向得分,基于得分进行全样本内控制组和对照组的分组匹配。因此,因变量在每一组内的差距完全是由于样本的处理效应引起,可以有效地避免样本选择误差。然而,这一类型的方法最多只能达到组间特征变量均值的平衡,无法处理在方差、协方差、高阶交互项等方面存在的数据不平衡性(Iacus,

King, & Porro, 2011, 2012)。

为了避免上述问题,研究者发展出一类新的非参数数据匹配方法,称为“单调不平衡性控制(Monotonic imbalance bounding, MIB)”。MIB方法具有一系列优异的统计特性,其中发展较为成熟的一种方法称为“泛精确匹配(Coarsened Exact Matching, CEM)”(Iacus et al., 2012)^①。泛精确匹配相当于对数据进行预处理,通过数据筛选,减少两个组别间特征变量的非平衡性,使得控制组和对照组保留下来的个体在特征变量上具有相似的分布,从而满足因果推断所要求的两组别分布的重叠性假设。在实际分析中,研究者根据对特征变量的理解,将每个特征变量划分为互不重叠的区间,从而将原先连续或者离散的特征变量转化为分类变量,这个步骤称为数据的粗化处理(data coarsening)。之后,利用这些分类变量对数据进行精确的匹配(exact matching)。在匹配时,研究者仅保留那些同时拥有两个组别个体的区间,通过事先确定的变量区间控制匹配的成功率和匹配后的平衡性(杨振宇 & 张程, 2016)。实际上,泛精确匹配相当于对原始数据加权,从而删除未能匹配的数据单元(Iacus et al., 2012)。

为了提高模型的拟合优度,笔者的目标是最小化多元非平衡性指标(multivariate imbalance measure)。该统计量(L_1)测度两组在全部特征变量上的整体平衡,取值在0和1之间。若控制组和对照组的分布完全不重合,则该指标的取值为1;若完全重合,则取值为零,较小的多元非平衡性指标意味着较好的组间平衡性。CEM泛精确匹配可降低干预效果估计对模型的依赖,部分解决了两组别间个体差异导致的内生性偏差。在取得良好的组间平衡性后,研究者可以借助多元回归分析控制匹配后的个体特征,并以泛精确匹配的权重为加权值来估计干预效果(Iacus et al., 2012)。

在泛精确匹配中,本文选择的变量包括母亲学历、母亲职业、家庭收入、户籍和省份。为了得到有效的匹配数量和改善多元非平衡性指标(Iacus et al., 2012),笔者对母亲职业类别进行了合并,将7个类别合并为5大类别。此外,笔者将家庭收入进行十等分后带入匹配模型,同时将省份划分为东部、中部、东北部和西部。在闫慧和孙立立(2012)对数字鸿沟影响因素的综述中,性别、职业、教育、收入、城乡、地区是最为关键的决定因素,因此本文也选择这些变量作

^① 这种方法成立所需的假设较少(仅需要干预的分配与可观测的自变量无关),而且它保障匹配后的控制组与对照组的组内差距不会大于研究者事先选择的范围。当研究者调整一个变量的平衡性时,其他变量的平衡性不会受到影响。与倾向性匹配得分等方法相比,泛精确匹配能够显著降低数据的不平衡性、模型依赖性、估计误差、数据总体方差和均方差。该方法符合统计学的一致性原则、对测量误差不敏感、能够自动将数据限制在共同区间范围内,并提高计算的效率(Iacus, King, & Porro, 2009; Iacus et al., 2012; King et al., 2011; Stuart, 2010)。

为匹配的基础^①。本文采用了 STATA 中的“CEM”程序进行泛精确匹配^②,然后将 CEM 得到的泛精确匹配权重带入模型(1),采用加权最小二乘法来估计母亲使用互联网对家庭教育投资规模的影响。除了估计总体样本中互联网使用的投资效果,我们也采用该方法估计了互联网使用对非农村户籍城市家庭、农村户籍进城务工家庭和农村户籍未迁移家庭教育投资的拉动作用。

(三) 影响机制分析

影响机制是分析互联网的教育投资效应的重要环节。利用 CFPS2010 年数据,周广肃(2016)分析了互联网使用对家庭金融投资行为的影响。该研究提出互联网使用对家庭投资的影响机制可能三个途径:降低交易成本、消除有限参与机会限制、增强社会互动。本文借鉴了上述思路,尝试分析互联网对家庭教育投资影响的机制。

首先,本文提出母亲使用互联网可能降低教育投资的交易成本,从而鼓励家庭加大投资。交易成本是家庭进行投资时不可忽视的因素。与小额的投资相比,大额投资往往以丰富的信息、广泛的咨询及审慎的权衡为基础,交易成本相对较高。互联网使用可以为家长提供更多的信息及在线咨询的机会,有助于改善教育投资方面的信息不对称问题。因此,互联网使用有可能通过降低交易成本而对大额教育投资产生更大的影响。为了检验这一影响机制,本研究利用分位数回归的方法,探究互联网使用对不同分位数的家庭教育投资的影响(Koenker & Bassett, 1978; 陈强, 2014)。为了估计“q 分位数回归系数”,本文采用下述模型进行估计,即

$$\min_{\beta_q} \sum_{i=1: y_i \geq x_i' \beta_q} q | y_i - x_i' \beta_q | + \sum_{i=1: y_i < x_i' \beta_q} (1-q) | y_i - x_i' \beta_q | \quad (2)$$

其中, x_i 包含母亲使用互联网的虚拟变量、省份虚拟变量、家庭社会经济地位变量和家庭及子女特征变量等。本文估计了母亲使用互联网对家庭教育投资 10 分位数、25 分位数、50 分位数、75 分位数和 90 分位数的影响。

其次,使用互联网可能消除教育投资——尤其是校外教育投资的参与障碍,提升家庭教育投资。互联网在信息传播方面的优势可能会帮助父母获取更多的、有价值的教育投资信息。一方面,丰富的消费者反馈信息可能消除教育服务质量方面的信息不对称的问题。家长投资课外教育补习最大的障碍之一是对教育质量的不了解,在众多服务品类相似的服务提供商中间进行选择,需

^① 本文尝试加入额外的控制变量,并未能使得多元平衡性指标显著变小,而且使得匹配成功的样本数量显著下降。因此,本文仅选择 5 个变量作为参与匹配的特征变量。作者感谢匿名评审人对匹配变量的建议。

^② CEM 的 STATA 应用程序可以从以下网址下载, <https://gking.harvard.edu/cem>。泛精确匹配的具体原理参见 Gary King 与合作者的系列研究论文(Iacus et al., 2009, 2012; King et al., 2011; Stuart, 2010)。

要搜集大量的信息。“家长帮”等在线教育平台的出现为家长提供了有关教育服务质量的反馈信息,父母可以通过互联网使用接触到这些信息,改进投资行为。另一方面,在线课程的兴起有助于克服时间和空间的隔阂,帮助少儿远程参与各类教育活动,相当于消除了参与障碍,提高父母对校外教育投资的参与^①。

本文通过检验互联网使用对课外辅导参与及课外辅导时间的影响来验证这一影响机制。具体分析模型如下:本文采用二元逻辑斯特模型,考察互联网使用对是否参与课外辅导的影响;采用泊松计数模型,考察互联网接入对参与课外辅导种类的影响;采用多元线性回归模型分析互联网使用对参与课外辅导时间的影响;最后,本文采用二元逻辑斯特模型,分别讨论互联网使用对学校课程辅导和才艺培养参与的影响。

四、数据、样本和变量

(一) 数据来源和样本处理

本文数据来自北京大学中国社会科学调查中心开展的“中国家庭追踪调查”(China Family Panel Studies,以下简称CFPS),这是一项全国性、综合性的社会跟踪调查项目,旨在反映中国社会、经济、人口、教育和健康等方面的变迁。该项目于2010年起正式实施基线调查。CFPS调查的目标总体为中国25个省、直辖市、自治区中的家庭户和家庭户中的所有家庭成员。这25个省、直辖市、自治区的人口约占全国总人口(不含港澳台地区)的95%,可被视作具有全国代表性的样本(谢宇,胡婧炜,张春泥,2014)。

本研究使用CFPS2014年的追踪调查数据,调查内容涵盖了社区、家庭和个人三个层次。社区层次调查涵盖政治环境、村/居面貌、基础设施、人口、资源、交通、医疗卫生和财政收支等方面;家庭层次调查包含了家庭的结构与成员关系、生活条件、社会交往、收入支出和资产状况等内容;个人层次调查涵盖了个人的教育、职业、收入、婚姻、心理与生理状况、观念与态度等情况。2014年调查中,对个人互联网使用行为、家庭背景信息及家庭教育投资行为进行了详细调查,为本文提供了良好的数据支持。

CFPS调查中少儿定义是年龄为16岁以下的儿童和青少年。研究者以CFPS的少儿数据库为基础,通过匹配社区库、家庭关系库、家庭库和成人库,获得了每位少儿父母的信息、家庭信息、所在地区信息,建立了少儿与父母匹配的样本库。本文分析的义务教育阶段少儿指16岁以下在小学或者中学就读的儿

^① 搜狐教育:《2016年中国教育行业白皮书》,2017年1月,http://www.sohu.com/a/123945717_484992。

童和青少年,因此样本中删除了未在学校就读的少儿。CFPS2014年的调查数据中,少儿数据库的样本总量为8617人,本文删除非义务教育阶段少儿4272人,余下样本中义务教育阶段的少儿共计4345人。本文删除与父亲或母亲同住8个月以下的1258名少儿后^①,样本数量降低到3087人,占少儿数据库总样本数量的35.8%。经过筛选后样本的特征基本与少儿库总体特征类似。筛选前后母亲特征、家庭收入和子女特征的分布未产生明显的改变,筛选后的样本依然具有代表性。本文分析样本的平均年龄为11岁,最小年龄为5岁,最大为15岁;处于小学阶段的样本占76.4%,初中阶段的样本占23.6%;男生样本占53.1%,女生样本占46.9%,与少儿库总体情况类似。

(二) 变量选择和数据填充

家庭教育投资是本文的主要因变量。在CFPS2014数据库中,本文选取家庭过去12个月投入在每名少儿上的教育支出来衡量家庭教育投资,该投资包括学杂费、交通费、伙食费、课外辅导费、择校费等。

本文的核心自变量是母亲的互联网使用。以往关于互联网使用的研究常选择“是否使用互联网”来衡量互联网的使用情况(李雅楠、谢倩芸,2016;毛宇飞、曾湘泉,2016;周广肃,2016),本文也选取“您/你是否上网?”来衡量母亲的互联网使用。在稳健性检验和影响机制的探讨中,本文还选取母亲业余上网时间以及互联网这一信息渠道的重要性等变量进行稳健性检验和影响机制的分析。

依据以往对家庭教育投资影响因素的研究(迟巍,2013;迟巍 et al., 2011;钱晓烨,迟巍, & 史瑶,2015),本文将控制变量分为三类,分别是家庭和母亲特征(包括人均家庭年收入、母亲最高学历、母亲职业、家庭子女个数)、所在地区(包括所在省份及城乡)以及子女特征(包括子女性别、民族、哪级学校、是否就读重点校、是否就读重点班、是否择校、是否寄宿、父母自评语文成绩、父母自评数学成绩、城乡迁移状况)。变量详细说明和描述统计结果参见表1。

^① 为了保证父母对子女受教育活动的有效参与和影响,本文只保留了每年与父亲或母亲同住8个月以上的少儿。笔者认为不与父母同住少儿获得的教育投资,不会受到母亲互联网使用的影响,因此不应纳入研究样本。为了检验这一数据处理方式的合理性,后文将选择每年与父亲或母亲同住小于8个月的亲子分离样本进行证伪检验。

表 1 变量的描述统计(CFPS2014 全国追踪样本)

变量名称	母亲不使用互联网		母亲使用互联网		T 检验/卡方检验
	均值	标准差	均值	标准差	
家庭教育投资					
年家庭教育总支出(元)	2175.09	3395.23	5044.76	6623.99	-15.29***
无择校费年家庭教育总支出(元)	2004.18	3119.25	4781.32	6927.27	-14.82***
年课外补习支出(元)	381.78	1431.80	2113.04	4351.98	-15.88***
课外补习投资					
学年中参加辅导班比例(%)	11.3%	0.32	35.8%	0.48	236.75***
周课外辅导时间(小时)	1.27	5.46	3.45	7.95	-8.43***
课外辅导种类平均数(种)	0.12	0.35	0.42	0.62	-16.42***
学年中参加课程辅导比例(%)	9.55%	0.29	26.1%	0.44	131.69***
学年中参加才艺培养比例(%)	1.81%	0.133	12.8%	0.33	150.55***
周学校课程辅导时间(小时)	1.15	5.32	2.47	6.80	-5.59***
周才艺培养时间(小时)	0.001	0.067	0.073	0.741	-4.31***
互联网使用					
母亲使用互联网比例(%)			29.1%	0.45	
周业余上网时间(小时)			7.27	9.42	
学习目的的重要性			1.31	0.78	
社交目的的重要性			1.09	0.79	
娱乐目的的重要性			1.06	0.81	
互联网信息渠道的重要性	0.23	0.58	1.32	0.82	-38.79***

(续表)

变量名称	母亲不使用互联网		母亲使用互联网		T 检验/卡方检验
	均值	标准差	均值	标准差	
家庭和母亲特征					
母亲学历层次	2.08	0.98	3.47	1.28	726.39***
母亲职业类型	4.29	1.85	3.32	1.91	705.77***
年人均家庭纯收入(元)	9500.21	11351.5	18211.0	23899.1	-13.22***
居住城市比例(%)	34.8%	0.476	69.6%	0.460	288.53***
子女数(人)	2.15	0.91	1.53	0.65	378.78***
家庭规模(人)	5.05	1.72	4.55	1.71	7.13***
子女特征					
男性子女(%)	53.1%	0.49	53.5%	0.49	0.039
少数民族(%)	13.6%	0.34	4.50%	0.21	50.75***
初中学生(%)	24.2%	0.43	23.7%	0.43	0.072
择校学生(%)	14.6%	0.35	16.2%	0.37	1.16
重点学校(%)	19.5%	0.39	28.9%	0.45	30.95***
重点班级(%)	9.16%	0.29	7.54%	0.26	1.98
少儿城乡迁移状况	2.49	0.72	1.75	0.84	492.90***
自评语文成绩等级	2.20	0.95	1.94	0.89	46.91***
自评数学成绩等级	2.24	1.02	1.87	0.92	85.16***
寄宿少儿(%)	13.4%	0.34	6.42%	0.25	28.43***
样本数	836		2041		

注:1.最后一列为连续变量的 T 检验结果或离散变量的卡方检验结果。

2.显著性水平: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ 。

研究中使用的部分变量存在数据缺失问题。关键控制变量中,母亲职业、母亲最高学历、母亲是否使用互联网、家庭人均收入四个变量的缺失比例较高,缺失率分别为10.1%、12.3%、14.1%、7.4%。有鉴于此,本文采用了多种常用的方法进行数据填充(Ette, Chu, & Alaa, 2006)。首先,本文采用父亲的相关信息来填补母亲的职业、学历和互联网使用变量的缺失。这种做法出于两方面的考虑。一方面,人口和婚姻研究领域的诸多研究表明,我国夫妻双方的职业和学历存在高度的正相关关系(陆益龙, 2009; 张翼, 2003),母亲的社会经济背景与父亲应高度相关;另一方面,采用配偶情况进行数据填充是家庭调查数据分析的常用做法。其次,本文使用样本中人均家庭收入的中位数来替代家庭人均收入变量的缺失值。数据填充完成后,母亲职业、母亲最高学历和母亲是否使用互联网等三个变量的缺失比例分别下降为3%、5%和6.8%,大幅度降低了由于变量缺失带来的样本缺失情况。家庭人均收入变量不再存在缺失值。独立样本T检验表明,填充后样本均值与填充前无显著差异,说明填充效果较好。

五、家庭教育投资差距的实证分析

(一) 母亲互联网使用的教育投资效果

近年来,学者开始采用泛精确匹配方法取代倾向得分匹配,进行干预效果的估计。例如,有研究者(Schurer, Alspach, Macrae, & Martin, 2015)利用CEM方法和新西兰健康管理机构的数据,估计了情绪失控对新西兰医疗保健成本的影响。另一项医学研究利用泛精确匹配方法来识别心脏病死亡患者的确切死因(Stevens, King, & Shibuya, 2010)。在教育领域,也有研究者采用泛精确匹配的方法,利用美国西部一所大学的数据分析了参与由学生同伴引导的“补充教学”活动对大学生课程成绩的影响(Guarcello et al., 2017)。杨振宇和张程(2016)在分析我国城乡迁移对农村籍父母教育观念和教育行为的影响时,利用CEM方法进行数据预处理,解决了自我选择带来的组别间数据不平衡问题。

本文尝试采用CEM方法解决使用和未使用互联网家庭间特征变量不平衡问题,即首先通过泛精确匹配提高组别间的平衡性,然后只保留配对成功、位于共同支持区间的个体,再利用加权回归的方法来识别互联网使用的处置效应。如前所述,本文首先采用母亲和家庭特征等变量进行CEM匹配,附录表1和表2汇报了全样本、非农村户籍少儿样本、农村户籍流动少儿样本和农村户籍未迁移少儿样本的匹配结果和平衡性检验结果。泛精确匹配较为成功地保留了足够数量的配对样本。在全体样本中,匹配成功的样本为1284人,未成功样本为

1593人,成功匹配率为45%^①。该表同时展示了CEM匹配的效果。全体样本中匹配前后的多元非平衡性指标从0.745下降为0.264,这表明在匹配后控制组和对照组的分布重合率达到了73.6%。同理,城镇少儿的 L_1 从0.719下降到了0.252;城乡迁移少儿的 L_1 从0.736下降到0.174,降幅最为明显。此外,CEM的平衡性检验表明,匹配后主要控制变量的均值差异不再显著,而且在四个分位数点上的组别间差异也不再显著,组间平衡性大幅度提升(平衡性检验结果参见附录表2)。

本文的核心任务是考察母亲互联网使用对少儿获得的家庭教育投资的影响。表2报告了模型(1)的最小二乘法估计结果,以及基于CEM匹配样本的、采用CEM权重加权的最小二乘法估计结果。

表2 母亲互联网使用对家庭教育总投资的影响

	最小二乘法模型	泛精确加权回归模型
母亲使用互联网	0.20*** (0.06)	0.20** (0.08)
母亲学历		
小学	0.05 (0.07)	0.04 (0.19)
初中	0.12 (0.07)	0.17 (0.20)
高中职高技校	0.13 (0.10)	0.33 (0.22)
大学专科	0.40*** (0.14)	0.03 (0.32)
大学本科	0.28 (0.19)	
硕士	1.74*** (0.20)	
博士	1.88*** (0.21)	
家庭人均年收入	0.07*** (0.02)	0.12** (0.05)
母亲职业		

^① 在CFPS2014数据中的836位母亲使用互联网的少儿中,CEM为54%的少儿匹配成功;在母亲未使用互联网的2041位少儿中,41%的少儿成功匹配成功。在使用互联网家庭中,农村未迁移少儿的匹配成功率最高,达到了70%;而在未使用互联网家庭中,非农村户籍少儿匹配成功率最高(47%)。

	(续表)	
	最小二乘法模型	泛精确加权回归模型
国家机关、党群组织、企业事业单位负责人	0.34** (0.14)	0.13 (0.23)
专业技术人员	-0.01 (0.13)	-0.24 (0.26)
办事人员和有关人员	0.51*** (0.14)	0.74** (0.32)
商业、服务人员	0.08 (0.09)	-0.04 (0.14)
农林牧渔水利业人员	-0.17** (0.09)	-0.32** (0.16)
生产运输操作及有关人员	-0.07 (0.09)	-0.10 (0.14)
军人及其他不便分类职业	-0.13 (0.19)	-0.42* (0.25)
子女为男性	0.00 (0.05)	-0.10 (0.09)
子女年龄	0.01 (0.01)	0.02 (0.02)
子女为少数民族	0.10 (0.09)	0.05 (0.21)
子女为初中生	0.28*** (0.08)	0.12 (0.14)
子女就读重点学校	0.32*** (0.06)	0.33*** (0.09)
子女就读重点班	0.23*** (0.08)	0.41*** (0.16)
子女参与择校	0.36*** (0.06)	0.33*** (0.12)
农村户籍—城乡迁移少儿	-0.19*** (0.07)	-0.03 (0.11)
农村户籍—未迁移少儿	-0.40*** (0.07)	-0.28** (0.13)
子女属于寄宿生	1.01*** (0.07)	0.95*** (0.12)

	(续表)	
	最小二乘法模型	泛精确加权回归模型
父母自评子女语文成绩	0.04 (0.03)	-0.06 (0.05)
父母自评子女数学成绩	-0.00 (0.03)	0.06 (0.05)
家庭子女个数	-0.12*** (0.03)	-0.25*** (0.06)
省份虚拟变量 观测值数	是 2,584	是 1,174
调整后 R 平方	0.397	0.410

注:1. 母亲学历的对照组为文盲;母亲职业的对照组为无工作;子女户籍对照组为非农户籍。模型控制了省份虚拟变量,对照组是北京。

2. 括号内是稳健标准误。

3. 显著性水平:* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ 。

模型(1)的因变量为家庭教育总投资的对数,泛精确加权回归结果显示使用互联网母亲的子女获得的家庭教育投资较其他人高 20%。结果验证了假设 1,即互联网使用具有显著的教育投资拉动效果。这表明家庭信息资本已经成功转化为人力资本投资。该效果不仅仅是家庭收入提高带来的“收入效应”:即便在控制家庭收入后,互联网使用的投资促进效果仍然显著。在以下的影响机制分析中,本文将具体讨论互联网使用如何影响家庭教育投资。

除互联网使用之外,其他因素对家庭教育投资的影响也值得关注。首先在家庭社会经济地位方面,母亲职业、家庭人均收入均对家庭教育投资有显著的影响。母亲担任办事人员和有关人员、来自高收入家庭的少儿,其获得的家庭教育投资总额显著高于其他少儿,这与以往发现一致(丁小浩 & 薛海平, 2005; 曲颖 & 薛海平, 2015)。在匹配成功的学历样本中,母亲学历对家庭教育投资没有显著影响。其次,在家庭及子女特征方面,就读重点班级和重点学校、参与择校的少儿获得的家庭教育投资分别显著的高于其他人。农村户籍学生获得的家庭教育投资显著低于非农户籍的学生。农村户籍流动少儿获得的家庭教育投资比非农户籍少儿显著低 19%,农村户籍且未迁移到城镇的学生获得的家庭教育投资最少,比非农户籍的学生低 40.2%。该结果表明,家庭教育投资的城乡差异持续存在。

(二) 效果异质性和稳健性分析

为了验证假设 2,本文进行了分样本回归,探讨互联网使用对家庭教育投资的影响在居住地、户籍方面的异质性。表 3 汇报了最小二乘法 and CEM 泛精确加权回归模型的结果。

表 3 母亲互联网使用对家庭教育总投资影响的城乡差异

	最小二乘法模型			泛精确加权回归模型		
	非农村 户籍	农村 户籍— 流动	农村 户籍— 未迁移	非农村 户籍	农村 户籍— 流动	农村 户籍— 未迁移
母亲使用互联网	0.35*** (0.10)	0.01 (0.12)	0.08 (0.10)	0.36** (0.16)	0.11 (0.16)	0.14 (0.14)
其他控制变量						
子女性别、年龄、民族、学段	是	是	是	是	是	是
子女就读重点校、重点班、 择校、寄宿	是	是	是	是	是	是
父母自评子女语文成绩、数 学成绩	是	是	是	是	是	是
母亲学历	是	是	是	是	是	是
母亲职业	是	是	是	是	是	是
家庭收入和子女个数	是	是	是	是	是	是
省份虚拟变量	是	是	是	是	是	是
观测值数	635	598	1,351	260	228	540
调整后 R 平方	0.321	0.383	0.317	0.163	0.554	0.373

注:1. 母亲学历的对照组为文盲;母亲职业的对照组为无工作;子女户籍对照组为非农村户籍。模型控制了省份虚拟变量,对照组是北京。

2. 括号内是稳健标准误。

3. 显著性水平:* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ 。

泛精确加权回归模型的结果表明,母亲互联网使用仅对非农户籍少儿获得的家庭的教育投资具有显著的正向影响(回归系数为 0.36,在 5%的水平上显著);对农村户籍流动少儿、留在农村的农村户籍少儿没有显著作用。农村户籍家庭的子女即使流动到城镇,也不会因母亲的互联网使用而获得更多的教育投资。最小二乘法估计结果与此完全一致。上述分析结果验证了假设 2,互联网使用可以显著提升家庭教育投资规模。遗憾的是,这一积极效应目前仅存在于城镇户籍家庭,并未显著影响进城务工家庭和农村未迁移家庭。换言之,互联网红利仅仅延伸到已经在教育投资中占据优势的城镇家庭;即便农村家庭或者进城务工人口家庭中的母亲开始利用互联网,这一行动本身并未改变家庭的教育投资行为。考虑到家庭教育投资与子女教育成就之间的显著相关关系(胡咏梅,范文凤, & 丁维莉, 2015; 李佳丽 & 胡咏梅, 2017),互联网的普及可能会恶化中国城乡的教育成就鸿沟,而不是弥合已有的城乡差距。

为了进一步考察亲子分离对互联网使用与家庭教育投资关系的影响,本文考察了亲子分离样本(即将每年与父母共同居住时间少于 8 个月的少儿作为样本)中母亲互联网使用的投资效应。表 4 呈现了亲子分离样本的分析结果。

表4 亲子分离样本检验结果

	最小二乘法模型	泛精确加权回归模型
母亲使用互联网	-0.16 (0.11)	0.22 (0.18)
其他控制变量		
省份虚拟变量	是	是
家庭和母亲特征	是	是
子女特征	是	是
观测值数	616	145
调整后R平方	0.428	0.603

注:1. 母亲学历的对照组为文盲;母亲职业的对照组为无工作;子女户籍对照组为非农户籍。模型控制了省份虚拟变量,对照组是北京。

2. 括号内是稳健标准误。

3. 显著性水平:* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ 。

表4中泛精确加权回归结果显示,在亲子分离样本中,母亲互联网使用对子女获得的家庭教育投资不存在显著的影响。该结果表明本文对样本的处理方式较为合理,即回归分析中应删除此类样本。这个发现从另一个维度解释了本研究表3发现的互联网效应异质性。在进城务工人口和农村户籍未迁移人口家庭中,亲子长期分离的现象较为普遍。CFPS2014年数据中,92.1%的亲子分离样本属于农村少儿。在亲子分离家庭中,长期分居导致父母无法有效参与子女的教育活动,从而会影响父母对子女的教育期望与投资^①。显然,即使亲子分离家庭的父母跨越了数字鸿沟、实现了互联网的接入和使用,由于长期的分居,这些农村迁移或未迁移家庭的父母也无法将其信息优势转化为子代的人力资本积累优势。

以往文献表明(闫慧 & 孙立立, 2012; 周广肃, 2016),单纯使用“是否使用互联网”这一变量难以全面刻画个体的互联网使用行为,它只对应于第一道数字鸿沟,即互联网的接入,它无法描述第二道和第三道数字鸿沟,即互联网的技能和应用(赵联飞, 2015)。业余上网时间可以衡量个体对互联网的依赖程度;对互联网这一信息渠道的态度可以衡量互联网在个体信息获取方面的重要性。因此,本文采用母亲业余上网时间、母亲对互联网这一信息渠道的态度分别代替“母亲是否使用互联网”这一核心自变量进行稳健性检验。回归结果参见表5,控制变量与表2相似。

采用最小二乘法模型,表5表明母亲互联网使用时长对家庭教育投资具有显著的正向影响,随着使用时间的增加,家庭教育投资显著增长。此外,与持中

^① CFPS2014年数据中,亲子分离家庭中父母对子女的教育期望低于亲子同居家庭(29.5%亲子分离的父母对子女的教育期望低于大学本科,24.1%的亲子同居家庭具有相似教育期望);亲子分离家庭中父母为子女教育储蓄的比例(19.3%)也低于其他家庭(22.6%)。

立态度的母亲相比,认为互联网这一信息渠道不重要的家庭其教育投资水平显著的低 19%;持重视态度的母亲所在家庭的教育投资水平并未显著提高。这些发现验证了本文的假设 3,即母亲的互联网使用时长和对互联网作为信息渠道的积极看法会显著影响家庭教育投资。综上所述,当采用“母亲互联网使用”的其他代理变量时,本文的基本结论仍然成立,即家庭之间的数字鸿沟正在转化为教育投资方面的鸿沟,互联网扩大而非弥合了以往文献呈现出的中国家庭之间的教育投资差异(钱晓烨 et al., 2015)。

表 5 母亲互联网使用强度和依赖性的教育投资效应

	模型 1	模型 2
母亲互联网使用时长	0.01** (0.00)	
互联网信息渠道的重要性:不重要		-0.19** (0.08)
互联网信息渠道的重要性:重要		0.05 (0.09)
其他控制变量		
省份虚拟变量	是	是
家庭和母亲特征	是	是
子女特征	是	是
观测值数	2470	2442
调整后 R 平方	0.395	0.400

注:1. 母亲学历的对照组为文盲;母亲职业的对照组为无工作;子女户籍对照组为非农户籍。互联网信息渠道的重要性的对照组为中立。模型控制了省份虚拟变量,对照组是北京。

2. 模型 1 的解释变量为母亲互联网使用时长;模型 2 的解释变量为互联网信息渠道的重要性(虚拟变量),对照组是中立态度。

3. 括号内是稳健标准误。

4. 显著性水平:* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ 。

六、影响机制讨论

本文借鉴周广肃(2016)对互联网使用对家庭金融投资行为影响的分析思路,尝试从降低交易成本、消除有限参与机会限制这两个维度,分析互联网使用对家庭教育投资影响的机制。

(一) 降低交易成本

交易成本是家庭进行教育投资时不可忽视的因素。已有文献深入探讨了个人的风险偏好和交易成本对大学选择和专业选择的影响(Chen et al., 2001;

Diaz-Serrano, Hartog, Nilsson, Van Ophem, & Yang, 2016; Hartog, Ding, & Liao, 2014; Hartog, Sun, & Ding, 2010), 尚未有研究分析降低交易成本是否会促进家庭的教育投资。本研究认为互联网使用有助于降低交易成本, 促进大额教育投资, 但是对小额教育投资的影响不显著。为了检验这一影响机制, 本研究利用分位数回归的方法, 探究互联网使用对不同教育投资水平家庭的影响。表6呈现了使用互联网使用对处于教育投资分布的第10分位数、第25分位数、第50分位数、第75分位数和第90分位数家庭投资规模的影响。

表6 降低交易成本机制检验:分位数回归结果

	家庭教育投资不同分位数点估计				
	10分位数	25分位数	50分位数	75分位数	90分位数
母亲使用互联网	0.11 (0.07)	0.20*** (0.06)	0.20*** (0.06)	0.25*** (0.05)	0.23*** (0.06)
其他控制变量					
省份虚拟变量	是	是	是	是	是
家庭和母亲	是	是	是	是	是
子女特征	是	是	是	是	是
观测值数	2,584	2,584	2,584	2,584	2,584
伪R平方	0.226	0.259	0.271	0.249	0.239

注:1. 母亲学历的对照组为文盲;母亲职业的对照组为无工作;子女户籍对照组为非农户籍。模型控制了省份虚拟变量,对照组是北京。

2. 括号内是稳健标准误。

3. 显著性水平:* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ 。

分位数回归的结果表明,互联网使用对小额教育投资没有显著影响,对于中高投资有显著影响,且在高额投资中的影响更大。在第10分位数点上,互联网使用产生了不显著的正向影响。随着家庭教育投资额度的增加,母亲使用互联网对教育投资的拉动效果逐步提高,且变得显著。在第25分位数点和第50分位数点上,互联网使用的教育投资效果为20%,在75分位数点上的效果为25%,在90分位数点上的作用为23%。由此可见,互联网使用对大额家庭教育投资产生了更大的影响。这可能是由于通过互联网,父母得以获得与教育投资相关的、更多的信息,或者因为选择在线教育服务而降低了服务费用,从而倾向于购买更多和更高价的教育服务。

(二) 消除有限参与障碍

互联网在信息传播方面的优势可能会帮助父母获取更多的、有价值的教育投资信息,从而消除了教育投资中的有限参与障碍。本文提出,通过使用互联网,家庭更有可能参与线上或者线下的课外辅导活动。此外,家庭有可能投资更多类型的课外辅导活动,包括学校课程补习和才艺培养课程。伴随着互联网

提供的更多的课外辅导机会,少儿参与补习或者才艺发展的时间也会逐步增加。本文通过检验互联网使用对课外辅导参与及课外辅导时间的影响来验证这一影响机制,结果呈现在表7中。

回归分析验证了上述影响机制假设。首先,在控制其他变量的条件下,母亲的互联网使用有效的提高了子女参与课外辅导的可能性。与未使用互联网的家庭相比,使用互联网家庭参与课外辅导、参与学校课程辅导、参与才艺培训的可能性均显著的高于其他家庭(比值比分别为1.64、1.55和1.96)。特别值得注意的是,新信息技术的使用提高了家庭的投资强度。泊松计数模型的回归结果显示,跨越了数字鸿沟的家庭倾向于投资更多种类的课外辅导。同时,除了常规的学校课程补习,这些家庭还愿意投资于高成本的才艺培养。

表7 互联网使用对课外辅导参与的影响

	二元逻辑 斯特模型	泊松计 数模型	二元逻辑 斯特模型		最小二 乘法模型	
	参与课 外辅导	课外辅 导种类数	参与学校 课程辅导	参与 才艺培养	学校课程 辅导时间	才艺培 养时间
母亲使用互联网	1.64*** (0.23)	1.44*** (0.14)	1.55*** (0.23)	1.96** (0.51)	0.75** (0.34)	0.50*** (0.03)
其他控制变量						
省份虚拟变量	是	是	是	是	是	是
家庭和母亲特征	是	是	是	是	是	是
子女特征	是	是	是	是	是	是
观测值数	2,647	2,650	2,648	2,303	2,643	2,524
预测准确百分比	83.94%		86.74%		94.40%	
调整后 R 方					0.059	0.012
伪 R 平方	0.233	0.195	0.204	0.272		

注:1. 母亲学历的对照组为文盲;母亲职业的对照组为无工作;子女户籍对照组为非农户籍。模型控制了省份虚拟变量,对照组是北京。

2. 括号内是稳健标准误。

3. 显著性水平:* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ 。

在参与校外教育投资的家庭中,使用互联网家庭投资于学校课程补习和才艺培养的时间也显著更多。这些发现表明拥有信息资本优势的家庭不仅更倾向于参与校外教育投资,而且愿意购买更多和更高价的补充性教育服务。表7还显示相比于对学校课程的投资,母亲使用互联网对才艺培养投资的促进作用更大。这一发现有力地支持了互联网使用消除有限参与障碍的假设。与学校课程辅导相比,家长在才艺培养投资方面受制于更严重的信息不对称问题,因为才艺培养这类教育服务的质量更难以直接观察。互联网带来的广泛、充分、即时的信息有助于家长克服信息的制约,促进了家庭在艺术培养方面的投资。

由此可见,新信息技术的使用不仅提高了家庭教育投资的可能性,而且提高了投资的规模和力度,甚至加大了投资的范围。在控制家庭收入等社会经济背景条件后,互联网仍然能够有效的提高家庭教育投资,这表明它不仅仅是家庭“收入效应”的副产品,而是真正改变了家庭教育投资模式。简言之,从需求方面来看,互联网使用解决了信息不对称的问题,为消除有限参与提供了机会;从供给方面来看,“互联网+教育”的兴起提供了更多的家庭教育投资机会,为消除有限参与障碍提供了可能性。

七、结论与讨论

本文着重探讨了社会优势阶层如何凭借其信息优势、通过多种渠道强化人力资本投资优势。本研究得出两个基本结论。

第一,数字鸿沟有可能转变为教育鸿沟。本文发现除了家庭社会经济背景,互联网使用成为扩大家庭间人力资本投资差距的重要推动力之一。考虑到我国传统的弱势群体—低收入家庭、农村户籍的进城务工人员、农村居民的互联网使用比例和家庭教育投资规模本来就低于城市居民,这种高低收入之间、城乡之间的数字鸿沟有可能进一步转化为家庭教育投资的鸿沟,强化优势阶层在人力资本投资方面的优势,扩大不同阶层之间教育机会的差距。通过在人力资本投资方面的先发优势,信息优势阶层还有可能将这种优势转化为更高的教育成就,从而成功的将“数字红利”转化为“教育红利”(Bulman & Fairlie, 2016)。

第二,缩小数字鸿沟未必能弥合教育鸿沟。知识鸿沟假说认为社会经济背景决定了个体接触和使用信息技术的可能性,决定了信息资本积累方面的差距(Dimaggio et al., 2001)。部分社会弱势群体有可能在互联网普及的过程中跨越数字鸿沟,实现互联网的接入和使用。遗憾的是,我们发现这些跨越了数字鸿沟的弱势群体—接近三分之一的进城务工人员家庭和接近七分之一的农村未迁移人口家庭,并未能成功的将家庭信息优势转化为人力资本投资方面的优势。相反,多数获取了信息优势的城市家庭能够成功的将信息资本转化为人力资本投资。

由此可见,信息优势未必能直接转化为人力资本,也未必能促进社会流动,成功的转化与家庭原有的社会经济背景有很大关系。在中国,户籍制度是社会分层的基础,也是制约社会流动的核心要素(李强, 2011)。互联网的出现固然为社会流动提供了新的机遇,但是它很难打破户籍制度下城乡的教育差距,尤其是城市流动人口面临有限的教育投资机会。因此,缩小数字鸿沟、帮助弱势阶层接入和使用互联网仍不足以弥合不同阶层间的教育鸿沟。

本文的重点不是讨论少儿在校或者在家互联网使用对教育成就的影响,而是关注作为教育投资主体的父母的互联网接入与使用对家庭代际人力资本投

资的影响。本文拓展了以往教育经济学的研究视野,将分析重心从“信息技术对个体教育成就的因果性影响”,拓展为“信息技术如何放大了家庭教育投资的不平等分布”。前者仅仅关注数字鸿沟对个体人力资本积累的作用,而后者着眼于社会分层的过程,考察从社会阶层到信息技术资本积累、直至信息技术资本转化为教育投资和人力资本积累、并最终巩固社会分层的动态过程。

本文仍存在一些局限。首先,虽然研究者采用泛精确匹配方法在一定程度上控制了家庭互联网使用的内生性问题,但是受到此方法本身的局限,本文未能识别互联网使用对家庭教育投资的因果性影响。未来研究可以考虑采用外生的政策变化所导致的互联网使用差异提高来识别其因果效应。其次,本文主要分析了当期家庭和个人特征等因素对教育投资的影响,没有考虑前期教育投入和少儿教育经历等因素的作用。家庭教育投资是一个积累的过程,后期投入与前期投入相关。未来的研究可能考虑利用 CFPS 等追踪调查提供的面板数据,分析前期家庭特征与教育投入等因素当期教育投资的影响。

参考文献

- [1] 陈纯槿,顾小清.互联网是否扩大了教育结果不平等——基于 PISA 上海数据的实证研究[J].北京大学教育评论,2017(1):140—53.
- [2] 陈纯槿,郅庭瑾.信息技术应用对数字化阅读成绩的影响——基于国际学生评估项目的实证研究[J].开放教育研究,2016(4):57—70.
- [3] 陈强.高级计量经济学及 Stata 应用[M].北京:高等教育出版社,2014.
- [4] 陈玉宇,吴玉立.信息化对劳动力市场的影响:个人电脑使用回报率的估计[J].经济学:季刊,2008(4):1149—1166.
- [5] 程千里,张炜华.数字鸿沟与学生学业水平——以 PISA 2012 上海地区的调查数据为例[J].青年学报,2016(3):49—56.
- [6] 迟巍.我国城镇家庭教育支出研究[M].北京:清华大学出版社,2013.
- [7] 迟巍,钱晓焯,吴斌珍.家庭教育支出平等性的实证研究[J].教育与经济,2011(4):34—37.
- [8] 丁小浩,翁秋怡.职业权力与家庭教育支出——基于政治经济学视角的实证分析[J].教育研究,2015(8):33—41.
- [9] 丁小浩,薛海平.我国城镇居民家庭义务教育支出差异性研究[J].教育与经济,2005(4):39—44.
- [10] 龚继红,钟涨宝.流动儿童与城市本地儿童的家庭教育差异及影响因素——基于武汉市 853 份样本的分析[J].学习与实践,2015(3):106—14.
- [11] 郭秋菊,靳小怡.农民工的母亲对子女重大决策的影响研究[J].社会,2016(5):222—41.
- [12] 胡安宁.统计模型的“不确定性”问题与倾向值方法[J].社会,2017(1):186—210.
- [13] 胡鞍钢,王蔚,周绍杰,鲁钰锋.中国开创“新经济”——从缩小“数字鸿沟”到收获“数字红利”[J].国家行政学院学报,2016(3):3—13.

- [14] 胡咏梅,范文凤,丁维莉. 影子教育是否扩大教育结果的不均等——基于 PISA 2012 上海数据的经验研究[J]. 北京大学教育评论, 2015(3): 29—46.
- [15] 金小红,王静美. 城乡结合部留守儿童的行为问题:亲子分离或家庭教育的影响——基于定量分析的研究[J]. 教育研究与实验, 2015(3):43—48,57.
- [16] 卡斯特,曼纽尔. 网络社会的崛起[M]. 夏铸九,王志弘等译. 北京:社会科学文献出版社,2006.
- [17] 李佳丽,胡咏梅. 谁从影子教育中获益? ——兼论影子教育对教育结果均等化的影响[J]. 教育与经济, 2017(2): 51—61.
- [18] 李强. 社会分层十讲 [M]. 北京:社会科学文献出版社,2011.
- [19] 李升. “数字鸿沟”:当代社会阶层分析的新视角[J]. 社会, 2006(6): 81—94.
- [20] 李湘萍. 义务教育阶段择校行为与教育机会分布公平性研究——基于中国 18 个城市居民家庭教育选择支出的实证分析[J]. 教育研究, 2008(3): 67—72.
- [21] 李雅楠,谢倩芸. 互联网使用对工资及工资不平等的影响[A]. 中国劳动经济学年会会议论文集[M]. 北京,2016.
- [22] 梁琦. 义务教育阶段家庭教育投资对子女成绩影响的实证研究[D]. 清华大学, 2012.
- [23] 陆益龙. “门当户对”的婚姻会更稳吗? ——匹配结构与离婚风险的实证分析[J]. 人口研究, 2009(2): 81—91.
- [24] 毛宇飞,曾湘泉. 互联网使用是否促进了女性就业? ——基于 CGSS 数据的经验分析 [A]. 中国劳动经济学年会会议论文集[M]. 北京,2016.
- [25] 潘光辉. “撤点并校”、家庭背景与入学机会[J]. 社会, 2017(3):131—162.
- [26] 钱晓烨,迟巍,史瑶. 义务教育阶段城镇家庭教育支出的构成及不平等:来自 2007 和 2011 的实证证据[J]. 教育与经济, 2015(6): 25—33.
- [27] 曲颖,薛海平. 我国义务教育阶段课外补习私人成本研究[J]. 上海教育科研, 2015 (4): 21—26.
- [28] 世界银行. 2016 年世界发展报告:数字红利 [M]. 北京:清华大学出版社,2017.
- [29] 搜狐教育. 2016 年中国教育行业白皮书 [EB/OL]. http://www.sohu.com/a/123945717_484992.
- [30] 唐有财,符平. 亲子分离对留守儿童的影响——基于亲子分离具体化的实证研究 [J]. 人口学刊, 2011(5):41—49.
- [31] 王静美. 城乡结合部留守儿童行为影响因素实证研究:亲子分离与家庭教育[D]. 华中师范大学, 2015.
- [32] 谢宇,胡婧炜,张春泥. 中国家庭追踪调查:理念与实践[J]. 社会, 2014(2): 1—32.
- [33] 闫慧,孙立立. 1989 年以来国内外数字鸿沟研究回顾:内涵、表现维度及影响因素综述[J]. 中国图书馆学报, 2012(5): 82—94.
- [34] 杨建. 留守儿童社会化问题的家庭归因研究——基于武汉市城乡结合部留守儿童的实证分析[D]. 华中师范大学, 2014.
- [35] 杨振宇,张程. 城乡迁移对农村籍父母教育观念与行为的影响[J]. 清华大学教育研究, 2016(4): 71—82.
- [36] 殷红霞. 我国农村家庭教育投资行为研究[M]. 北京:中国社会科学出版社,2010.
- [37] 张锦华,杨晖,沈亚芳等. 不确定性对城乡家庭教育支出倾向的影响研究[J]. 复旦

- 教育论坛, 2014(6): 83—88.
- [38] 张翼. 中国阶层内婚制的延续[J]. 中国人口科学, 2003(4): 39—47.
- [39] 赵联飞. 中国大学生中的三道互联网鸿沟——基于全国12所高校调查数据的分析[J]. 社会学研究, 2015(6): 145—68.
- [40] 中国互联网信息中心. 第39次中国互联网络发展状况统计报告[EB/OL]. http://www.cac.gov.cn/2017-01/22/c_1120352022.htm.
- [41] 周广肃. 互联网使用、市场摩擦与家庭风险金融资产投资[A]. 第四届中国劳动经济学者论坛会议论文集[M]. 北京, 2016.
- [42] 庄家焯, 刘爱玉, 孙超. 网络空间性别不平等的再生产: 互联网工资溢价效应的性别差异以第三期妇女地位调查为例[J]. 社会, 2016(5): 88—106.
- [43] Blau, P. (1977). *Inequality and heterogeneity*. New York: Free Press.
- [44] Bray, M. (2014). The impact of shadow education on student academic achievement: Why the research is inconclusive and what can be done about it. *Asia Pacific Education Review*, 15(3), 381—389.
- [45] Bray, M., Ding, X., & Huang, P. (2004). Reducing the burden on the poor: Household costs of basic education in Gansu, China. *International Studies in Educational Administration*, 26(4), 445—447.
- [46] Bray, M., & Kobakhidze, M. N. (2014). Measurement issues in research on shadow education: Challenges and pitfalls encountered in TIMSS and PISA. *Comparative Education Review*, 58(4), 590—620.
- [47] Bray, M., & Kwo, O. (2013). Behind the façade of fee-free education: Shadow education and its implications for social justice. *Oxford Review of Education*, 39(4), 480—497.
- [48] Bulman, G., & Fairlie, R. W. (2016). Technology and education: Computers, software, and the internet. In E. Hanushek, & L. Woessmann (Eds.), *Handbook of economics of education* (Vol. 5) (pp. 230—280). Amsterdam, The Netherlands: Elsevier B. V.
- [49] Castells, M. (1996). *Rise of the network society: The information age: Economy, society and culture*. Blackwell Publishers, Inc.
- [50] Chen, S. H., Bils, M., Hanushek, E., Fox, J., Kolay, S., & Lee, Y., et al. (2001). *Is investing in college education risky?* University at Albany, SUNY, Department of Economics.
- [51] Chi, W., & Qian, X. (2016). Human capital investment in children: An empirical study of household child education expenditure in China, 2007 and 2011. *China Economic Review*, 37, 52—65.
- [52] Crook, C. J. (1995). The role of mothers in the educational and status attainment of Australian men and women. *Journal of Sociology*, 31(2), 45—73.
- [53] Diaz-Serrano, L., Hartog, J., Nilsson, W., Van Ophem, H., & Yang, P. (2016). *Student earnings expectations: Heterogeneity or noise?* Social Science Electronic Publishing.

- [54] Dimaggio, P., & Hargittai, E. (2001). *From the 'digital divide' to 'digital inequality'*. Working paper series NO. 15. Princeton University.
- [55] DiMaggio, P., Hargittai, E., Celeste, C., & Shafer, S. (2004). From unequal access to differentiated use: A literature review and agenda for research on digital inequality. In Neckerman, K. (Ed.), *Social inequality* (pp. 355—400). New York: Russell Sage Foundation.
- [56] Donkoh, S. A., & Amikuzuno, J. A. (2011). The determinants of household education expenditure in Ghana. *Educational Research & Reviews*, 6(8), 570—579.
- [57] Donohue, G. A., Tichenor, P. J., & Olien, C. N. (1975). Mass media and the knowledge gap a hypothesis reconsidered. *Communication Research*, 2(1), 3—23.
- [58] Ette, E. I., Chu, H., & Alaa, A. (2006). *Data imputation. Pharmacometrics: The science of quantitative pharmacology*. John Wiley & Sons, Inc.
- [59] Golder, S. A., & Macy, M. W. (2014). Digital footprints: Opportunities and challenges for online social research. *Annual Review of Sociology*, 40(1), 129—152.
- [60] Guarcello, M. A., Levine, R. A., Beemer, J., Frazee, J. P., Laumakis, M. A., & Schellenberg, S. A. (2017). Balancing student success: Assessing supplemental instruction through coarsened exact matching. *Technology Knowledge & Learning*, 4, 1—18.
- [61] Hartog, J., Ding, X., & Liao, J. (2014). Is earnings uncertainty relevant for educational choice? An empirical analysis for China. *Education Economics*, 22(5), 471—483.
- [62] Hartog, J., Sun, Y., & Ding, X. (2010). University rank and bachelor's labour market positions in China. *Economics of Education Review*, 29(6), 971—979.
- [63] Haywood, T. (1995). *Info-Rich—Info-Poor: Access and exchange in the global information society*. London: Bowker-Saur.
- [64] Iacus, S. M., King, G., & Porro, G. (2009). CEM: Software for coarsened exact matching. *Journal of Statistical Software*, 30(9), 1—27.
- [65] Iacus, S. M., King, G., & Porro, G. (2011). Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association*, 106(493), 345—361.
- [66] Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, 20(1), 1—24.
- [67] King, G., Nielsen, R., Coberley, C., Pope, J. E., Wells, A., Abadie, A., Meirowitz, A. (2011). Comparative effectiveness of matching methods for causal inference. *Political Analysis*. <https://gking.harvard.edu/publications/comparative-effectiveness-matching-methods-causal-inference>
- [68] Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1), 33—50.
- [69] Mcvey, M. (2012). Shadow education: Private supplementary tutoring and its implications for policy makers in Asia. *International Review of Education*, 58(6), 809—

- 811.
- [70] Norris, P. (2001). *Digital divide: Civic engagement, information poverty, and the internet worldwide*. Cambridge, UK: Cambridge University Press.
- [71] Reeder, A. L., & Conger, R. D. (1984). Differential mother and father influences on the educational attainment of black and white women. *Sociological Quarterly*, 25 (2), 239—250.
- [72] Rosenbaum, P. R., & Rubin, D. B. (1983). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society*, 45 (2), 212—218.
- [73] Schurer, S., Alspach, M., Macrae, J., & Martin, G. (2015). The medical care costs of mood disorders: A coarsened exact matching approach. *Economic Record*, 92 (296), 81—93.
- [74] Stevens, G. A., King, G., & Shibuya, K. (2010). Deaths from heart failure: Using coarsened exact matching to correct cause-of-death statistics. *Population Health Metrics*, 8 (1), 6.
- [75] Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science A Review Journal of the Institute of Mathematical Statistics*, 25 (1), 1—21.
- [76] Tichenor, P. J., Donohue, G. A., & Olien, C. N. (1970). Mass media flow and differential growth in technology. *Public Opinion Quarterly*, 34 (2), 159—170.

附录表格

表 1 泛精确匹配结果

	全部样本		非农村户籍		农村户籍流动		农村户籍未迁移	
	使用	未使用	使用	未使用	使用	未使用	使用	未使用
匹配成功样本数	453	831	142	129	108	134	152	458
未匹配成功样本数	383	1210	280	146	79	330	64	827
匹配前多元不平衡性指标 L_1	0.745		0.719		0.736		0.734	
匹配后多元不平衡性指标 L_1	0.264		0.252		0.174		0.282	

注:泛精确匹配中使用的变量包括母亲学历、母亲职业、人均家庭收入、城乡和省份。

表 2 泛精确匹配前、后的平衡性检验

	多元非 平衡性 指标 L1	均值差	最值和分位数点均值差				
			最小值	25%	50%	75%	最大值
泛精确匹配前							
母亲学历	0.438	1.379	0	2	1	1	3
母亲职业	0.432	0.894	0	1	1	1	0
家庭人均收入对数	0.311	0.741	2.485	0.765	0.563	0.622	0.942
地区	0.185	0.434	0	1	0	0	0
城乡	0.355	0.355	0	0	1	0	0
泛精确匹配后							
母亲学历	2.2e-16	6.2e-15	0	0	0	0	.
母亲职业	1.4e-16	3.1e-15	0	0	0	0	0
家庭人均收入对数	0.057	0.044	4.159	0.017	0.003	0.029	0.635
地区	4.3e-16	2.2e-16	0	0	0	0	0
城乡	7.8e-16	5.6e-16	0	0	0	0	0