

家庭收入和教育支出的关联分析

——基于中国城镇住户调查数据的研究

万相昱 唐亮 张晨*

内容提要 教育支出和收入之间存在着何种关系，一直是近期研究关注的焦点。基于中国城镇住户调查数据，我们构建了多元回归模型和断点回归模型，来检验教育支出与家庭收入之间的关系。研究表明，中国城镇居民的教育支出并不存在高收入家庭和低收入家庭的不一致特征。同时，教育支出和家庭收入之间呈现了显著的倒U型关系，说明低收入家庭看待教育的回报更加缺乏理性。总体上来看，中国城镇居民对待教育并非理性人的态度，因此导致较高的教育支出水平。

关键词 教育支出 家庭收入 断点回归

一 引言

家庭教育支出的决策行为本质上是跨期风险决策，其行为影响着家庭的收入、消费以及生活状态。在教育支出的相关研究中，一个主流的观点是，当家庭面临信贷约束时，其财富和收入状况就会对家庭的教育开支产生重大影响（Becker & Tomes, 1994; Lochner & Monge-Naranjo, 2012）。学者们的观点在于，信贷约束使得不同家庭的教育支出存在重大区别，这些观点在印度、美国等国家得到了研究者的很好验证。

* 万相昱，中国社会科学院数量经济与技术经济研究所，电子邮箱：45440549@qq.com；唐亮，东北师范大学商学院，电子邮箱：tangl123@nenu.edu.cn；张晨，中国社会科学院数量经济与技术经济研究所，电子邮箱：2777913876@qq.com。本文是教育部青年基金项目“宏观金融不稳定——微观机理和传导扩散”（批准号：12YJC790173）和中国社会科学院创新工程项目的阶段性成果。

在中国,大多数研究者认为,家庭收入是影响教育支出最重要的因素。此外,父母对子女的期望、住房价值等因素都会影响家庭教育支出,也有学者研究了家庭的文化背景、信贷约束等问题对教育支出的影响,大多得到显著相关的结论。

从经济理论的角度来看,传统观念认为,家庭经济状况和教育支出之间的关系是重要的。从消费的角度看待教育支出,若家庭进行教育支出决策时,视教育支出为普通商品,则教育支出和家庭经济条件正相关;若教育支出被视为奢侈品,则收入越高的家庭将享受更高的教育支出。若从投资的角度看待教育支出,考虑到投资的回报率和教育支出的成本,家庭将进行跨期风险决策。显然,对待教育支出的看法不同,则家庭经济条件和教育支出的关系也就不同。然而,从中国家庭的角度来看,家庭的教育支出显然更加功利,即倾向于以投资的视角看待教育支出;然而,和理性人对投资的态度不同,中国家庭看待教育支出更多地关注“幸存者效应”,即高估教育支出带来的远期回报。这意味着中国家庭教育支出的波动性较小,经济条件对教育支出的解释可能并不显著。

若考虑到家庭收入对教育支出的限制,则可以预期低收入家庭在收入提高后,会提高教育支出水平;而当教育支出达到较高水平后,家庭的教育支出水平将不再继续增加。因此,可以预期,教育支出和收入之间存在着倒U型的曲线关系。若这种关系成立,意味着中国家庭的教育支出一直都处于较高水平,且收入仅仅是教育支出的瓶颈,但并不影响教育支出的增长;当收入超过最高教育支出水平后,教育支出将维持不变,这将和传统的相关研究存在较大区别。若家庭教育支出显著地受到收入的影响,那么高收入群体和低收入群体对待教育支出的态度将会存在区别。因此,可以针对家庭收入水平高低,分别研究不同群体的教育支出函数;在这一点上,断点回归可以很好地帮助我们分析家庭收入高低对教育支出的影响机制差异,如前文所述,经济条件对教育支出的解释可能并不显著,因此,断点回归的结果可能是并不显著的。为了进一步描述经济条件和教育支出的关系,我们构建了收入和教育支出关系的多元回归模型,并侧重分析教育支出和收入的非线性关系;我们发现,教育支出和收入之间存在着显著的倒U型关系,这验证了我们的假设。

由于家庭教育支出可以划分为自主性教育支出和非自主教育支出,前者指并非学校提供的教育支出,即家庭自发选择的学校外的教育服务,后者则是指子女所受学校教育而进行的教育支出。因此,在实证分析中,我们针对自主性教育支出和总教育支出分别进行研究,考察两者的差异。文章的结构安排如下:第二部分是相关文献综述;第三部分是数据和模型设定;第四部分是实证检验;文章的最后是结论。

二 相关理论和文献

关于家庭教育支出影响因素的研究，研究者分别从收入、家庭特征（主要是家庭成员的年龄、性别、受教育程度）等角度进行了分析；其中，大多数研究者集中研究了收入对教育支出的影响机制。

（一）收入对于教育支出的影响

李普亮和贾卫丽（2010）通过对广东省 446 个农村家庭进行问卷调查，发现家庭在校子女数量与教育支出呈现显著的负相关关系，而家庭收入水平与教育支出的正相关关系较为显著。蹇昊（2010）采用 2009 年城乡居民家庭的统计数据，实证检验了城乡居民收入差距和教育支出之间的格兰杰因果关系，发现从短期来看，城镇居民的教育支出集中在能够短期获益的教育培训等支出，农村的短期教育支出显著增加了人均纯收入，但是长期来看教育支出的增加抑制了收入水平的提高。李亚伟和刘晓瑞（2011）基于 2006 年中国健康与营养调查（CHNS）数据对中国居民家庭教育支出的影响因素进行分析，他们采用分层模型发现，社区居民的收入水平显著影响家庭教育支出。余红玲和万俊毅（2012）对农村家庭教育支出变动的原因进行探究，发现家庭人均纯收入的增加会显著增加教育支出，但是达到一定数量时，教育支出的增速要显著低于收入的增速，此时教育支出和收入呈现负相关关系。谷宏伟和杨秋平（2013）对家庭收入、子女受教育程度的期望值和教育支出的关系进行实证分析，发现这两个因素是影响教育支出的关键。

张锦华等（2014）通过建立基于家庭效用最大化的教育支出选择模型，发现家庭收入的不确定性和投资教育所获收益的不确定性对教育支出倾向产生负向调节作用，较为贫困的家庭这种调节作用更加明显。谷宏伟和杨秋平（2014）采用大连市家庭调查数据，对收入与家庭教育支出的关系进行实证检验，发现收入与家庭教育支出呈现显著的正相关关系。钱晓焯等（2015）采用 2007 年和 2011 年国家统计局关于家庭教育支出的专项调查数据，对义务教育阶段的家庭教育支出不平等性进行分析，对比发现家庭收入越低，教育支出所占的比例越高，低收入家庭的教育负担越来越重。周雪涵和张羽（2015）采用济南市高中生的家庭教育支出调查数据，实证发现家庭教育支出与家庭收入之间并不存在显著的相关关系。高澎涛（2015）采用因子分析方法研究中国各个地区的农村家庭收入结构，通过双对数回归发现，工资性收入对于农村家庭教育支出的影响最大，认为收入结构的地区性差异能够解释中部、东部和西部地区的

农村教育支出差异。周红莉和冯增俊（2016）从恩格尔定律出发，发现1981-2009年中国城乡家庭收入与教育支出呈现正相关关系，且教育需求弹性较大。易行健等（2016）采用2002年和2009年中国城镇住户调查数据，对教育支出的影响因素进行研究，发现在教育支出的各项费用中，培训费用受家庭可支配收入的影响最大。

（二）其他家庭特征对于教育支出的影响

李亚伟和刘晓瑞（2011）基于2006年中国健康与营养调查（CHNS）数据发现，户主受教育水平的提高会显著增加家庭的教育支出。谷宏伟和杨秋平（2014）发现，收入和子女数量与家庭教育支出呈现显著的正相关关系。赵静（2014）将养老保险纳入到教育支出的分析框架中，采用2002-2009年中国城镇居民调查数据，实证发现养老保险缴费率对于教育支出有显著正向影响，子女不同教育阶段的教育支出存在显著差异，户主的受教育水平也显著影响家庭教育支出。陈永伟等（2014）采用2010年中国家庭追踪调查（CFPS）数据，对家庭住房价值与教育支出的关系进行研究，发现住房价值和家庭教育支出之间呈现显著的倒U型关系，且当住房价值过高和过低时对于教育支出的影响都不显著，只有处于中间阶段时这种显著关系才会存在。

周雪涵和张羽（2015）采用济南市高中生的家庭教育支出调查数据实证发现，家庭社会经济地位、中考成绩和学生性别与教育支出显著相关，但是对于有关学校的相关教育费用，男女性别并无显著差异。易行健等（2016）采用2002年和2009年中国城镇住户调查数据实证发现，教育支出与户主的年龄呈现显著的U型关系，女孩能够获得比男孩更多的教育资源，另外，城镇人口规模、家庭成员受教育程度对于教育支出也有显著的影响。另外，还有探究政府行为对家庭教育支出影响的研究，袁诚等（2013）在教育支出影响因素的研究中引入地方政府财政支出，采用2002-2009年九个省份的城镇住户调查数据实证发现，政府教育支出对于家庭教育支出、义务教育学杂费和家教费用具有显著的替代效应，且对于家教费用的替代效应在不同收入家庭有着显著的不同，中高收入家庭这种替代效应转化为互补效应。

通过对已有文献的归类，学术界对于家庭教育支出影响因素的研究主要分为家庭收入和家庭特征（如成员受教育程度、子女性别、家庭成员年龄等）两个方面，且对于家庭收入和家庭教育支出关系的研究居多。大部分研究得出的结论较为一致，即家庭收入的提高将会显著增加家庭教育支出，但是余红玲和万俊毅（2012）的研究发现，家庭人均纯收入和教育支出可能存在倒U型关系。关于家庭特征对教育支出影响的文献相对较少，也没有得到具有普遍性的结论。另外，在实证方法上，已有的研究主要集中在多元回归分析，探究的是各种因素和家庭教育支出的相关关系，即使存在显著相关

性也无法说明因果关系的存在，比如家庭收入提高后，教育支出是否会提高。和以往多数研究有所不同的是，本文与余红玲和万俊毅（2012）的观点较为一致，认为家庭收入和教育支出之间可能存在倒 U 型关系，因此首先对收入和家庭教育支出的相关关系进行实证检验。如果存在相关关系，则进一步检验这种相关关系是否可能是因果关系。

三 数据和模型

（一）样本和数据

研究样本来自于国家统计局 2009 年进行的中国城镇住户调查（UHS），调查地区涵盖北京、浙江、广东、辽宁、陕西和四川六个省（市），共计 224 个县（市）。调查内容主要包括城镇居民家庭成员的基本情况、家庭基本情况、现金收支和消费支出等情况。根据本文的研究内容，主要从调查数据中提取教育费用（消费支出表）、家庭成员的人数、受教育程度、年龄、性别（家庭成员基本情况表）以及家庭收入和个人收入情况（现金收支表），详细数据说明如下。

问卷中关于教育支出的调查分为两部分，第一部分是教材，包括课本及参考书、教育软件和其他教材；第二部分是教育费用，包括非义务教育学杂费、义务教育学杂费、托幼费、成人教育费、家教费、培训班、学校住宿费以及其他教育费用。家庭成员的基本信息中，受教育程度由与文化程度相关的问题获得，问卷将文化程度分为 9 个层次，1 表示未上过学，2 表示扫盲班，3 表示小学，4 表示初中，5 表示高中，6 表示中专，7 表示大学专科，8 表示大学本科，9 表示研究生（硕士和博士）。本文将问卷中的文化程度转换为受教育年限，文化程度 1、2 的受教育年限为 0 年，文化程度 3 的受教育年限为 6 年，文化程度 4 的受教育年限为 9 年，文化程度 5、6 的受教育年限为 12 年，文化程度 7 的受教育年限为 15 年，文化程度 8 的受教育年限为 16 年，文化程度 9 的受教育年限为 19 年。另外，其他信息（如性别、年龄、收入等）均可以直接从问卷调查数据中获得。

（二）变量及其度量

本文的主要研究内容是教育支出和家庭经济条件的关系，因此，研究变量主要包括：教育支出变量、家庭经济条件变量以及其他可能影响教育支出的变量，如家庭成员受教育程度、子女性别、子女受教育阶段等等。本文根据问卷对教育支出的调查，将教育支出分为自主性教育支出和非自主性教育支出。非自主性教育支出主要指的是完成义务教育所必须的花费，也就是问卷中的义务教育学杂费；由于义务教育的强制

性，这一部分教育费用原则上是所有拥有适龄儿童和青少年的家庭必须支出的，往往和家庭的收入并无明显相关性。自主性教育支出指的是，家庭可以根据自身的经济条件或者其他因素自主进行调控的支出，比如子女的培训班费用、子女的学习工具费用等，这一部分费用在不同收入水平的家庭中可能存在差异，是本文的研究重点。不同家庭处于上学阶段的子女数量可能不尽相同，因此，我们将教育支出除以家庭上学人数，即用家庭人均教育支出来度量有关教育支出的变量。

本文用家庭收入来度量家庭的经济条件，由于家庭规模的不同，所以用家庭人均收入来度量家庭的经济条件更加合理。对于可能影响教育支出的其他变量，本文控制了家庭最高学历、父母学历差距、子女性别、子女受教育阶段、父母的相对收入和有收入成员的比重。家庭最高学历指的是家庭成员所获得的最高学历，将其转化为受教育年限进行度量；父母学历差距指的是，父母文化程度如果相差 2 个等级及以上就认为差距较大，此变量为虚拟变量，如果差距较大，记为 1，否则记为 0；子女性别也是虚拟变量，如果全为女性，则记为 1，否则记为 0；子女受教育阶段，指的是年龄最大的子女的受教育阶段，将其转换为受教育年限进行度量；父母的相对收入同样是虚拟变量，如果妻子的收入高于丈夫，那么记为 1，否则记为 0；家庭有收入成员比重，用家庭中有收入的成员数量除以家庭总成员数量来度量。变量及其度量如表 1 所示。

表 1 变量及其度量

	变量	度量	符号
被解释变量	家庭人均总教育支出	家庭总教育支出除以上学人数	<i>Aep</i>
	家庭人均自主性教育支出	家庭自主性教育支出除以上学人数	<i>Autoaep</i>
解释变量	家庭经济条件	家庭总收入除以人数	<i>Ainc</i>
控制变量	家庭最高学历	家庭学历最高成员的受教育年限	<i>Hedulev</i>
	父母学历差距	如果父母文化程度差距 2 个等级及以上，等于 1，否则等于 0	<i>Edugap</i>
	子女性别	全为女性，等于 1，否则等于 0	<i>Childsex</i>
	子女受教育阶段	年龄最大的子女受教育年限	<i>Childedu</i>
	父母相对收入	妻子收入高于丈夫，等于 1，否则等于 0	<i>Relinc</i>
	有收入成员比重	有收入成员数量除以总人数	<i>Incratio</i>

(三) 研究设计

1. 多元回归

多元回归的目的是为了检验家庭经济条件和教育支出是否存在非线性关系，在线性回归模型的基础上加入了家庭经济条件变量（即 *Ainc*）的二次项，模型如式（1）

所示：

$$Aep_i = \alpha_0 + \alpha_1 Ainc_i + \alpha_2 Ainc_i^2 + \alpha_3 Hedulev_i + \alpha_4 Edugap_i + \alpha_5 Childsex_i + \alpha_6 Childed_u_i + \alpha_7 Relinc_i + \alpha_8 Incratio_i + \varepsilon_i \quad (1)$$

其中，被解释变量 Aep_i 为家庭人均总教育支出；解释变量中， $Ainc_i^2$ 为 $Ainc_i$ 的平方项。模型 (2) 的被解释变量 $Autoaep_i$ 为家庭人均自主性教育支出：

$$Autoaep_i = \alpha_0 + \alpha_1 Ainc_i + \alpha_2 Ainc_i^2 + \alpha_3 Hedulev_i + \alpha_4 Edugap_i + \alpha_5 Childsex_i + \alpha_6 Childed_u_i + \alpha_7 Relinc_i + \alpha_8 Incratio_i + \varepsilon_i \quad (2)$$

2. 断点回归

断点回归方法是一种拟随机实验方法，用于检验因果关系。在断点回归中，存在一个变量，当这个变量大于某一个临界值时，那么个体就可能受到处置；小于该临界值时，个体就可能不会受到处置。如果个体受到处置的概率为 1，就是精确断点回归；如果受到处置的概率小于 1，就是模糊断点回归。由于本文认为，当收入低于某一临界值时家庭经济条件相对较差，反之就较好，因此可以采用精确断点回归方法。当处理状态 D_i 是变量 x_i 的确定不连续函数时，就可以使用精确断点回归：

$$D_i = \begin{cases} 1, & \text{if } x_i \geq x_0 \\ 0, & \text{if } x_i < x_0 \end{cases} \quad (3)$$

假设变量 Y_{0i} 和变量 x_i 之间的关系是线性的，并且因果效应是常数，那么：

$$E[Y_{0i} | x_i] = \alpha + \beta x_i \quad (4)$$

$$Y_{1i} = Y_{0i} + \rho \quad (5)$$

其中， Y_{0i} 表示未接受处置的结果变量， Y_{1i} 表示接受处置后的结果变量， ρ 为常数。通过整理得到：

$$Y_i = \alpha + \beta x_i + \rho D_i + \eta_i \quad (6)$$

其中， ρ 为我们需要估计出的因果效应，这里的虚拟变量 D_i 和以往虚拟变量回归不同的是， D_i 与 x_i 相关，并且是 x_i 的不连续函数。

如果 $E[Y_{0i} | x_i]$ 是非线性的，那么估计方程就变为：

$$Y_i = f(x_i) + \rho D_i + \eta_i \quad (7)$$

我们可以用 p 次多项式来模型化 $f(x_i)$ ，于是：

$$Y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \dots + \beta_p x_i^p + \rho D_i + \eta_i \quad (8)$$

对于式 (8)，因果效应估计的有效性依赖于多项式模型能否精确地刻画 $E[Y_{0i} | x_i]$ 。如果不能，那么看上去发生的跳跃其实只不过是模型刻画错误带来的。为了避免这样

的错误,我们只考虑在不连续的邻域里的数据,也就是考虑区间 $[x_0 - \Delta, x_0 + \Delta]$, Δ 是一个很小的正数,于是有:

$$E[Y_i | x_0 - \Delta < x_i < x_0] \cong E[Y_{0i} | x_i = x_0] \quad (9)$$

$$E[Y_i | x_0 < x_i < x_0 + \Delta] \cong E[Y_{1i} | x_i = x_0] \quad (10)$$

因此:

$$\begin{aligned} & \lim_{\Delta \rightarrow 0} E[Y_i | x_0 < x_i < x_0 + \Delta] - E[Y_i | x_0 - \Delta < x_i < x_0] \\ & = E[Y_{1i} - Y_{0i} | x_i = x_0] \end{aligned} \quad (11)$$

也就是说,在 x_i 足够小的邻域中比较 Y_{1i} 和 Y_{0i} 的均值就可以得到因果效应估计,这样就与 $E[Y_{0i} | x_i]$ 的具体形式无关。另外,采用非参数的局部线性回归等局部回归方法对靠近临界值的点赋予更大的权重,则能够避免临界值周围数据样本少而带来的估计不准确问题。断点回归对临界值两端采用局部多项式回归的方法进行处置效应估计,而并不针对全样本进行估计,这样能够很好地控制临界值周围的噪声。将局部多项式回归的拟合结果通过图形展现出来,可以对临界值周围的效应情况有更加直观的认识。因果效应的检验有两个途径,第一个是在临界值两侧分别进行局部线性回归,分别计算其方差和协方差来构造检验统计量,第二个方法是将临界值两侧的局部线性回归放置于同一个回归方程中直接进行检验:

$$\min \sum_{i=1}^N (Y_i - \alpha - \beta \cdot (x_i - c) - \rho \cdot D_i - \gamma \cdot (x_i - c) \cdot D_i)^2 \quad (12)$$

需要指出的是,这里 x_i 的取值范围为 $[c - \Delta, c + \Delta]$,也就是只对 $c - \Delta \leq x_i \leq c + \Delta$ 这一区间的样本进行目标函数(即公式(12))的最小化。

四 实证检验

实证检验包括三部分,首先是对研究所涉及变量的描述性统计;接下来是多元回归的结果,用于检验教育支出和家庭经济条件之间的关系;最后是断点回归的结果,用于检验家庭经济条件是否构成教育支出的原因。

(一) 描述性统计

表2给出了研究变量的描述性统计,剔除问卷数据不完整的样本,最终用于研究的样本数量为2729个。首先是教育支出变量的统计,人均自主性教育支出(Autoaep)的平均值为4056.24元,最小值为50元,最大值为144618元,说明自主性教育支出在不同家庭之间具有较大的差异,标准差为7011.16元,数值较大。另外,人均总教育

支出 (*Aep*) 的标准差为 7435.65 元, 和人均自主性教育支出的水平较为接近; 两者平均值的差距大约为 300 元, 差距并不大, 这似乎能够说明非自主性教育支出 (义务教育费用) 在家庭之间差距并不是很大, 而且绝对值较小。衡量家庭经济条件的家庭人均收入 (*Ainc*) 最小值为 360 元, 最大值为 393490 元, 标准差为 22097.65 元, 说明不同家庭的经济条件差距较大。

表 2 变量描述性统计

	<i>Autoaep</i> (元)	<i>Aep</i> (元)	<i>Ainc</i> (元)	<i>Hedulev</i> (年)	<i>Edugap</i>	<i>Childsex</i>	<i>Childeddu</i> (年)	<i>Relinc</i>	<i>Incratio</i>
平均值	4056.239	4390.622	12649.879	13.999	0.221	0.362	10.067	0.274	0.623
中位数	2042.840	2336.914	6616.560	15	0	0	9	0	0.667
众数	500	500	4000	15	0	0	6	0	0.667
标准差	7011.162	7435.649	22097.646	2.368	0.415	0.481	3.640	0.446	0.142
最小值	50	235	360	6	0	0	0	0	0
最大值	144618	144618	393490	19	1	1	19	1	0.979
观测数	2729	2729	2729	2729	2729	2729	2729	2729	2729

资料来源: 根据 2009 年中国城镇住户调查数据计算得到。

另外, 家庭成员最高受教育年限 (*Hedulev*) 的最大值为 19 年, 即学历为研究生; 而最小值是 6 年, 学历为小学; 众数为 15 年, 说明样本中大部分家庭的最高学历是大学专科。夫妻之间学历差距 (*Edugap*) 的众数为 0, 差距大的样本 (即 *Edugap* 等于 1) 占比约为 22.1%, 说明大部分家庭的夫妻学历差距并不大。子女性别 (*Childsex*) 全为女性的样本占比为 36.2%。子女受教育阶段 (*Childeddu*) 的众数为 6 年, 也就是仍处于小学阶段; 平均值约为 10 年, 标准差为 3.64 年, 集中在中学阶段。夫妻相对收入 (*Relinc*) 变量中, 妻子收入高于丈夫的样本数量占总样本量的 27.4%, 比例并不高。家庭有收入成员比重 (*Incratio*) 的统计结果显示, 家庭中有收入成员的比重均值是 62.3%, 众数是 66.7%, 说明大部分家庭的这项指标比较接近。

(二) 多元回归

为了检验教育支出和家庭经济条件之间非线性关系 (倒 U 型) 的存在, 分别用家庭人均教育总支出和家庭人均自主性教育支出对家庭人均收入及其平方项 (均取自然对数) 进行回归。如果倒 U 型关系存在, 那么家庭人均收入二次项的系数将会显著小于 0。

表 3 列示了家庭人均教育总支出对家庭人均收入回归的结果, 家庭人均收入 (取对数) 的系数显著大于 0, 而其二次项 (取对数) 的系数显著小于 0, 说明家庭人均

教育总支出和家庭人均收入之间的倒 U 型关系是显著存在的。家庭人均教育总支出并不会随着家庭收入的增加而一直增加，当家庭经济条件提高到一定程度时，家庭人均教育总支出会减少。家庭成员最高学历变量的系数显著大于 0，说明家庭成员的受教育程度越高，家庭的人均教育总支出越高。夫妻学历差距变量的系数显著小于 0，说明夫妻之间的学历差距越大，家庭人均教育总支出就越低。子女性别变量的系数显著大于 0，说明子女全是女性的家庭人均教育总支出要高于子女是其他性别组合的类型。子女受教育阶段变量的系数显著大于 0，表明子女的学习阶段越高，家庭人均教育总支出就越高。家庭有收入成员比重变量的系数显著大于 0，说明有收入的家庭成员越多，家庭人均教育总支出就越高。最后，父母相对收入变量的系数估计结果并不显著。

表 3 家庭人均教育总支出对家庭人均收入的回归结果

变量	系数	t 值	P 值
家庭人均收入的对数	0.151 ***	6.12	0.000
家庭人均收入二次项的对数	-0.012 ***	-8.33	0.000
家庭成员最高学历	0.117 ***	9.07	0.000
夫妻学历差距	-0.173 **	-2.47	0.013
子女性别	0.125 **	2.11	0.035
子女受教育阶段	0.032 ***	3.95	0.000
父母相对收入	0.042	0.66	0.510
有收入成员比重	2.631 ***	12.89	0.000
常数项	2.545 ***	8.71	0.000
调整后的 R ²	0.134		
F 值	53.6 ***		

注：*、**和***分别表示在 10%、5% 和 1% 水平下显著。
资料来源：根据 2009 年中国城镇住户调查数据计算得到。

表 4 是家庭人均自主性教育支出对家庭人均收入回归的实证结果。和表 3 类似，倒 U 型关系依然显著存在，其余解释变量中，除夫妻学历差距并不显著地影响家庭人均自主性教育支出外，变量的符号和显著性均和表 3 类似，说明家庭人均自主性教育支出和总教育支出在总体层面上呈现了一致的关系。上述实证结果表明，家庭教育支出和收入之间存在的关系和已有研究的观点不同，收入和教育支出之间存在着显著的倒 U 型关系，而非简单的线性关系。

表 4 家庭人均自主性教育支出对家庭人均收入的回归结果

变量	系数	t 值	P 值
家庭人均收入的对数	0.182 ***	6.91	0.000
家庭人均收入二次项的对数	-0.011 ***	-7.32	0.000
家庭成员最高学历	0.116 ***	8.47	0.000
夫妻学历差距	-0.104	-1.40	0.161
子女性别	0.191 ***	3.05	0.002
子女受教育阶段	0.038 ***	4.38	0.000
父母相对收入	0.058	0.87	0.386
有收入成员比重	2.195 ***	10.11	0.000
常数项	2.308 ***	7.42	0.000
调整后的 R ²	0.113		
F 值	44.54 ***		

注：*、**和***分别表示在10%、5%和1%水平下显著。

资料来源：根据2009年中国城镇住户调查数据计算得到。

(三) 家庭经济条件对教育支出的因果关系检验

我们用教育支出对家庭经济条件进行断点回归，用于检验二者之间是否具有因果关系。本文将临界值设为12000元^①，用于判断是否高收入家庭更倾向于高的教育支出。图1是家庭人均收入和人均教育总支出的断点回归图（图中曲线是局部多项式回归的拟合曲线）。横坐标表示家庭人均收入的自然对数减去临界值的自然对数，纵坐标是人均教育总支出的自然对数，由于所选择的临界值是12000元，所以横坐标等于0的位置就是家庭人均年收入为12000元的情况。从图1中可以看到，此处并无明显的断点存在，也就是0附近的人均教育总支出并无明显的差异，从图中并不能说明高收入家庭更倾向于高的人均教育总支出。从图中可以看到其他家庭人均收入水平附近的人均教育总支出均不存在明显的跳跃情况，说明将家庭人均收入的其他水平作为断点时，均不能检测出人均教育总支出存在显著差异。

图1只是直观地呈现了检验结果，而对于临界值两端的因果效应估计结果如表5所示。本文给出三种估计方法，误差修正方法能够消除传统估计方法中渐进误差的偏误，但是往往只对大的带宽有效；稳健估计法使得误差修正方法能够在小带宽情况下

^① 本文将最初的断点设定为家庭人均年收入的平均水平（本文家庭人均年收入的均值为12649.88元，断点回归图中的曲线是以每个样本点作为断点进行估计得到的，因此，从图中也可以看到除了均值外其他样本点作为断点时的情况）。

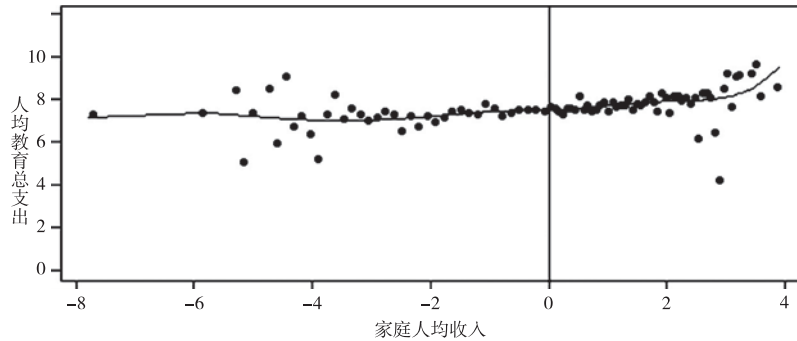


图1 家庭经济条件和家庭人均教育总支出的断点回归图

注：横坐标表示家庭人均收入的自然对数减去临界值的自然对数；纵坐标表示人均教育总支出的自然对数。
资料来源：根据2009年中国城镇住户调查数据计算得到。

也能够有效。从表5的估计结果及其显著性可以看到，三种估计方法均不能得到显著的因果效应，也就是高收入家庭并没有倾向于更高的教育支出。

表5 处置效应估计：家庭经济条件和家庭人均教育总支出的因果关系检验

估计方法	效应估计	标准误	z 值	P 值
传统估计	0.057	0.176	0.325	0.745
误差修正	0.077	0.176	0.434	0.662
稳健估计	0.077	0.21	0.365	0.715

资料来源：根据2009年中国城镇住户调查数据计算得到。

断点回归的第二部分是对家庭经济条件和家庭人均自主性教育支出的因果关系检验（图2横坐标和纵坐标的计算方法同图1）。本文认为自主性教育支出往往直接和家庭收入有关，这一部分是家庭根据自身的经济条件和其他因素自行调整的，收入高的家庭可能更加倾向于对子女的教育投入，比如买更加昂贵的学习工具、上高质量的培训班等等。无论从断点回归图（见图2）还是处置效应的估计结果（见表6）来看，和教育总支出的情形类似，高收入家庭并没有倾向于更高的自主性教育支出。

表6 处置效应估计：家庭经济条件和家庭人均自主性教育支出的因果关系检验

估计方法	效应估计	标准误	z 值	P 值
传统估计	-0.06	0.178	-0.336	0.734
误差修正	-0.072	0.178	-0.404	0.686
稳健估计	-0.072	0.213	-0.337	0.734

资料来源：根据2009年中国城镇住户调查数据计算得到。

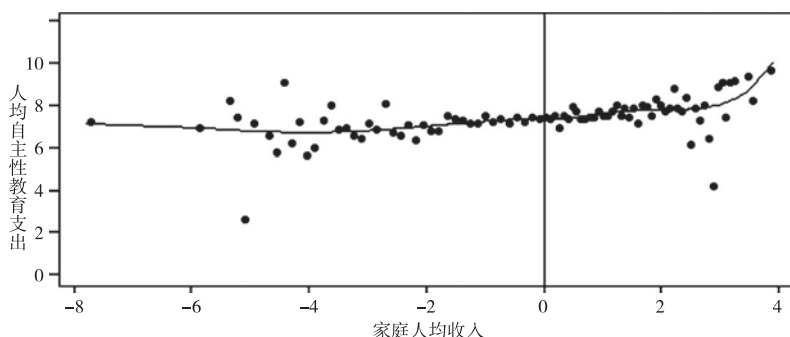


图2 家庭经济条件和家庭人均自主性教育支出的断点回归图

注：横坐标表示家庭人均收入的自然对数减去临界值的自然对数；纵坐标表示人均自主性教育支出的自然对数。

资料来源：根据2009年中国城镇住户调查数据计算得到。

五 结论

教育支出和收入之间的关系究竟呈现什么样的规律？一直以来，研究者们分别从城乡差异、预算约束、信贷约束、财富、收入等多个角度进行分析，然而，收入和教育支出的关系仍然存在着很多争论。在这个基础上，我们利用中国城镇住户调查数据，对教育支出和收入之间的关系进行多元回归估计和断点回归估计，得到了以下结论。

第一，高收入家庭和低收入家庭的教育支出并不存在显著的断点。高收入家庭的教育支出会更多这一假设被很多研究者所接受，然而我们的实证研究表明，这一断点并不存在。这意味着，家庭对教育支出的态度和理性经济人的假设并不一致，家庭视教育支出为最重要的投资，教育支出往往存在一个上限，并不能随着收入提高而一直增加，低收入家庭会努力使自身的教育支出接近这一上限，而高收入家庭虽然收入比较高，但是已经达到上限并不能够继续提高教育支出。

第二，收入和教育支出之间存在倒U型关系。当家庭收入水平较低时，收入水平提高会提高教育支出，这种提高往往和收入水平增长速度相一致；当收入足以支付较高教育支出水平后，教育支出占收入的比重开始降低，这意味着教育支出存在着上限，而无论高收入家庭还是低收入家庭，其目的都是为了给子女提供最高水平的教育。

第三，收入较高的家庭对待教育支出更加理性。教育支出所带来的回报和家庭成本之间的权衡是理性决定教育支出的经济前提。由于收入和教育支出存在倒U型关系，

低收入家庭进行教育支出付出了更高的机会成本，而其教育支出的回报却是不变的，因此，低收入家庭的高教育支出不符合理性人的假设，而高收入家庭的教育支出呈现下降的趋势，表明其对待教育支出的态度更趋理性。

参考文献：

- 陈永伟、顾佳峰、史宇鹏 (2014), 《住房财富、信贷约束与城镇家庭教育开支——来自 CFPS2010 数据的证据》, 《经济研究》第增 1 期, 第 89 - 101 页。
- 高澎涛 (2015), 《基于因子分析研究收入结构对农户家庭教育支出的影响》, 《金融经济月刊》第 12 期, 第 126 - 129 页。
- 谷宏伟、杨秋平 (2013), 《收入、期望与教育支出：对当前中国家庭教育投资行为的实证分析》, 《宏观经济研究》第 3 期, 第 68 - 74 页。
- 谷宏伟、杨秋平 (2014), 《收入和子女数量对城市家庭教育投资行为的影响——基于大连市的实证研究》, 《宏观经济研究》第 5 期, 第 127 - 134 页。
- 蹇昊 (2010), 《城乡居民的教育投入与收入差距的因果分析——基于江苏省数据的格兰杰检验》, 《江苏商论》第 7 期, 第 39 - 41 页。
- 李普亮、贾卫丽 (2010), 《农村家庭子女教育投资的实证分析——以广东省为例》, 《中国农村观察》第 3 期, 第 73 - 85 页。
- 李亚伟、刘晓瑞 (2011), 《我国居民家庭教育支出的影响因素分析》, 《统计与决策》第 17 期, 第 88 - 91 页。
- 钱晓焯、迟巍、史瑶 (2015), 《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》, 《教育与经济》第 6 期, 第 25 - 33 页。
- 易行健、张家为、杨碧云 (2016), 《家庭教育支出决定因素分析——来自中国城镇住户调查数据的经验证据》, 《南方人口》第 3 期, 第 21 - 35 页。
- 余红玲、万俊毅 (2012), 《农村家庭教育支出变动成因探析》, 《中国农学通报》第 28 卷第 2 期, 第 150 - 156 页。
- 袁诚、张磊、曾颖 (2013), 《地方教育投入对城镇家庭教育支出行为的影响——对我国城镇家庭动态重复截面数据的一个估计》, 《经济学动态》第 3 期, 第 29 - 35 页。
- 张锦华、杨晖、沈亚芳、韩阳 (2014), 《不确定性对城乡家庭教育支出倾向的影响研究》, 《复旦教育论坛》第 6 期, 第 83 - 88 页。

- 赵静 (2014), 《养老保险对家庭教育支出的影响——基于世代交叠模型的分析》, 《中国经济问题》第4期, 第75-87页。
- 周红莉、冯增俊 (2016), 《恩格尔定律下中国家庭收入与教育投入关系的实证研究》, 《当代教育科学》第3期, 第46-49页。
- 周雪涵、张羽 (2015), 《高中阶段家庭教育成本及其影响因素分析》, 《清华大学教育研究》第5期, 第110-117页。
- Becker, Gary & Nigel Tomes (1994). Human Capital and the Rise and Fall of Families. In Gary Becker (ed.), *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. Chicago: University of Chicago Press, pp. 257-298.
- Lochner, Lance & Alexander Monge-Naranjo (2012). Credit Constraints in Education. *Annual Review of Economics*, 4 (1), 225-256.

The Correlation Analysis of Household Income and Education Expenditure: A Research Based on China Urban Household Survey

Wan Xiangyu¹, Tang Liang² & Zhang Chen¹

(Institute of Quantitative & Technical Economics, Chinese Academy of Social Sciences¹;
School of Business, Northeast Normal University²)

Abstract: The relationship between education expenditure and income has been a focus of recent research. Based on China Urban Household Survey data, we build a multivariable regression model and use a regression discontinuity design to examine the relationship between education expenditure and family income. The results show that the education expenditure of high-income families and low-income families of China's urban residents does not exist inconsistent characteristics. Meanwhile, there is a significant inverted U-shape relationship between education expenditure and family income. It shows that low-income families treat the return of education less rationally. Overall, the attitude to education expenditure of China's urban residents is irrational, thus leading to a high level of education expenditure.

Keywords: education expenditure, family income, regression discontinuity

JEL Classification: C21, D12, I21

(责任编辑: 周敏丹)