

基于多重计量偏误的农村代际收入流动分位回归研究*

韩军辉 龙志和

【摘 要】文章利用中国健康与营养调查数据中的农村家庭数据,在综合考虑暂时性收入偏误、生命周期偏误、同住选择和工作选择偏误后运用分位回归对中国农村家庭的代际收入弹性进行估计。实证结果表明,在收入分布的两端,由于个体特征的差异导致农村家庭呈现出较强的代际收入流动性;而在中位数附近则呈现出较低的代际收入流动性。

【关键词】分位回归 计量偏误 代际收入流动

【作 者】韩军辉 太原理工大学经济管理学院,讲师;龙志和 华南理工大学工商管理学院,教授。

一、引 言

目前中国收入差距过大的事实引起了社会各界的关注,而已有研究关于收入差距的成因解释主要集中在现行分配制度的宏观分析上,较少关注和分析代际收入流动问题。事实上,当前中国居民社会经济地位的差异不仅仅体现在“当代人”的生活水平层面,而且随着社会各阶层间的利益分化还出现了较为明显的代际传递和转移。

代际收入流动反映了父辈持久收入对子代持久收入的影响程度。已有文献一般将持久收入看做不随时间变化的恒量,用父辈持久收入对子代持久收入的回归系数(即代际收入弹性)来表示这种效应。在早期的实证研究中,多数学者主要利用某单一年份的收入作为持久收入的替代变量(Soltow, 1965; Sewell 等, 1975)。Solon (1992)明确指出这种方法会产生向下的偏误,这一点在 Mazumder (2001)的研究中也得到证实。纠正计量偏误的另一思路是使用工具变量。但每个与父辈收入高度相关的变量几乎均与子代的收入相关。因此 Solon (1992)认为,不理想的工具变量将会导致估计结果产生向上的偏误。如何较为精准地估算持久收入是相关研究中不可回避的关键问题。

考察代际收入流动性需要注意的另一个问题是应该选取父辈和子代哪一年的收入。Haider 等(2006)认为,当前收入与持久收入在整个生命周期的不同阶段会呈现不同的关

* 本文得到教育部人文社会科学研究青年基金项目(10YJC790075)资助。

系。Grawe(2003)通过实证分析发现,在年老时测量父亲的收入会导致估计结果被压低,同时他们也找到了一些非常有限的依据来支撑代际收入弹性系数与被测年龄成正比这一假设。少数研究考虑了由于样本选择而引起的偏误。具体包括同住选择偏误和工作选择偏误。研究大多运用赫克曼两步法,而事实上,该方法主要是针对不可观测变量选择的矫正,并没有考虑可观测变量的情况。同住选择主要是由于使用了短面板数据造成的。同样,运用赫克曼两步法对“过分节俭”的代际收入流动方程矫正同住偏误也不适合。

随着计量方法的不断改进,学者对代际收入弹性的估算越来越准确。然而,上述文献仅是在“过分”强调某单一计量偏误的条件下对代际收入弹性的均值进行估算,无法度量在收入分布中不同分位数上的收入弹性。事实上,个体特征(如能力、努力等)存在一定差异。利用均值回归得出的代际收入弹性系数不能全面刻画两代人收入之间的传递景象,因此有必要利用分位回归对其进行分析。国内运用该方法的相关研究比较少。王海港(2005)、何晓琦等(2006)、姚先国等(2007)利用相关数据运用最小二乘法对中国居民的代际收入弹性系数进行了估计。魏颖(2009)利用分位回归研究发现,在控制子代受教育年限变量之前,农村代际收入弹性随着分位数的上升呈现出先增后减的“倒U形”变化趋势。该研究虽然在方法上进行了拓展,但只单纯考虑分位回归,而没有将各种样本选择问题纳入分析的框架,更没有考虑暂时性收入偏误和生命周期偏误。

基于上述分析,本文借鉴 Cheti Nicoletti(2008)研究代际职业流动的思路,同时考虑暂时性收入偏误、生命周期偏误、工作选择偏误及同住偏误,利用中国健康和营养调查(CHNS)数据中的农村家庭数据对中国农村居民代际收入弹性进行估算。

二、数据说明

(一) 变量说明

本文所用数据来自中国健康和营养调查(CHNS),选取的变量主要涉及广西、贵州、黑龙江、河南、湖北、湖南、江苏、辽宁和山东共9省份的农村家庭户主与子代的收入、职业、教育、年龄、性别和身体健康状况等。收入变量包括调查问卷中的户主和子代各自的工资收入、奖金、补贴(包括副食补贴、保健补贴、洗理费、书报费、房屋补贴和其他补贴)、务农收入、家庭园艺收入、畜牧收入、渔业收入等项目。教育变量为受教育年限。职业变量按照 Goldthorpe 职业分类表进行等级分类。身体健康状况变量采用国际通用的身体质量指数。该指数用体重(kg)与身高(cm)平方的比值表示。年龄变量进行中心化处理。性别变量用0表示女性,1表示男性。

(二) 主要变量描述性统计

为了检验并纠正各种计量偏误,本文将所有农村家庭样本分为全样本和子样本。全样本是将中国健康与营养调查中的个体教育、个体收入、个体职业及个体健康等数据文件按照 wave=1989、1991、1993、1997、2000、2004、2006 进行处理得到相应年份的数据;并将上述年份个体教育、个体收入、个体职业及个体健康等数据文件中的户主(父亲或母亲)与子代

信息进行分离;最后将同属一个家庭的父辈与子代的数据合并得到父辈与其子代历年的综合数据信息。本文使用的是一组非平衡面板数据。1997 年数据中不仅增加了新成立的 家庭,同时还增加了一些家庭代替不再参加调查的家庭。为了保持调查对象的一致性,本文将该 年新增家庭部分删除,导致该年调查样本量较小。另外,本文假设全样本存在计量偏误,因 此,全样本中历年各变量存在一定的缺失值(见表 1)。

由表 1 可知,在每一轮调查中子代受教育的年限均高于父辈,而且呈现出一定的代际 传递性。“高素质的子代”几乎与“高素质的户主”相对应。本文所涉及的职业变量按照 Goldthorpe 职业分类表进行等级分类。该职业分类表中取值为 2 的职业为农民、渔民或猎 人,取值为 3 的职业为非技术工人或熟练工人(如普通工人、伐木工等)。表 1 中所列出的职 业类型变量表明农村家庭的户主和子代基本从事传统农业或充当普通工人。由工资理论可 知,个体收入主要取决于人力资本(通常用教育变量表示),同时与个体的职业类型密切相 关。因此,上述事实间接说明农村家庭中两代人之间可能存在收入上的继承或传递性。但如 果直接用收入变量来描述这种继承或传递性,则存在一定的问题。因为相关研究文献中所 用收入变量为持久收入,而表 1 中所涉及的收入变量实际上为可观测收入。按照经济学的 理论,可观测收入等于持久收入与暂时性收入之和。因此,若要详细考察代际收入流动性, 必须解决暂时性收入偏误问题并对持久收入进行合理估算。

从理论上讲,父辈与子代的收入都应该在当期收入最接近其持久收入的时点年份进行 测量。由表 1 可以看出,户主平均年龄在 1989 年约为 34 岁,随后逐渐增加。到 2006 年,户 主平均年龄增加到约 54 岁。在相应的 7 轮调查中,子代平均年龄分别为 14 岁、19 岁、20 岁、32 岁、20 岁、24 岁和 26 岁。因此,在估计代际收入弹性时,到底应该选取户主和子代哪一年的 收入是一个值得考虑的问题。图 1 和图 2 分别给出了户主和子代的可观测收入一年龄曲 线。从可观测收入一年龄曲线来看,户主和子代的收入一年龄曲线均呈现倒 U 形,而且在每 一年龄时点对应不同的收入。若不考虑生命周期偏误而“随意”选取某一年龄时点的可观测 收入进行估计将会得出多种估计结果。

1989 年调查数据中有 93% 的户主为男性,这一比例的最低值出现在 1997 年的调查数 据中,但仍然高达 88%。从表 1 可知,在 1989 年和 2000 年调查数据中,儿子所占比例分别 约为 78% 和 70%,女儿所占比例较小。而在另外 5 轮调查中,女儿所占比例有所增加。如 1991 年女儿与儿子几乎各占 50%。从人口学的角度来看,本文样本的选取具有一定的合理 性,但这一事实暗示可能存在工作选择偏误。另外,由于所用数据为短面板数据并且多为 成年人,所以户主和子代的身体健康状况(用身体质量指数表示)变化幅度不大。

在全样本基础上,文章选取在 7 轮调查中至少出现 3 次且共同居住在同一家庭的户主 和子代构成子样本。这样做是为了把同住效应“强加”给子样本,并保持家庭的代表性在一 个较长时间内保持不变。子样本的选取主要依据中国健康与营养调查中相关问题的回答。 如 2000 年住户调查中的“家庭成员基本情况登记表”有“现在还住在家里吗?”“现在他还是 该家庭的成员吗?”等问题,根据对此类问题的回答,我们可以对子代与户主的同住状况做出

表 1 全样本中历年主要变量描述性统计

变量	父辈			子代		
	样本量	均值	方差	样本量	均值	方差
1989 年						
性别	3340	0.932	0.251	2502	0.778	0.416
年龄	1563	34.360	6.842	1036	14.088	11.938
受教育年限	3138	5.389	4.149	1629	5.822	3.304
身体健康	1553	21.421	2.320	965	18.580	3.124
职业类型	3340	2.848	2.541	2502	0.877	1.820
对数收入	3340	7.848	1.042	2502	7.418	1.065
1991 年						
性别	2575	0.889	0.314	1959	0.575	0.494
年龄	2279	44.366	13.923	1658	18.961	17.122
受教育年限	2556	6.017	4.059	1600	6.310	3.176
身体健康	2263	21.494	2.743	1642	18.354	3.612
职业类型	2575	3.250	2.970	1959	1.259	2.098
对数收入	2575	7.806	0.978	1959	7.385	1.031
1993 年						
性别	2433	0.893	0.310	1789	0.596	0.491
年龄	2128	46.354	13.588	1498	20.439	17.571
受教育年限	2384	6.237	4.061	1506	6.551	4.033
身体健康	2114	21.747	2.810	1481	18.615	3.574
职业类型	2433	3.172	2.793	1789	1.370	2.177
对数收入	2069	7.831	1.071	1789	7.496	1.168
1997 年						
性别	384	0.888	0.316	266	0.680	0.467
年龄	384	54.168	9.692	266	32.007	14.121
受教育年限	384	5.253	3.704	261	7.658	3.725
身体健康	339	21.589	2.659	197	21.637	2.756
职业类型	384	2.646	2.469	266	3.008	2.146
对数收入	384	7.999	1.033	266	7.856	1.156
2000 年						
性别	2792	0.913	0.282	1860	0.701	0.490
年龄	2439	49.597	13.346	1428	19.972	14.997
受教育年限	2417	6.131	4.377	1382	6.187	4.695
身体健康	2407	22.733	3.125	1318	19.313	3.832
职业类型	2792	2.931	2.979	1860	1.510	2.463
对数收入	2271	8.322	1.103	1860	8.239	1.170
2004 年						
性别	2431	0.904	0.294	949	0.623	0.485
年龄	2431	52.559	12.841	947	23.755	17.074
受教育年限	2424	7.107	4.057	896	7.361	4.456
身体健康	2299	28.272	4.645	870	25.108	1.268
职业类型	2431	2.415	3.059	949	1.378	2.481
对数收入	1835	8.454	1.178	346	8.038	1.602
2006 年						
性别	2396	0.906	0.292	792	0.628	0.484
年龄	2396	53.963	12.512	792	26.289	18.534
受教育年限	2033	7.248	4.486	519	7.798	3.692
身体健康	2271	24.599	2.435	720	21.093	6.858
职业类型	2396	2.520	2.966	792	1.549	2.591
对数收入	1810	8.694	1.227	792	8.373	1.589

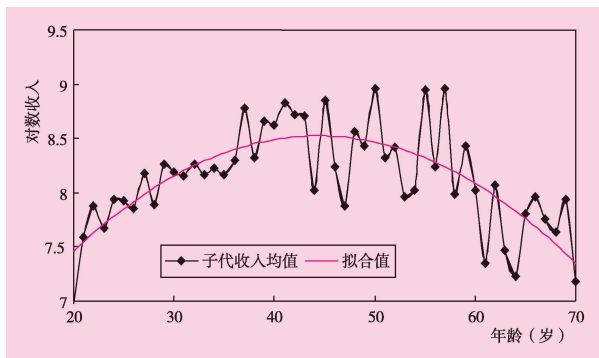


图1 子代可观测收入一年龄曲线

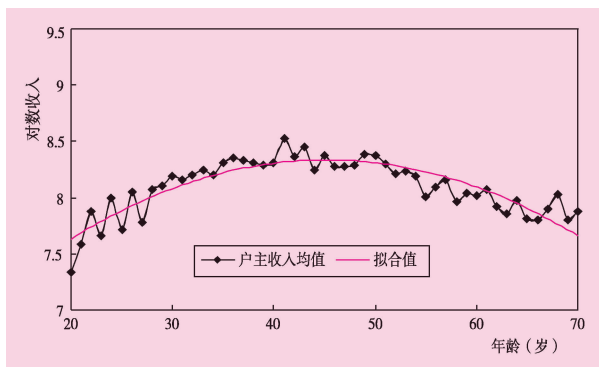


图2 户主可观测收入一年龄曲线

判断,为选取子样本提供依据。限于篇幅未列出子样本主要变量情况。

三、基于多重计量偏误的分位回归模型设定

相关文献中一般将父辈收入对子代收入的均值回归系数作为代际收入弹性。该方法不能度量在收入分布中不同分位数上的收入弹性。实际上,个体特征(如能力、努力等)存在一定差异。在其他条件相同的情况下,具有较强能力或者付出较高的努力后,子女往往会获得较高的收入。这可能导致父辈对“有出息”的子女给予更多的关爱。或者说,父辈收入对子代收入的影响程度可能会因为子代能力或努力程度的差异而表现出不同的“传递场景”。如果这一假设正确的话,那么利用均值回归得出的代际收入弹性系数将不能全

面刻画两代人收入之间的传递景象。另外,本研究中可能存在暂时性收入偏误、生命周期偏误、工作选择偏误和同住偏误。为此,有必要在考虑上述计量偏误的基础上进行分位回归。

(一) 暂时性收入偏误与生命周期偏误分析

本文结合中国健康与营养调查数据的特点提出估算持久收入的方法。需要注意的是按照已有文献的研究思路,本文将持久收入看做是一个不随时间变化的恒量。这里假设 $y_{is} = \lambda_s \eta_{is} + u_{is}$, 其中, y_{is} 为个体 i 在 s 时期(某一年龄)的可观测收入; η_{is} 表示相应的持久收入部分; u_{is} 表示暂时收入冲击,并与持久收入部分不相关^①; λ_s 表示当期收入与持久收入部分的相关系数。该系数(λ_s)接近 1 的年龄段是我们选取数据的最佳区间。在确定选取数据的年龄区间后,可以将持久收入部分(η_{is})的历年均值作为计量模型中的持久收入变量。首先,将个体历年可观测收入对年龄和年龄平方进行回归,并求出回归残差;其次,以历年职业、受教育年限、身体质量指数、地区等变量作为自变量与上述残差进行回归求出拟合方程的预测值。在确定合理年龄区间的基础上,将历年预测值的均值作为持久收入。之所以这样处理,是因为我们认为持久收入是由一系列不随时间变化的因素(如性别、出生地、受教育程

^① 这一假设与 Anders 等(2006)的研究有较大差别。

度、职业等)决定的,如果对众多因素进行多元回归估计持久收入可能会出现共线问题。求解当期可观测收入对年龄和年龄平方的回归残差主要是试图把影响当期收入的随时间变化的因素剔除掉,将不随时间变化的因素所决定的收入部分看做是持久收入部分(η_{is})。另外,考虑到所使用的是 1989~2006 年横跨 18 年的中国健康与营养调查数据,在这一时期,中国社会经济状况发生了剧烈而深刻的变化。这里将历年持久收入部分(η_{is})的均值作为个体持久收入。

从经验上来判断,女性与男性的可观测收入一年龄曲线应该存在一定差异。为此,这里将全样本按照性别划分为两组,并运用上述步骤计算 $\lambda_s \approx 1$ 的年龄区间及持久收入。经计算男性和女性的历年持久收入部分(η_{is})与其可观测收入(y_{is})在 20~70 岁之间大致呈现以下变动关系(见图 3、图 4)。单纯从图形上无法准确推断 $\lambda_s \approx 1$ 的年龄区间。这里我们对男性个体 λ_s 的曲线与 $\lambda_s=1$ 这两条曲线通过 t 检验进行差异比较(见表 2)。经过多次重复尝试,发现当男性个体年龄大于 34 岁时历年持久收入部分(η_{is})与可观测收入(y_{is})在 0.01 水平上($p=0.007$)不存在统计意义上的差异,即可以采用男性个体大约在 34 岁以后的历年持久收入部分(η_{is})的均值作为其持久收入。

对于女性个体,虽然经过多次重复尝试,但仍旧没能找到一个合理的年龄区间去满足 $\lambda_s \approx 1$ 的条件。这说明女性个体的收入在整个生命周期中起伏变化比较大。也正是由于该原因,本文将在下面的实证研究中考虑样本选择问题。为了处理数据的方便,这里参照男性个体的年龄区间,将女性个体的年龄选择区间也规定为大于 34 岁。

(二) 工作选择偏误与同住偏误分析

借鉴 Cheti Nicoletti(2008)研究代际职业流动的思路,对于代际收入流动,这里将第 θ 个分位数上的分位回归方程写为:

$$y_i^c = a_\theta + b_\theta y_i^p + \varepsilon_{i\theta} \tag{1}$$

其中, $Quant_\theta(\varepsilon_{i\theta} | y_i^p) = 0$, y_i^c 表示考虑了生命周期偏误后子代持久收入的对数; y_i^p 代表父辈相应的持久收入对数; i 代表父辈和子代所在的家庭。

事实上,我们所考虑的工作选择和同住选择都会使样本存在缺失值。当存在

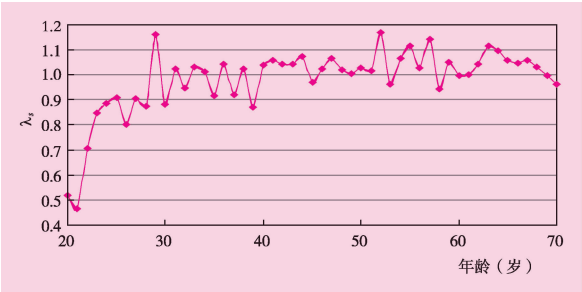


图 3 男性年龄与系数 λ_s 之间的关系变动

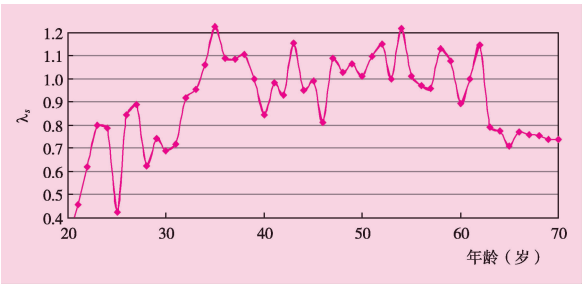


图 4 女性年龄与系数 λ_s 之间的关系变动

表 2 男性个体年龄区间的 t 检验

变量	样本量	均值	标准误差
λ_s	36	1.029	0.010
$\lambda_s=1$	36	1	0
差异	36	0.029	0.010

缺失值时,可以通过虚拟变量将求解式(1)的问题转化为:

$$\frac{1}{n} \sum_i S_i [\theta - I(y_i^c \leq a_\theta + b_\theta y_i^p)] [y_i^p] = 0 \quad (2)$$

其中, S_i 为虚拟变量且通过指示函数 $S_i = I(S_i^* > 0)$ 与潜变量 (S^*) 相联系。该指示函数表示个体的收入记录是否能被观测到及是否观测到父辈与子代至少在一轮调查中共同居住。 S 的具体取值为 0 和 1。在考虑同住选择时,如果观测到父辈与子代至少在一轮调查中共同居住在一起,则 $S=1$, 否则 $S=0$; 而在考虑工作选择时,如果我们可以观测到个体的收入记录,则 $S=1$, 否则 $S=0$ 。其中的潜变量 (S^*) 是个体选择决策方程的函数,即:

$$S^* = \alpha + \beta X_i + v_i \quad (3)$$

其中, X_i 为个体特征向量。一般情况下 y_i^p 是 X_i 子集,但在考察同住选择时 X_i 不包含 y_i^p 。如果 S_i 的取值依赖于某些观测变量,上述最小化问题得出的估计量可能是有偏的。

当 $\varepsilon_{i\theta}$ 与 X_i 相关,即可观测变量存在样本选择时可以采用倾向得分的倒数,即 $1/\hat{P}$ 作为权重对式(2)进行加权估计,则:

$$\frac{1}{n} \sum_i \frac{S_i}{P_i} [\theta - I(y_i^c \leq a_\theta + b_\theta y_i^p)] [y_i^p] = 0 \quad (4)$$

其中, $P_i = \text{pr}(S_i = 1 | X_i) = \text{pr}(v_i > -\alpha - \beta X_i)$ 为倾向得分。

解决可观测变量选择问题的另一种思路是分位回归调整估计。例如,在代际收入流动的简约式方程中子代的受教育程度被遗漏,但该变量既与子代参与工作的意向有关,又会影响子代的收入。此种情况,利用经典的赫克曼两步纠偏法不合适。此时,可将式(1)改写为:

$$y_i^c = a_\theta + b_\theta y_i^p + \delta_\theta X_i + v_{i\theta} \quad (5)$$

其中, X_i 为个体的特征属性变量; δ_θ 为新加入解释变量的系数;其他变量及系数与式(1)相同。如果式(5)满足线性假设,并且对可观测变量不存在样本选择问题,那么, $\text{Quant}_\theta(v_{i\theta} | y_i^p, X_{i\theta}) = 0$ 成立。我们可以通过分位回归调整估计法进行估计。该方法实际是对由于遗漏变量引起的偏误进行矫正。

当 $\varepsilon_{i\theta}$ 依赖于 v_i 时,即不可观测变量存在选择问题。第 θ 个条件分位数可写为:

$$\text{Quant}_\theta(y_i^c | y_i^p, S_i = 1) = a_\theta + b_\theta y_i^p + \text{Quant}_\theta(\varepsilon_{i\theta} | y_i^p, S_i = 1) \quad (6)$$

一般来说 $\text{Quant}_\theta(\varepsilon_{i\theta} | y_i^p, S_i = 1)$ 不等于 0。如果 $(\varepsilon_{i\theta} \perp y_i^p, X_i | P_i, S_i = 1)$ 成立,则有:

$$y_i^c = a_\theta + b_\theta y_i^p + F_\theta(P_i) + \varpi_{i\theta} \quad (7)$$

其中 $\varpi_{i\theta}$ 满足 $\text{Quant}_\theta(\varpi_{i\theta} | X_i, S_i = 1) = 0$, $F_\theta(P_i)$ 为倾向得分的某种形式的函数,主要运用 Heckman 两步法,即运用倾向得分预测值的多项式作为纠正项进入代际收入流动方程对不可观测变量的选择性进行控制。

当样本中同时存在同住选择和工作选择时,应该考虑双变量概率选择模型。这里用 S_{1i} 、 S_{2i} 分别表示样本中是否满足工作选择条件和同住选择条件的两个虚拟变量。当满足这两个条件时, S_{1i} 、 S_{2i} 的取值为 1;其他情况为 0。令 $p_i = p_i(X_i) = \text{pr}(S_{1i} = 1, S_{2i} = 1 | X_i)$ 为同住选择和工作

选择的联合倾向得分。这里仍然可按照前面的分析思路。首先估计双变量概率选择模型;第二步将联合倾向得分的倒数作为权重利用加权最小二乘法估计代际收入流动方程。

四、实证研究结果

应用上述各种模型需要一定的前提条件,因此,在实证研究之前需要对各种假设条件进行检验。表 3 对上述涉及的主要假设进行检验。

在表 3 中,假设 1 中假定全样本与子样本中代际收入弹性无显著性差异,而 Walt 检验拒绝了该假设($p=0.040$),这说明子样本中的同住效应不可忽视。如果使用一般的均值回归可能会产生计量偏误。假设

表 3 估计量假设检验

序号	估计量	假设	Walt 检验	p 值
1	均值回归	$(y^c \perp s y^p)$	4.220	0.040
2	控制函数两步法	$(s \perp y^p y^c, X)$	1.040	0.307
3	控制函数两步法	$(y^c \perp X y^p)$	217.950	0.000
4	倾向得分加权回归	$(s \perp y^p, y^c X)$	0.354	0.562

2 主要考察在全样本中父辈收入与表示同住的虚拟变量 S 是否独立。Walt 检验没有拒绝原假设,即全样本中不存在同住偏误,在实证中可以使用控制函数两步法对全样本中的工作选择偏误进行纠正。假设 3 试图验证子代收入是否与代际收入方程中遗漏的相关变量(比如子女教育、职业等)独立。Walt 检验高度拒绝了原假设。这意味着可以利用回归调整估计法对代际收入弹性进行估计,而原假设中的控制函数两步法将不再适用。最后,Walt 检验假设 4 成立,即 y^p, y^c 在全样本中与表示同住的虚拟变量 S 独立。这时可以利用倾向得分加权回归纠正估计偏误。表 4 给出了全样本的估计结果,表 5 给出了子样本中各种模型估计情况。

从表 4 可以看出,无论是单方程均值回归还是分位回归,在全样本不考虑工作选择偏误的情况下得出的代际收入弹性与表 5 子样本忽略选择偏误时得出的代际收入弹性相比,子样本中的同住选择存在较大偏误。这说明在实证研究中不能忽视同住选择造成的偏误。表 5 此部分中的均值回归及各分位点的系数估计值均大于表 4 此部分的均值回归和分位

表 4 全样本代际收入弹性估计

	忽略工作选择偏误	纠正工作选择偏误		
	单方程回归	控制函数两步法	倾向得分加权回归	回归调整估计
均值回归	0.266*	0.332***	0.308***	0.168**
10 分位	0.036	0.159**	0.005*	0.010
30 分位	0.071**	0.281***	0.209**	0.044**
40 分位	0.089**	0.304***	0.237*	0.058**
50 分位	0.124***	0.313***	0.260**	0.103***
60 分位	0.205***	0.393***	0.317**	0.120***
70 分位	0.300***	0.466***	0.398**	0.117***
90 分位	0.555***	0.557***	0.607**	0.067

注: *、**、*** 分别表示在 10%、5%、1%的水平上显著;分位回归的标准差通过 100 次 bootstrap 计算得到。

回归的估计结果。这一结果说明在本研究中由于忽略同住选择而造成了向上的计量偏误。表 5 给出了在忽略工作选择的情况下,利用控制函数两步法、倾向得分加权回归和回归调整估计等方法对同住选择进行矫正的结果。其中,倾向得分加权回归对同住选择进行纠偏的效果较好,其次是控制函数两步法,而回归调整估计的效果最差。在对工作选择与同住选择偏误同时进行矫正时,倾向得分加权回归的估计结果与忽略各种样本选择时所得结果差异不大,这说明该方法对子样本中的工作选择与同住选择偏误没有起到很好的纠偏作用。而控制函数两步法则与忽略样本选择时所得结果表现出明显差异。这一结果说明在本研究中考虑 4 种计量偏误的必要性。

表 5 子样本代际收入弹性估计

	忽略工作选择 偏误和同住偏误	只考虑同住偏误			同时纠正工作选择偏误和 同住偏误	
	单方程回归	控制函数 两步法	倾向得分 加权回归	回归调整估计	控制函数 两步法	倾向得分 加权回归
均值回归	0.369	0.294***	0.350**	0.314***	0.294***	0.350**
10 分位	0.104**	0.102**	0.119**	0.276***	0.196**	0.145**
30 分位	0.220***	0.192**	0.157***	0.473***	0.332**	0.206***
40 分位	0.229***	0.239***	0.185*	0.478***	0.674***	0.223**
50 分位	0.264***	0.234**	0.189**	0.522***	0.775**	0.287***
60 分位	0.376**	0.428***	0.284***	0.501***	0.773*	0.285**
70 分位	0.526***	0.537***	0.267**	0.439***	0.654**	0.366**
90 分位	0.676**	0.626**	0.218**	0.244***	0.561**	0.303**

注:同表 4。

另外,运用控制函数两步法进行均值回归得到的代际弹性系数为 0.294,且在 1%水平上显著;而采用分位回归得到的各分位点的代际收入弹性变化较大。这说明父辈收入对处在不同收入水平上的子代表现出不同的影响程度。在低分位点和高分位点,代际收入弹性较小;而在 50 和 60 分位点上,代际收入弹性最大,分别为 0.775 和 0.773。从总体上看,农村代际收入弹性随着分位数的上升呈现出先增后减的倒 U 形变化趋势。这一结果与魏颖(2009)的研究类似。在收入分布的两端,由于个体特征的差异导致农村家庭呈现出较强的代际收入流动性;而在中位数附近则呈现出较低的流动性。研究结果说明个体间的特征存在一定差异,并对个体收入有重要影响。

在上述模型中,子代属性变量(如受教育年限、职业、身体健康状况等)虽然在不同分位的收入水平中表现出不同的效应,但这种效应在各个模型之间不存在明显差异。职业变量对于处于高分位收入的子代影响比较大,在 10 分位上回归系数大约为 0.18($p<0.05$);而后随着分位数的提高,职业对子代收入的影响逐渐增大。受教育年限的贡献率大约从 40 分位开始随着子代收入的增加而提高。这说明在较高分位数上,随着受教育年限的提高,子代收入将不断提高。而身体健康状况则在各模型中几乎没有通过显著性检验。需要指出的是农

村居民(特别是富裕家庭、乡村干部)的收入在一定程度上很难准确计量,这可能对估计结果产生一定的影响。

五、结 语

本文将中国健康与营养调查数据分为全样本和子样本,在综合考虑了暂时性收入偏误、生命周期偏误、同住选择和工作选择偏误后对农村家庭的代际收入弹性进行估计。通过比较多种方法的估计结果,发现倾向得分加权估计法对于纠正同住选择引起的偏误具有较好的效果;而对于综合偏误的矫正,控制函数两步法则优于其他方法。从总体上看,农村代际收入弹性随着子代收入分位数的上升呈现出先增后减的“倒U形”变化趋势。这说明在较低分位数部分,父辈收入的变动对子代收入的边际效应较小;在中位数附近,父辈收入对子代收入的边际效应达到最大;随着分位数的提高,这种收入传递效应开始减小。

对于处在低分位数上的农村子女而言,其收入与父辈收入“基本无关”。一般来说,农村家庭对教育型人力资本的需求本来就比较弱,尤其是对那些读书无望的子女,家庭投资就更加消极。在中位数附近,呈现出较高的代际收入弹性,父辈收入对子代收入的影响达到最大,表现出强烈的代际收入传递性。

参考文献:

1. 何晓琦等(2006):《跨代收入流动及其评估》,《统计研究》,第6期。
2. 王海港(2005):《中国居民收入分配的代际流动》,《经济科学》,第2期。
3. 魏颖(2009):《中国代际收入流动与收入不平等问题研究》,中国财政经济出版社。
4. 姚先国等(2007):《中国代际收入流动与传递路径研究:1989~2000》, http://www.lepp.org.cn/soft/images/100296_1.pdf。
5. Anders Bohlmark and Matthew J. Lindquist (2006), Life-Cycle Variations in the Association between Current and Lifetime Income: Replication and Extension for Sweden. *Journal of Labor Economics*. 24(4):879-900.
6. Cheti Nicoletti (2008), Multiple Sample Selection in the Estimation of Intergenerational Occupational Mobility. Institute for Social and Economic Research ISER Working Paper WP2008-20, 1-30.
7. Grawe, N. (2003), Life Cycle Bias in the Estimation of Intergenerational Income. Mobility. Statistics Canada Analytical Studies Branch Working Paper Series, No 207.
8. Haider, S. and Solon, G. (2006), Life Cycle Variations in the Association Between Current and Life Time Earnings. *American Economic Review*. 96(4):1308-1320.
9. Mazumder, B. (2001), Earnings Mobility in the US: A New Look at Intergenerational Mobility. Federal Reserve Bank of Chicago Working Paper NOWP-01-18: 1-38.
10. Sewell, W. and Hauser, R. (1975), *Education, Occupation and Earnings: Achievement in the Early Career*. New York: Academic Press.
11. Solon, G. (1992), Intergenerational Income Mobility in the United States. *The American Economic Review*. 82(3):393-408.
12. Soltow (1965), *Toward income equality in Norway*. University of Wisconsin Press.

(责任编辑:朱 萍)