中国居民代际收入流动性趋势研究[[1]](#footnote-2)

袁青青 刘泽云

摘要：本文运用2002年、2013年和2018年中国家庭收入调查数据，在考虑收入测量偏误和同住样本选择偏差的基础上，测算了我国1970-1995年出生群体的代际收入流动性趋势。研究发现：（1）我国居民代际收入流动性呈下降趋势，代际收入秩回归系数从“70后”的0.397上升至“80后”的0.435，进而上升到“90后”的0.493。（2）代际收入流动性趋势具有性别和户口差异。其中，男性和农业户口的代际收入流动性呈下降趋势，而女性没有显著变化，非农业户口则在“70后”至“80后”有显著下降。（3）不同类型收入的代际流动性趋势略有不同，“90后”工资性收入和主要工作收入的代际流动性与“80后”没有显著差异，但劳动收入和可支配收入均呈下降趋势。（4）同住样本选择偏差对分性别和工资性收入、主要工作收入代际流动性的敏感度较大，但也没有改变我国居民代际收入流动性总体呈下降趋势的结论。本文为提高我国社会流动性和促进共同富裕的公共政策制定提供了有益启示。

关键词：代际收入流动性 代际收入秩回归系数 同住样本选择偏差

中图分类号：F241.2；F063.4 JEL：D31；J18；J62

**Trends in Intergenerational Income Mobility in China**

YUAN Qingqing LIU Zeyun

（Beijing Normal University，Beijing，China）

**Abstract:** Using the data of Chinese Household Income Project survey in 2002, 2013 and 2018, this paper investigates the trends in intergenerational income mobility for cohorts born between 1970 and 1995, based on the correcting in the measurement errors of income and the co-residency bias of samples. The results show that: (1) the intergenerational income mobility shows a downward trend, and the intergenerational income rank regression coefficient rises from 0.397 of the 1970-1979 birth cohort to 0.435 of the 1980-1989 birth cohort, and then to 0.493 of the 1990-1995 birth cohort. (2) The trends of intergenerational income mobility are different by gender and *Hukou*. Among them, the intergenerational income mobility of men and agricultural *Hukou* shows a downward trend, while there is no significant change for women, and the non-agricultural *Hukou* decreases significantly from the birth cohort of 1970-1979 to 1980-1989. (3) The intergenerational mobility trends of different types of income are slightly different. The intergenerational mobility of wage and main work income of the 1990-1995 birth cohort is not significantly different from that of the 1980-1989 birth cohort, but both labor income and disposable income show a downward trend. (4) The co-residency bias of samples is more sensitive to the intergenerational mobility of gender, wage and main work income, but it does not change the overall trends of declining intergenerational income mobility. This study provides useful enlightenment for public policy-making to improve social mobility and achieve common prosperity in China.

**Keywords:** Intergenerational Income Mobility; Intergenerational Income Rank Regression Coefficient; Co-residency Bias

一、引言

代际流动性是衡量机会平等的重要指标，也是度量社会流动性的重要维度，合理的代际流动性有利于社会公平和经济效率提升。目前，已有充分的证据表明，自上世纪80年代以来我国居民收入的横向不平等程度已上升到较高水平（Luo et al，2020），但对于代际收入流动性趋势，也即纵向的收入不平等程度如何变化还没有定论（王学龙、袁易明，2015；Fan et al，2021）。如果一个收入处于较高横向不平等程度的社会，其代际收入流动性也处于较低水平，将容易陷入低流动或不平等陷阱，从而不利于社会稳定和经济可持续发展。同时，我国已开启全面建设社会主义现代化国家的新征程，共同富裕是社会主义的本质要求，是中国式现代化的重要特征，确保人人都有实现向上流动的机会是共同富裕的重要前提。因此，研究我国居民代际收入流动性趋势具有重要现实意义。

本文运用2002年、2013年和2018年中国家庭收入调查数据（Chinese Household Income Project，CHIP），在考虑收入测量偏误和同住样本选择偏差的基础上，测算了我国“70后”、“80后”和“90后”的代际收入流动性趋势，考察了代际收入流动性趋势的性别和户口异质性。同时，比较了不同类型收入的代际流动性及其变化差异，并进一步探讨了同住样本选择偏差问题。本文的研究有助于认识改革开放以来的宏观经济政策和社会环境对我国居民代际收入流动性的系统影响，为提高我国社会流动性和促进共同富裕的公共政策制定提供有益启示。

与现有文献相比，本文的贡献在于：第一，对1990-1995年出生群体的代际收入流动性进行测算，首次把“90后”年轻人群纳入分析范围，给出了我国居民代际收入流动性趋势的最新判断。第二，测算了劳动收入、工资性收入、可支配收入和主要工作收入等不同类型收入的代际流动性，发现不同收入代表性群体的代际流动性差异，也为以往研究得出的不同结论提供了可能的解释。第三，基于CHIP数据中与户主不同住成年子女的收入信息，直观考察了我国居民代际收入流动性测算中的同住样本选择偏差问题，并运用Heckman两步法发现不同样本和收入类型的代际流动性对同住样本选择偏差的敏感度不同，揭示了测量代际收入流动性时考虑同住样本选择偏差的重要性。

二、文献综述

在代际流动性主题研究中，常用指标有收入、教育、职业和主观阶层评价等，本文关注代际收入流动性的原因是：收入的代际流动性更能反映经济福利的代际传递。相比较而言，教育、职业的代际流动性研究容易忽略教育和职业阶层内部的差异，而主观阶层评价则面临可比性的难题。因此，本文关注了代际收入流动性趋势的研究[[2]](#footnote-3)。以下将从我国居民代际收入流动性趋势的相关研究和测算代际收入流动性的方法与计量问题两个方面对现有文献进行评述。

（一）我国居民代际收入流动性趋势的相关研究

关于我国居民代际收入流动性趋势的研究通常有两个视角：一是基于出生队列（杨娟、张绘，2015；王学龙、袁易明，2015；亓寿伟，2016；Fan et al，2021），二是基于数据调查年份（陈琳、袁志刚，2012；何石军、黄桂田，2013；杨沫、王岩，2020）。其中，基于出生队列的研究有助于考察教育体制、劳动力市场和再分配政策等宏观环境的系统影响。从这一视角来看，基于出生队列的研究相比于调查年份的研究更具现实意义，所以在拥有行政大数据和长面板数据的国外研究中，也多从出生队列视角考察代际收入流动性的变化趋势（Lee & Solon，2009；Chetty et al，2014a；Richey & Rosburg，2017；Jácome et al，2021）。基于此，本文也关注出生队列视角的代际收入流动性趋势。

目前，关于我国居民代际收入流动性的出生队列趋势研究数量还比较少，并且对最新的趋势变化缺乏认识。同时，由于数据和方法的差异，已有研究结论并不一致。如王学龙和袁易明（2015）运用CHNS 数据，研究发现我国上世纪60年代至80年代出生群体的代际收入流动性呈先下降后上升的趋势。同样使用CHNS数据，亓寿伟（2016）却发现我国1949-1990年出生群体的代际收入流动性总体呈上升趋势。而使用CHIP1995-2007数据，杨娟和张绘（2015）测算了我国城镇居民的工资性收入的代际流动性变化，发现我国1967-1973、1974-1980和1981-1987年出生群体的代际收入流动性呈上升趋势。总的来看，以上研究均发现我国“70后”至“80后”代际收入流动性呈上升趋势，但Fan et al（2021）运用CFPS数据，研究发现我国1970-1988年出生群体的代际收入流动性呈下降趋势。

基于此，本文运用CHIP数据测算我国居民代际收入流动性趋势，并且不同于以往运用截面数据时仅对部分住户和个别收入的研究。一方面，本文考察了全国居民的代际收入流动性趋势。另一方面，考虑到不同类型收入的代际流动性研究所关注的群体不同，本文比较了不同类型收入的代际流动性及其变化差异。

（二）测算代际收入流动性的方法与计量问题

代际收入流动性可以分为全局性流动和局部性流动指标，也可以分为相对流动和绝对流动指标。其中，以全局性相对流动指标运用最为广泛（Deutscher & Mazumder，2021）[[3]](#footnote-4)。为便于代际收入流动性趋势的直观比较和分析，本文研究的代际收入流动性为相对流动指标。

测算相对代际收入流动性的具体方法有代际收入弹性（intergenerational income elasticity，IGE）、代际收入相关系数（intergenerational income correlation，IGC）和代际收入秩回归系数（intergenerational income rank regression coefficient，IGRC）等[[4]](#footnote-5)。其中，代际收入弹性方法运用最早且最广泛，但是自从Chetty et al（2014b）等运用收入秩（收入排序）方法的研究以来，代际收入秩回归系数方法在近年来的研究中逐渐占据重要位置，尤其被多项研究证实优势突出，并提倡运用（Nybom & Stuhler，2017；Kyzyma & Groh-Samberg，2020）。具体来看，相比于代际收入弹性，这一方法具有以下优点：一是对衰减偏差和生命周期偏误的敏感度较小，在不同样本和方法运用中得出的结论更稳健（Nybom & Stuhler，2017；Kyzyma & Groh-Samberg，2020）；二是收入秩相比于收入水平在人的生命周期中更稳定，考察的代际收入的线性关系更明显，同时能减小收入极端值的影响（Dahl & DeLeire，2008；Chetty et al，2014b；Nybom & Stuhler，2017）。因此，本文以代际收入秩回归系数作为估计代际收入流动性的基准方法。

目前，这一方法在我国代际收入流动性的新近研究中得到运用，如王伟同等（2019）运用CHIP数据对收入秩方法和弹性方法的比较研究，杨沫和王岩（2020）运用CHNS数据对代际收入流动性的趋势和机制分析，刘琳和赵建梅（2020）、郑筱婷等（2020）运用CFPS数据分别对社会网络、城市产业的就业扩张影响代际收入流动性的研究。而关于代际收入流动性的出生队列趋势研究，只有Fan et al（2021）运用该方法对CFPS数据进行的测算。本文的研究则丰富了基于截面数据运用收入秩方法测算代际收入流动性趋势的研究。

在测算代际收入流动性时，由于使用短面板或截面数据，往往需要处理收入测量偏误和样本选择偏差两个问题。欧美等发达国家因为具有长面板数据和税收系统数据，能够较好克服以上计量偏误（Chetty et al，2014b；Nybom & Stuhler，2017）。但在我国的研究中，虽然收入测量偏误受到重视并得到一定程度纠正，但样本选择偏差得以纠正的研究很少。考察发现只有杨娟和张绘（2015）、Fan et al（2021）使用了Heckman两步法纠正这一问题，朱诗娥和唐瑾（2018）则使用包含不同住样本收入的数据进行研究。比较来看，Heckman两步法的优势在于对样本的要求相对较低，即没有要求不同住样本的收入信息，但需要具有是否同住的排他性约束变量；而后者则需要不同住样本的收入信息。在CHIP调查中，由于只有CHIP2013城镇数据和CHIP2018数据有不同住样本的收入信息，所以本文运用Heckman两步法纠正样本选择偏差。但在进一步分析中，也运用CHIP2018数据具有收入信息的不同住样本，再次考察了同住样本选择偏差对研究的影响。

三、数据与方法

（一）数据说明

1.CHIP数据

本文数据来源于中国家庭收入调查，该调查由北京师范大学中国收入分配研究院等单位组织，已连续于1988、1995、2002、2007、2013和2018年开展了6次全国住户调查，调查内容包括住户成员的个人特征、工作和收入等信息，以及住户家庭的基本特征、收支和资产等信息。其中，2002、2007和2013年除了城镇住户和农村住户调查外，还包括单独的流动人口调查。同时，2002年的城镇住户、2007年的部分农村住户、2013和2018年的城乡住户调查均可以获得两套收支数据，即一个来自CHIP问卷调查，一个来自国家统计局住户收支调查，且后者以日记账形式记录，这也是CHIP数据不同于国内其他数据库的一个重要方面，收入的可信度更高。另外，2007、2013（城镇）和2018年的CHIP数据调查了与户主不同住成年子女的基本信息，包括年龄、教育和工作等。其中，2013和2018年的调查中还询问了与户主不同住成年子女的主要工作收入信息，这也是CHIP数据与国内其他数据库的不同之处。

本文的实证分析基于CHIP2002、2013和2018三个调查年份的混合截面数据。其中，没有使用CHIP1988和2007数据的原因是收入口径无法统一，而没有使用CHIP1995数据的原因是1970年及之后出生的样本量较少。

2.样本

本文的样本定义如下：

第一，通过匹配同一家庭父母与子女的信息，获得户主及其配偶作为父代，其子女作为子代的样本，以及户主及其配偶作为子代，其父母作为父代的样本。首先，对于子代，虽然部分研究只关注儿子（Gregg et al，2017），但也有研究指出，女性劳动参与率在提高，并且婚姻在代际传递中具有放大作用，应该将女性样本也包括进去（Olivetti & Paserman，2015），因此本文对儿子和女儿均进行分析。其次，对于一个家庭有多个子女的情况，一些研究做重复观测值处理（韩军辉、龙志和，2011；杨沫、王岩，2020），一些研究仅保留独生子女或仅保留年长子女（Solon，1992；何石军、黄桂田，2013），也有研究计算同一个家庭多个子女收入排名的均值（Abramitzky et al，2021）。本文的基准分析以重复观测值处理，而对于只保留年长子女和计算多个子女收入排名均值的情况进行稳健性检验，但对于只分析独生子女的研究，本文认为会存在较大偏误，故不做此处理。最后，对于父代，通常选择父亲收入作为父代收入（Solon，1992；杨沫、王岩，2020），或者父母平均（加总）收入（Chetty et al，2014a；Fan et al，2021）等。考虑到父亲和母亲在家庭中承担的责任不同，本文使用了父母平均收入来衡量父代收入。

第二，将子女出生年份限定在1970-1995年，调查时的年龄为23-35岁，并剔除父母与子女年龄差距小于16岁的样本，以及父母年龄超过60岁的样本。首先，子女年龄范围的限定考虑了两个方面：一是对于删除23岁以下样本的原因是个人即使接受大学本科教育，多数也在23岁时离开了学校，所以避免了选择更低年龄时造成对接受更高教育者的样本遗漏。根据2015年的全国1%人口抽样调查数据，1980-1992年出生（23-35岁）人口仍在校的比例为2.02%，而根据CHIP2018数据，1983-1995年出生（23-35岁）人口仍在校的比例为2.25%。而一些研究包括了20岁甚至16岁的样本，显然过于年轻。二是对于删除35岁以上子女，原因在于限定更小的年龄区间可使样本更具可比性，而且超过35岁的样本较少。另外，考虑到部分研究以更小的年龄区间（25-35岁）进行了分析（王伟同等，2019；杨沫、王岩，2020；Fan et al，2022），本文也在稳健性检验部分将子女年龄区间进一步缩小。其次，父母年龄选择删除60岁以上样本，虽然有研究为增加样本量等原因将父母年龄上限提高到64岁（Fan et al，2021）或者65岁（杨沫、王岩，2020），但本文认为将60岁以上父母包括进来时，会忽略掉因退休而不在劳动力市场中的部分群体，进而可能导致更大的估计偏误，因此设定了较小的父母年龄区间。

第三，限定子女和父母均在调查年份有收入和教育信息的样本，即不包括收入为0和教育年限异常的样本，同时剔除子女和父母在上学、已退休或处于失业/待业等其他不工作情况的样本。最终获得有效观测值5743个，其中，户主及其配偶作为子代的样本有180个，占比为3.13%。

（二）变量描述

1.变量定义

本文用到的核心变量为劳动收入，定义为个人年工资性收入和经营净收入的总和。具体来看，在CHIP2002数据中，城镇住户收入数据直接使用国家统计局提供的工资性收入和经营净收入指标，农村住户收入数据则由CHIP问卷调查的工资性收入加上按工作时间分摊到个人的家庭非农经营净收入和农业经营净收入。其中，关于个人农业经营净收入的获取，本文借鉴了汪小芹和邵宜航（2021）在运用CHIP1995-2013数据研究代际收入流动性时的处理办法[[5]](#footnote-6)。具体来看，个人农业经营净收入=（家庭成员在家庭农业经营中的劳动时间÷家庭农业经营的总时间）×家庭农业经营净收入（家庭农业经营净收入=农业经营收入-农业经营费用）。而CHIP2013和2018的城乡住户收入数据均来自国家统计局。其中，农村住户家庭农业经营净收入也根据劳动时间分摊到个人，而对于家庭非农经营净收入在调查时已记录到个人，无需再分摊处理。最终，本文获得用于研究1970-1995年出生群体口径一致的收入指标，作为考察代际收入流动性趋势的基准分析变量。

在进一步分析部分，为比较不同类型收入的代际流动性及其变化差异，本文还用到的收入变量有工资性收入、可支配收入和主要工作收入。变量定义和说明如下：（1）工资性收入，在CHIP2002、2013和2018数据中均可以直接获得这一指标，主要反映个人工作能力（Deutscher & Mazumder，2020）。但不同于劳动收入，测算工资性收入的代际流动性时，考察的群体将不包括只从事经营性工作（即非农经营性工作和农业经营性工作）的个人，所以预计得到的代际收入流动性相对要高。（2）可支配收入，定义为工资性收入、经营净收入、财产净收入和转移净收入的总和，反映总的经济福利。不同于前两类收入，预计可支配收入的代际流动性要低于工资性收入，但与劳动收入的大小不唯一。在CHIP2013和2018数据中，可支配收入的口径是一致的，并且可以直接获得这一指标，但CHIP2002数据没有与之口径统一的指标（因CHIP2002的财产性收入数据没有包括自有住房折算租金）。另外，需要说明的是，由于税收转移不会改变个人收入在人群中的位置，所以在采用收入秩的方法研究中，对可支配收入的考察可以等同于对总收入的考察（Chetty et al，2014b）。（3）主要工作收入，指个人年工资性收入或经营净收入等总额，在CHIP2013和2018数据中可由问卷直接获得。由于在分析不同住子女样本时用到了此项收入指标，所以为便于比较，本文将其考虑进来。以上工资性收入指标考察的群体范围相对较小，可支配收入和主要工作收入指标则不能考察“70后”群体，所以本文没有将这三类指标加入基准分析，而是在进一步的分析部分探讨其代际流动性及变化差异。

为剔除通货膨胀的影响，本文利用各省CPI数据将2002年和2013年的所有个人收入均平减至2018年。在本文中，收入变量最终以收入秩（收入百分位排序）衡量，并且在不同队列和组别的分析中，收入均按百分位进行重新排序。

除此之外，本文还用到的变量有性别、兄弟姐妹数量、婚姻状况、户口类型、住户类型、地区等。其中，性别设定男性为1，女性为0。婚姻状况设定已婚为1，未婚为0。户口指个人出生时的户口，即还原为个人户口转换之前的状态，设定非农业户口为1，农业户口为0。当前住户类型分为两类，城镇住户为1，农村住户为0。

2.描述统计

本文分析的总样本和各出生队列的描述统计结果见表1。首先，从总样本来看，子女（样本）的年收入（劳动收入）均值为15107元，年龄均值为26岁，受教育年限均值为10.4年，大致达到高一年级水平，男性比例73%[[6]](#footnote-7)，非农业户口比例12%，城镇住户比例24%，东部地区样本比例36%；父母的年收入均值为16376元，年龄均值为52岁，受教育年限均值为6.8年，基本达到初一年级水平。其次，从不同出生队列来看，“70后”和“90后”的样本相对年轻，样本的年收入、受教育年限、非农业户口比例、城镇住户比例均在不断增加，而兄弟姐妹数量和东部地区比例则在减少。其中，受教育年限均值从“70后”的9.2年提高到“90后”的11.8年，城镇住户比例从“70后”的13%提高到“90”后的41%。另外，“70后”的已婚样本占比47%。最后，从样本分布来看，“70后”样本均来自CHIP2002数据，而“80后”和“90后”的样本来自CHIP2013和2018两个调查年份的数据。

表1 变量描述统计（均值）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 变量 | 总样本 | 70后 | 80后 | 90后 |
| 子女特征 |  |  |  |  |
| 收入（元/年） | 15107 | 6607 | 19327 | 22304 |
| 年龄（岁） | 26.37 | 25.79 | 28.03 | 24.96 |
| 受教育年限（年） | 10.44 | 9.18 | 10.83 | 11.85 |
| 男性 | 0.73 | 0.71 | 0.76 | 0.69 |
| 非农业户口 | 0.12 | 0.11 | 0.12 | 0.15 |
| 城镇住户 | 0.24 | 0.13 | 0.25 | 0.41 |
| 东部地区 | 0.36 | 0.37 | 0.36 | 0.35 |
| 兄弟姐妹数量（包括自己） | 1.97 | — | 2.00 | 1.92 |
| 已婚 | 0.46 | 0.46 | — | — |
| 父母特征 |  |  |  |  |
| 收入（元/年） | 16376 | 6031 | 19127 | 28424 |
| 年龄（岁） | 51.73 | 52.06 | 52.59 | 50.03 |
| 受教育年限（年） | 6.77 | 5.92 | 7.16 | 7.55 |
| 观测值 | 5743 | 2249 | 2025 | 1469 |
| CHIP2002 | 2249 | 2249 | 0 | 0 |
| CHIP2013 | 1731 | 0 | 1445 | 286 |
| CHIP2018 | 1763 | 0 | 580 | 1183 |

注：表中数据为各变量均值，收入为劳动收入，已通过CPI调整为2018年的可比数据；“70后”、“80后”、“90后”分别指出生于1970-1979年、1980-1989年、1990-1995年的群体；子女年龄23-35岁，父母年龄41-60岁。

（三）研究方法

1.代际收入秩回归系数

本文运用代际收入秩回归系数测算代际收入流动性。具体模型设定如下：

 $Y\_{ci}=β\_{0}+β\_{1}Y\_{pi}+β\_{2}X\_{ci}+ε\_{i}$ （1）

其中，$Y\_{ci}$和$Y\_{pi}$分别代表子女和父母在各自队列的收入百分位排序（简称“收入秩”）。$β\_{1}$为代际收入传递系数（介于0-1之间），表示相对于父母收入分配地位的一个百分位变化，子女收入分配的百分位变化大小，衡量了子女和父母收入在各自收入分配地位的相关性（Chetty et al，2014b；Nybom & Stuhler，2017）。（$1-β\_{1}$）则反映了代际收入流动性。$X\_{ci}$为控制变量，包括父母和子女的年龄及年龄平方项。$ε\_{i}$为随机误差项。

需要说明的是，代际收入流动性研究的关注重点并不是因果关系。虽然Bowles & Gintis（2002）、Blanden et al（2007）等研究试图处理代际收入流动性研究中的内生性问题，即控制认知能力等变量，但本文同Richey & Rosburg（2017）等众多研究一样，将重点关注代际收入流动性的变化。

2.计量问题处理

（1）收入测量偏误

收入测量偏误包括两个方面。一是生命周期偏误（life-cycle bias），指收入在个人年龄较小或年龄较大时获得，会导致代际收入流动性的高估。对于生命周期偏误，本文的处理办法同Fan et al（2021）类似，做了以下三方面的处理：第一，在代际收入流动性测算时加入子女和父母的年龄及年龄平方项；第二，限制父母与子女取得收入时的年龄边界，即子女年龄足够大而不在上学，父母年龄足够小还未退休；第三，使用收入秩的方法，即运用子女和父母在各自队列中的收入排序进行估计，原因在于同龄人的相对收入排序在不同生命周期中变化相对较小（Nybom & Stuhler，2017），因此利用这一指标来代替持久收入排序。

二是衰减偏误（attenuation bias），指由于仅使用一年的调查数据可能出现短暂性的收入冲击问题，或者收入调查时面临的报告偏误问题，会导致代际收入流动性的高估（Solon，1992；Mazumder，2005）。对于衰减偏误，追踪数据可以利用多年收入均值来减小偏误，本文使用截面数据不能选择此方法，但基于收入秩的方法同样可以减小此类偏误的影响（Nybom & Stuhler，2017）。

（2）样本选择偏差

样本选择偏差通常包括两类。一是因流动人口缺失导致的偏误，即研究设计时忽略了流动人口，进而可能导致全国层面代际收入流动性的低估[[7]](#footnote-8)。本文使用CHIP数据包含城镇住户、农村住户和流动人口所有样本，所以不会面临这一问题。

二是同住样本选择偏差（co-residency bias），也称样本截断问题（sample truncation problem），即由于以家庭为单位的微观调查数据通常只有同一住户的父母和子女收入信息，使代际收入流动性研究无法考虑与父母不同住的子女，进而导致同住样本选择偏差问题（Emran et al，2018）[[8]](#footnote-9)。本文通过Heckman两步法纠正这一问题。具体步骤如下：

首先，估计选择方程，即子女是否与户主同住的Probit模型，并计算逆米尔斯比率（inverse mills ratio，IMR）。其次，将得到的逆米尔斯比率作为控制变量加入代际收入流动性测算的方程中。选择方程设定如下：

 $P\_{i}=π\_{0}+π\_{1}Z\_{i}+π\_{2}C\_{i}+v\_{i}$ （2）

其中，$P\_{i}$表示是否为户主的同住子女样本的虚拟变量，1是与户主同住的子女，0是不同住子女。$Z\_{i}$为子代的兄弟姐妹数量（CHIP2013和2018样本）或已婚变量（CHIP2002样本），作为方程的排他性约束变量；$C\_{i}$包括一系列控制变量，有子女性别、教育年限、子女和父母的年龄及年龄平方项、城乡住户类型、地区虚拟变量。

在上述选择方程中，运用两个外生识别条件的原因是：CHIP2013和2018数据的不同住样本没有婚姻状况信息，而CHIP2002数据没有兄弟姐妹数量信息，因此在估计同住样本的逆米尔斯比率时，前者使用兄弟姐妹数量做外生工具，后者使用婚姻状况做外生工具。同时，由于CHIP2002数据没有调查与户主不同住子女的信息，而CHIP2007数据与其相隔年份较近，且调查了与户主不同住子女的信息，因此借助CHIP2007数据来计算CHIP2002同住样本的逆米尔斯比率。在这里，选择兄弟姐妹数量做排他性约束变量的原因是：兄弟姐妹越多，个人要照顾父母的可能性越低，进而与父母同住的概率越低（Fan et al，2021）。而是否结婚也与子女是否与父母同住高度相关，通常来讲，已婚子女与父母同住的概率更低，因此也可以作为是否同住子女的排他性约束变量。

还需说明的是，虽然也有研究通过两样本两阶段最小二乘法（TS2SLS）来解决收入测量偏误和样本选择偏差问题，或者使用父母教育作为父母收入的工具变量，但基于工具变量的方法往往严重低估代际收入流动性（Solon，1992），同时工具变量的外生性也不满足，因此本文未考虑这一方法。

四、实证分析结果

（一）基准回归

表2为代际收入流动性的Heckman两步法估计结果和OLS方法估计结果。其中，第（1）至（3）列分别为“70后”、“80后”和“90后”的代际收入秩回归系数，第（4）和（5）列分别为“80后”与“70后”、“90后”与“80后”的代际收入回归系数差异，其中，组间系数差异检验使用了费舍尔组合检验方法。

表2显示：首先，从基于Heckman两步法估计的基准结果来看，我国“70后”的居民平均代际收入秩回归系数为0.397，“80后”为0.435，“90后”为0.493，且“80后”与“70后”、“90后”与“80后”的代际收入秩回归系数分别在10%和5%的水平上具有显著差异，表明我国居民代际收入流动性呈下降趋势。这一发现与Fan et al（2021）对1970-1988年出生群体代际收入流动性变化的研究发现一致。本文的发现一方面支持了关于我国居民代际收入流动性呈下降趋势的观点，另一方面将研究扩展到“90后”，为最新的代际收入流动性趋势判断提供了经验证据。其次，从未纠正同住样本选择偏差的OLS估计结果来看，代际收入秩回归系数比基准回归结果略高，但两者差异并不明显。同时，代际收入流动性也呈现显著下降的趋势。可见，虽然“70后”和“80后”基准回归的逆米尔斯比率系数显著，但总的来看，同住样本选择偏差在这里并未影响基准回归的结论。

表2 代际收入流动性趋势：基准回归结果

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | （1） | （2） | （3） | （4） | （5） |
| 70后 | 80后 | 90后 | 80后-70后 | 90后-80后 |
| 1.Heckman两步法 | 0.397\*\*\*(0.020) | 0.435\*\*\*(0.021) | 0.493\*\*\*(0.023) | 0.038\*[0.080] | 0.058\*\*[0.047] |
| 逆米尔斯比率 | -12.169\*\*\*(2.449) | 6.253\*\*(2.553) | 1.192(3.355) |  |  |
| 观测值 | 2249 | 2025 | 1469 |  |  |
| R2 | 0.168 | 0.201 | 0.250 |  |  |
| 2.OLS方法 | 0.399\*\*\*(0.020) | 0.446\*\*\*(0.020) | 0.494\*\*\*(0.022) | 0.047\*[0.063] | 0.048\*[0.053] |
| R2 | 0.159 | 0.199 | 0.250 |  |  |

注：\*\*\*、\*\*和\*分别表示在1%、5%和10%水平上显著；圆括号内为稳健标准误；方括号内为费舍尔组合检验的P值，重复抽样1000次。“70后”、“80后”、“90后”分别指出生于1970-1979年、1980-1989年、1990-1995年的群体。使用数据为CHIP2002、2013和2018，样本年龄为23-35岁；各回归方程的控制变量还有子女和父母的年龄及年龄平方项。

（二）稳健性检验

本文从3个方面进行稳健性检验：一是运用代际收入弹性方法，即替换收入秩为对数收入指标；二是仍然使用代际收入秩回归系数，但调整了样本；三是在代际收入秩回归时运用了滚动法估计。表3所示为前两种检验方法的估计结果，图1则为运用滚动法估计检验的结果。

1.代际收入弹性方法

代际收入弹性方法是测算代际收入流动性运用最多的方法，但与代际收入秩回归系数不同，代际收入弹性系数的变化还包含了代际收入不平等的变化，即从数量关系来看，代际收入弹性系数等于代际收入相关系数乘以子女收入与父母收入标准差的比值。因此，代际收入秩回归系数与代际收入弹性系数也有本质区别，即不同时期代际收入弹性系数的差异一方面来自代际收入秩回归系数的差异，另一方面也来自两代人各自收入差距的相对大小。这意味着代际收入秩回归系数的降低反映了阶层流动增加，但代际收入弹性系数的缩小并不一定来自阶层流动性的增加，也可能源自两者收入差距缩小。因此，本文也估计了代际收入弹性系数，以更全面反映代际收入流动性趋势。

从表3代际收入弹性方法估计结果来看，我国“70后”、“80后”和“90后”的代际收入弹性系数分别为0.502、0.605和0.685。其中，“80后”的系数显著高于70后，而“90后”的系数与“80后”在统计上没有显著差异。但总的来看，运用代际收入弹性方法测算的我国居民代际收入流动性仍然呈下降趋势。在这里，代际收入弹性系数要高于代际收入秩回归系数，表明子女收入的不平等程度要大于父母。同时，代际收入流动性变化的原因可能是收入不平等在加剧，或者是父母与子女之间其他差异的影响（Kyzyma & Groh-Samberg，2020）。另外，与Fan et al（2021）的结果相比，本文估计的代际收入弹性要更高，这可能源于两者数据来源及处理等方面的差异。

2.调整样本

本文关于样本的调整主要考虑了三个方面：一是子女和父母年龄范围的重新选择；二是对多子女样本的处理；三是仅使用户主子女为子代的样本。

首先，考虑到“90后”样本的年龄都在23-28岁之间，而“70后”和“80后”的年龄范围分别为23-32岁和24-35岁，为避免各出生队列收入年龄不对称带来的不可比性，本文将各出生队列的年龄都限定在23-28岁。表3显示，“80后”的代际收入秩回归系数仍显著高于“70后”，而“90后”和“80后”没有显著差异，但总的来看，代际收入流动性趋势下降的结论仍然稳健。其次，考虑到年龄限制过小可能导致样本代表性不足，因此，本文进一步将子女年龄限定在27-35岁，同样发现“80后”与“70后”之间有显著差异，而“90后”与“80后”之间没有显著差异，但总的结论仍然稳健。在这里，选择27岁的原因是能够确保“90后”有足够多的样本量进行研究，同时，这一年龄对于进入劳动力市场的群体来说，多数超过5年的工作经验，收入也更具代表性。最后，本文也将样本年龄限定在27-32岁，即更小的年龄范围使同一出生队列的群体更具可比性，同时，较低的年龄对于同住样本选择偏差也相对较小（Fan et al，2022）。从回归结果来看，基准结论也是稳健的。当然，我们也依次去除子女年龄为24、25和26岁的样本，得到基本稳健的结论[[9]](#footnote-10)。另外，关于父母年龄范围的调整，进一步限制在41-59岁、41-58岁、…、41-55岁时，同样得到了稳健结论（限于篇幅，仅列出父母为41-55岁的估计结果）[[10]](#footnote-11)。

对于多子女样本，本文首先将每个家庭保留一个年长子女，回归结果表明结论相当稳健。其次，本文将家庭多个子女收入排序的均值作为对应父母收入排序的子女收入排序，也得到稳健的估计结论。而对于仅使用户主子女作为子代的样本，同样估计得到了稳健结论。

表3 代际收入流动性趋势：替换对数收入和调整样本

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | （1） | （2） | （3） | （4） | （5） |
| 70后 | 80后 | 90后 | 80后-70后 | 90后-80后 |
| A.代际收入弹性方法 |  |  |  |  |  |
| 1.对数收入 | 0.502\*\*\*(0.032) | 0.605\*\*\*(0.046) | 0.685\*\*\*(0.056) | 0.103\*\*[0.010] | 0.080[0.131] |
| B.调整样本 |  |  |  |  |  |
| 2.子女23-28岁 | 0.416\*\*\*(0.021) | 0.480\*\*\*(0.027) | 0.483\*\*\*(0.024) | 0.064\*\*[0.035] | 0.003[0.469] |
| 观测值 | 1887 | 1077 | 1469 |  |  |
| 3.子女27-35岁 | 0.291\*\*\*(0.037) | 0.404\*\*\*(0.027) | 0.460\*\*\*(0.047) | 0.113\*\*[0.011] | 0.056[0.153] |
| 观测值 | 748 | 1247 | 366 |  |  |
| 4.子女27-32岁 | 0.291\*\*\*(0.037) | 0.399\*\*\*(0.030) | 0.459\*\*\*(0.047) | 0.109\*\*\*[0.006] | 0.060[0.149] |
| 观测值 | 748 | 1090 | 366 |  |  |
| 5.父母41-55岁 | 0.399\*\*\*(0.023) | 0.441\*\*\*(0.025) | 0.484\*\*\*(0.025) | 0.042\*[0.095] | 0.043[0.119] |
| 观测值 | 1605 | 1304 | 1291 |  |  |
| 6.保留一个年长子女 | 0.384\*\*\*(0.022) | 0.437\*\*\*(0.022) | 0.498\*\*\*(0.025) | 0.053\*[0.053] | 0.061\*\*[0.030] |
| 观测值 | 1890 | 1845 | 1274 |  |  |
| 7.多子女收入秩平均 | 0.394\*\*\*(0.022) | 0.439\*\*\*(0.022) | 0.497\*\*\*(0.025) | 0.046\*[0.059] | 0.058\*\*[0.043] |
| 观测值 | 1890 | 1845 | 1274 |  |  |
| 8.户主子女为子代 | 0.399\*\*\*(0.020) | 0.441\*\*\*(0.021) | 0.503\*\*\*(0.023) | 0.042\*[0.077] | 0.062\*\*[0.030] |
| 观测值 | 2192 | 1936 | 1435 |  |  |

注：使用数据为CHIP2002、2013和2018，样本年龄为23-35岁，采用Heckman两步法估计；各回归方程的控制变量还有子女和父母的年龄及年龄平方项、逆米尔斯比率。其他说明同表2。

3.滚动法估计

关于代际收入流动性趋势研究，运用滚动法可以考察较为准确的变化过程。通常来讲，逐年回归对变化过程的考察最为细致，但如果相近年份的代际收入流动性变化较小，也可以将多个年份进行平均[[11]](#footnote-12)。因此，为准确把握代际收入流动性的长期趋势，本文进一步采取了滚动法进行稳健性检验。图1给出了时间窗口为1-10年的估计结果，为便于观察，分两个图（a和b）进行展示。

可以发现，时间窗口较长的估计结果要更加平缓，而时间窗口较短的估计结果波动较大。需要强调的是，由于出生于1970、1971和1980-1982各年的样本量均不足100，所以导致在较短时间窗口（如小于6年）的估计时，1970和1981年附近估计结果的可信度不高，这也是图1中基于较长时间窗口期和较短时间窗口期样本估计结果差异较大的原因。同时，我们也不能根据5年及以下窗口期的估计结果直接得出代际收入流动性在1980年前后发生转折的结论。但这一问题并不影响较长窗口期的估计结果，并且可以确定的是，从1970-1995年，我国居民代际收入传递系数总体呈上升趋势，即代际收入流动性的整体趋势是下降的。这一发现与亓寿伟（2016）对我国1949-1990年出生队列的研究结论不同，我们认为可能的原因是两者在方法选择和数据处理等方面的差异。



（a）时间窗口为1-5年 （b）时间窗口为6-10年

图1 代际收入流动性趋势：滚动法估计

注：横轴为样本出生年份。“10年”线表示以十年为时间窗口进行滚动回归的结果，其中，1970年对应1970-1979年出生样本的估计结果，依次类推，1986年对应1986-1995年出生样本的估计结果，而1986年以后年份对应的时间窗口数逐年减1；其他线条的含义同上。使用的数据为CHIP2002、2013和2018，样本年龄为23-35岁，采用Heckman两步法估计。

综上，根据基准回归结果和稳健性检验结果，可以发现“70后”、“80后”和“90后”的代际收入流动性逐渐下降的结论是比较稳健的（尽管在表3的部分稳健性回归中，“90后”较“80后”代际收入流动性的差异不显著，但从数值上看，“90后”的代际收入秩回归系数明显大于“80后”的代际收入秩回归系数）。也就是说，总体来看，我国1970-1995年出生群体的代际收入流动性呈下降趋势。

（三）异质性分析

在基准分析部分，本文关注全国层面的代际收入流动性变化，以分析全国的机会平等情况，但也有必要对不同群体内部的代际收入流动性进行考察，以研究特定群体内部的机会平等状况。以下从性别和户口两个方面探讨了代际收入流动性变化的异质性问题。

1.性别差异

表4为代际收入流动性趋势的性别差异估计结果。首先，从男性样本的估计结果来看，“70后”、“80后”和“90后”的代际收入秩回归系数分别为0.358、0.375和0.438，并且“90后”系数较“80后”有显著增加，但“80后”与“70后”没有显著差异。尽管如此，男性的代际收入流动性总体呈下降趋势，即机会平等程度在下降。而对于女性样本，“70后”、“80后”和“90后”的代际收入秩回归系数分别为0.472、0.404和0.438，虽然“80后”的代际收入秩回归系数较“70后”和“90后”均比较小，但这些差异在统计上并不显著，因此，可以认为女性的代际收入流动性没有发生显著变化。对于以上关于代际收入流动性趋势的异质性发现，本文认为可能的原因是：计划生育政策等宏观环境变化对男性与女性的机会平等产生了不同影响。以教育机会平等为例，一方面，从子女教育获得与父母收入的相关性变化来看，女性的数值下降幅度要大于男性，意味着两者的机会平等情况在出生队列变化中可能具有不同特征。另一方面，生育政策会导致女性相对男性教育获得的差异缩小，即由于家庭子女数量减少，家庭对于女孩的教育等投资增加，促使女性的机会平等程度大幅提升，进而使女性表现出与男性群体不同的变化特征。

其次，从逆米尔斯比率系数的显著性来看，分性别估计如果不纠正样本选择偏差将导致严重的估计偏误。以男性样本为例，用OLS估计“80后”和“90后”的代际收入秩回归系数分别为0.429和0.489，虽然趋势没变，但严重低估了男性代际收入流动性。同理，各出生队列女性的代际收入秩回归系数也存在高估的情况。因此有必要纠正同住样本选择偏差问题。

最后，从男女代际收入秩回归系数的比较来看，“70后”和“80后”男性的代际收入流动性略高于女性，这一发现也与Fan et al（2021）类似，但“90后”男性和女性代际收入流动性没有明显差异。前期，从婚姻匹配和劳动力供给反应两个机制可以解释男性和女性代际收入流动性的差异（Raaum et al，2007），但随着女性与男性受教育水平差距缩小，女性劳动力市场参与率提高，以及生育政策对“重男轻女”思想的冲击等原因，可能使女性与男性在机会平等中的差异不断缩小。

表4 代际收入流动性趋势：性别差异

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | （1） | （2） | （3） | （4） | （5） |
| 70后 | 80后 | 90后 | 80后-70后 | 90后-80后 |
| 1.男性 | 0.358\*\*\*(0.024) | 0.375\*\*\*(0.025) | 0.438\*\*\*(0.032) | 0.017[0.322] | 0.063\*\*[0.049] |
| 逆米尔斯比率 | -4.845(5.140) | 55.591\*\*\*(7.449) | 47.061\*\*\*(12.303) |  |  |
| 观测值 | 1604 | 1549 | 1016 |  |  |
| R2 | 0.129 | 0.218 | 0.260 |  |  |
| 2.女性 | 0.472\*\*\*(0.035) | 0.404\*\*\*(0.044) | 0.438\*\*\*(0.048) | -0.068[0.112] | 0.034[0.287] |
| 逆米尔斯比率 | -16.924\*\*\*(4.413) | 51.723\*\*\*(8.304) | 41.609\*\*\*(10.846) |  |  |
| 观测值 | 645 | 476 | 453 |  |  |
| R2 | 0.273 | 0.307 | 0.297 |  |  |

注：使用数据为CHIP2002、2013和2018，样本年龄为23-35岁，采用Heckman两步法估计；各回归方程的控制变量还有子女和父母的年龄及年龄平方项。其他说明同表2。

2.户口差异[[12]](#footnote-13)

表5为代际收入流动性趋势的户口差异估计结果。首先，从非农业户口的估计结果来看，“70后”、“80后”和“90后”的代际收入秩回归系数分别为0.354、0.472和0. 385，并且“80后”系数较“70后”有显著增加，表明代际收入流动性有所下降，这一发现与Fan et al（2021）的研究类似。但“90后”与“80后”没有显著差异，而从系数大小来看，可能出现流动性逆转上升的情况，即非农业户口群体内部的机会平等程度增加。而对于农业户口，“70后”、“80后”和“90后”的代际收入秩回归系数分别为0.268、0.398和0. 464，并且“80后”与“70后”、“90后”与“80后”的系数至少在5%的水平上有显著差异，表明代际收入流动性呈下降趋势，即农业户口群体内部的机会平等程度在下降。

其次，从逆米尔斯比率系数的显著性和大小来看，同住样本选择偏差对于分户口估计的敏感度较小。

最后，从非农业户口和农业户口代际收入秩回归系数的比较来看，“70后”和“80后”的农业户口群体要低于非农业户口群体，这一发现不同于Fan et al（2021）的研究，但“90后”农业户口群体的代际收入流动性要小于非农业户口群体，即出现逆转现象，表明非农业户口群体的机会平等程度高于农业户口群体。

表5 代际收入流动性趋势：户口差异

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | （1） | （2） | （3） | （4） | （5） |
| 70后 | 80后 | 90后 | 80后-70后 | 90后-80后 |
| 1.非农业户口 | 0.354\*\*\*(0.062) | 0.472\*\*\*(0.061) | 0.385\*\*\*(0.066) | 0.118\*[0.081] | -0.087[0.149] |
| 逆米尔斯比率 | -16.303\*(8.912) | -1.721(6.787) | -6.689(8.839) |  |  |
| 观测值 | 253 | 241 | 216 |  |  |
| R2 | 0.165 | 0.240 | 0.174 |  |  |
| 2.农业户口 | 0.268\*\*\*(0.023) | 0.398\*\*\*(0.022) | 0.464\*\*\*(0.026) | 0.130\*\*\*[0.000] | 0.066\*\*[0.039] |
| 逆米尔斯比率 | -8.442\*\*\*(2.719) | 2.190(2.974) | 0.096(3.833) |  |  |
| 观测值 | 1996 | 1784 | 1253 |  |  |
| R2 | 0.075 | 0.166 | 0.223 |  |  |

注：农业户口和非农业户口均指个人最初获得的户口类型。使用数据为CHIP2002、2013和2018，样本年龄为23-35岁，采用Heckman两步法估计；各回归方程的控制变量还有子女和父母的年龄及年龄平方项。其他说明同表2。

（四）进一步的分析

1.不同收入的代际流动性

表6是基于工资性收入、可支配收入和主要工作收入测算的代际收入秩回归系数。其中，工资性收入样本来自CHIP2002、2013和2018，可支配收入和主要工作收入样本来自CHIP2013和2018，原因在于CHIP2002数据的财产性收入口径与后两个年份不一致，同时也没有调查个人主要工作收入。因此，对后两类收入的比较以“80后”和“90后”为主。

首先，从工资性收入的代际流动性来看，“80后”较“70后”有显著下降，变化趋势与劳动收入的代际流动性趋势相同，但“90后”与“80后”没有显著差异，这一变化不同于劳动收入。另外，从这两类收入的代际流动性大小来看，工资性收入的流动性往往更高，这一发现与众多研究一致（Deutscher & Mazumder，2020），即由于劳动收入中还包括经营性收入，尤其是农业经营净收入，会使代际传递程度更高，即代际流动性更低。其次，从可支配收入的代际流动性变化来看，“90后”较“80后”有显著下降，这一趋势与劳动收入的代际流动性趋势相同，并且流动性高于劳动收入的代际流动性。最后，从主要工作收入的代际流动性变化来看，“90后”与“80后”没有显著差异，这一发现与工资性收入相似。总的来看，虽然用不同收入衡量的代际流动性趋势略有不同，但同基准回归结论相近，没有证据表明我国居民代际收入流动性在增加。

对于同住样本选择偏差问题，从各方程的逆米尔斯比率系数来看，工资性收入和主要工作收入对同住样本选择偏差的敏感度较大，而劳动收入和可支配收入对同住样本选择偏差的敏感度较小。

表6 代际收入流动性趋势：不同收入指标

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | （1） | （2） | （3） | （4） | （5） |
| 70后 | 80后 | 90后 | 80后-70后 | 90后-80后 |
| 1.基准结果（劳动收入） | 0.397\*\*\*(0.020) | 0.435\*\*\*(0.021) | 0.493\*\*\*(0.023) | 0.038\*[0.080] | 0.058\*\*[0.047] |
| 2.工资性收入 | 0.297\*\*\*(0.045) | 0.413\*\*\*(0.024) | 0.395\*\*\*(0.029) | 0.116\*\*\*[0.002] | -0.018[0.315] |
| 逆米尔斯比率 | -9.062(6.574) | 13.448\*\*\*(2.837) | 13.926\*\*\*(3.759) |  |  |
| 观测值 | 520 | 1596 | 1116 |  |  |
| R2 | 0.113 | 0.219 | 0.226 |  |  |
| 3.可支配收入 | — | 0.347\*\*\*(0.020) | 0.410\*\*\*(0.020) | — | 0.063\*\*[0.015] |
| 逆米尔斯比率 |  | 4.423\*(2.404) | 0.971(2.690) |  |  |
| 观测值 |  | 2375 | 2294 |  |  |
| R2 |  | 0.138 | 0.174 |  |  |
| 4.主要工作收入 | — | 0.337\*\*\*(0.018) | 0.354\*\*\*(0.020) | — | 0.017[0.292] |
| 逆米尔斯比率 |  | -3.706\*(2.225) | -6.631\*\*(2.644) |  |  |
| 观测值 |  | 2808 | 2391 |  |  |
| R2 |  | 0.172 | 0.172 |  |  |

注：劳动收入和工资性收入的估计使用CHIP2002、2013和2018数据，可支配收入和主要工作收入使用CHIP2013和2018数据，未使用CHIP2002的原因是后两类收入口径不一致；样本年龄为23-35岁，采用Heckman两步法估计；各回归方程的控制变量还有子女和父母的年龄及年龄平方项。其他说明同表2。

2.再论同住样本选择偏差

同住样本选择偏差是代际收入流动性研究中不容忽视的一个重要问题。不同于国内其他家庭调查数据，CHIP2013城镇住户调查和CHIP2018城乡住户调查均询问了与户主不同住成年子女的主要工作月收入信息，因此可以直接考察同住样本选择偏差对代际收入流动性的影响。朱诗娥和唐瑾（2018）运用CHIP2013城镇住户家庭同住子女和不同住子女的主要工作月收入，发现只分析同住样本会严重低估城镇居民代际收入流动性。但是，考虑到只分析城镇住户并不能代表全国的情况，同时也会因为城乡人口流动现象的存在而出现估计偏误。因此，本文仅使用CHIP2018城乡所有住户家庭调查数据对此问题进行考察，同时分析代际收入流动性趋势。

表7为运用CHIP2018数据考察同住样本选择偏差的OLS估计结果，样本定义同基准分析。其中，分别测算了与户主同住成年子女样本和所有子女样本（同住子女+不同住子女）的代际收入流动性趋势。首先，可以发现仅分析同住样本时确实会低估代际收入流动性，表明同住样本选择偏差会导致我国居民代际收入流动性的低估。这一发现与前文运用Heckman两步法估计的结论一致，即通过与OLS估计结果的比较，发现不纠正同住样本选择偏差确实会导致代际收入流动性的低估。

其次，从变化趋势来看，两者均表明“80后”与“90后”的代际收入流动性未发生显著变化，但从数值来看，“90后”代际收入传递系数确实大于“80后”。这也同基准回归结果相近，即没有证据表明我国居民代际收入流动性在增加。

最后，需要说明的是，运用主要工作收入计算的代际收入流动性相对较高，可能的原因是这一研究样本并未包括主要从事农业生产经营的个体，而忽略从事农业生产经营的群体可能高估整体的代际流动水平。

表7 代际收入流动性趋势：同住样本选择偏差

|  |  |  |  |
| --- | --- | --- | --- |
|  | （1） | （2） | （3） |
| 80后 | 90后 | 90后-80后 |
| 1.同住样本 | 0.305\*\*\*(0.034) | 0.313\*\*\*(0.023) | 0.008[0.417] |
| 观测值 | 857 | 1925 |  |
| R2 | 0.094 | 0.110 |  |
| 2.同住+不同住样本 | 0.201\*\*\*(0.021) | 0.230\*\*\*(0.018) | 0.029[0.163] |
| 观测值 | 2292 | 3086 |  |
| R2 | 0.040 | 0.065 |  |

注：使用数据为CHIP2018，样本年龄为23-35岁，采用OLS估计；各回归方程的控制变量还有子女和父母的年龄及年龄平方项。收入指标为个人主要工作月收入。其他说明同表2。

五、结论与启示

代际收入流动关系社会公平和经济效率。本文运用2002年、2013年和2018年中国家庭收入调查数据，在考虑收入测量偏误和同住样本选择偏差的基础上，测算了我国1970-1995年出生群体的代际收入流动性趋势。研究发现：（1）我国1970-1995年出生群体的代际收入流动性呈下降趋势，代际收入秩回归系数从“70后”的0.397上升至“80后”的0.435，进而上升到“90后”的0.493。（2）异质性分析结果表明，男性和农业户口的代际收入流动性均呈下降趋势，但女性没有发生显著变化，非农业户口仅在“70后”至“80后”有显著下降，而“80后”与“90后”没有显著差异。（3）用不同类型收入衡量的代际流动性趋势略有不同。“90后”工资性收入和主要工作收入的代际流动性与“80后”没有显著差异，但劳动收入和可支配收入均呈下降趋势；而“80后”工资性收入和劳动收入的代际流动性较“70后”均呈下降趋势。总的来看，没有证据表明我国居民代际收入流动性在增加。（4）同住样本选择偏差对分性别、工资性收入和主要工作收入代际流动性的敏感度较大，但也没有改变我国居民代际收入流动性呈下降的总体趋势。

基于以上分析与发现，可以得出以下启示：第一，应该认清我国居民代际收入流动性趋势及特征，意识到社会阶层固化问题的严峻性。第二，应该积极采取措施提高代际收入流动性。一方面，异质性分析表明，男性群体和农业户口群体代际收入流动性下降是我国居民代际收入流动性下降的重要结构性原因，因此，应关注这部分群体内部的机会平等情况，提高其代际收入流动性。另一方面，不同收入的代际流动性比较表明，工资性收入比重的下降也可能导致我国居民代际收入流动性下降，因此，可以通过收入结构调整提高我国居民代际收入流动性。第三，应该认识到同住样本选择偏差对不同样本和不同类型收入代际流动性的敏感度不同，故测算代际收入流动性时有必要纠正这一计量偏误。

参考文献：

陈琳 袁志刚，2012：《中国代际收入流动性的趋势与内在传递机制》，《世界经济》第6期。

何石军 黄桂田，2013：《中国社会的代际收入流动性趋势:2000～2009》，《金融研究》第2期。

韩军辉 龙志和，2011：《基于多重计量偏误的农村代际收入流动分位回归研究》，《中国人口科学》第5期。

李任玉 陈悉榕 甘犁，2017：《代际流动性趋势及其分解:增长、排序与离散效应》，《经济研究》第9期。

刘琳 赵建梅，2020：《社会网络如何影响代际收入流动？》，《财经研究》第8期。

亓寿伟，2016：《中国代际收入传递趋势及教育在传递中的作用》，《统计研究》第5期。

汪德华 邹杰 毛中根，2019：《“扶教育之贫”的增智和增收效应——对20世纪90年代“国家贫困地区义务教育工程”的评估》，《经济研究》第9期。

王伟同 谢佳松 张玲，2019：《中国区域与阶层代际流动水平测度及其影响因素研究》，《数量经济技术经济研究》第1期。

汪小芹，2018：《中国社会代际流动趋势与结构分解》，《经济学动态》第11期。

汪小芹 邵宜航，2021：《我们是否比父辈过得更好:中国代际收入向上流动研究》，《世界经济》第3期。

王学龙 袁易明，2015：《中国社会代际流动性之变迁:趋势与原因》，《经济研究》第9期。

杨娟 张绘，2015：《中国城镇居民代际收入流动性的变化趋势》，《财政研究》第7期。

杨沫 王岩，2020：《中国居民代际收入流动性的变化趋势及影响机制研究》，《管理世界》第3期。

郑筱婷 袁梦 王珺，2020：《城市产业的就业扩张与收入的代际流动》，《经济学动态》第9期。

Abramitzky, R. et al(2021), “Intergenerational mobility of immigrants in the United States over two centuries”, *American Economic Review* 111(2):580-608.

Blanden, J. et al(2007), “Accounting for intergenerational income persistence: Noncognitive skills, ability and education”, *Economic Journal* 117(519):C43-C60.

Bowles, S. & H.Gintis(2002), “The inheritance of inequality”, *Journal of Economic Perspectives* 16(3):3-30.

Chetty, R. et al(2014a), “Is the United States still a land of opportunity? Recent trends in intergenerational mobility”, *American Economic Review* 104(5):141-147.

Chetty, R. et al(2014b), “Where is the land of opportunity? The geography of intergenerational mobility in the United States”, *Quarterly Journal of Economics* 129(4):1553-1623.

Dahl, M. W. & T.DeLeire(2008), “The association between children's earnings and fathers' lifetime earnings: Estimates using administrative data”, Discussion Paper No.1342-08, Institute for Research on Poverty, University of Wisconsin-Madison.

Deutscher, N. & B.Mazumder(2020), “Intergenerational mobility across Australia and the stability of regional estimates”, *Labour Economics* 66(C), https://doi.org/10.1016/ /j.labeco.2020.101861.

Deutscher, N. & B.Mazumder(2021). “Measuring intergenerational income mobility: A synthesis of approaches”, FRB of Chicago Working Paper, No. 09.

Emran, M.S. et al(2018), “When measure matters coresidency, truncation bias, and intergenerational mobility in developing countries”, *Journal of Human Resources* 53(3):589-607.

Fan, Y. et al(2021), “Rising intergenerational income persistence in China”, *American Economic Journal: Economic Policy* 13(1):202-230.

Fan, H. et al(2022), “Prevalence of SOEs and intergenerational income persistence: Evidence from China”, *World Economy* 45(1):276-291.

Gregg, P. et al(2017), “The role of education for intergenerational income mobility: A comparison of the United States, Great Britain, and Sweden”, *Social Forces* 96(1):121-152.

Harding, D.J. & M.D.Munk(2020), “The decline of intergenerational income mobility in Denmark: Returns to education, demographic change, and labor market experience”, *Social Forces* 98(4):1436-1464.

Jácome, E. et al(2021), “Mobility for all: Representative intergenerational mobility estimates over the 20th century”, NBER Working Paper, No. 29289.

Kyzyma, I. & O.Groh-Samberg(2020), “Estimation of intergenerational mobility in small samples: Evidence from German survey data”, *Social Indicators Research* 151(2):621-643.

Lee, C.I. & G.Solon(2009), “Trends in intergenerational income mobility”, *Review of Economics and Statistics* 91(4):766-772.

Luo, C. et al(2020), “The long-term evolution of national income inequality and rural poverty in China”, *China Economic Review* 62, <https://doi.org/10.1016/j.chieco.2020.101465>.

Mazumder, B.(2005), “Fortunate sons: New estimates of intergenerational mobility in the United States using social security earnings data”, *Review of Economics and Statistics* 87(2):235-255.

Nybom, M. & J.Stuhler(2017), “Biases in standard measures of intergenerational income dependence”, *Journal of Human Resources* 52(3):800-825.

Olivetti, C. & M.D.Paserman(2015), “In the name of the son (and the daughter): Intergenerational mobility in the united states, 1850-1940”, *American Economic Review* 105(8):2695-2724.

Raaum, O. et al(2007), “Marital sorting, household labor supply, and intergenerational earnings mobility across countries”, *B.E. Journal of Economic Analysis & Policy(Advances)* 7(2), Art.7.

Richey, J. & A.Rosburg(2017), “Changing roles of ability and education in US intergenerational mobility”, *Economic Inquiry* 55(1):187-201.

Simard-Duplain, G. & X.St-Denis(2020), “Exploration of the role of education in intergenerational income mobility in Canada: Evidence from the longitudinal and international study of adults”, *Canadian Public Policy* 46(3):369-396.

Solon, G.(1992), “Intergenerational income mobility in the United States”, *American Economic Review* 82(3):393-408.

1. 袁青青，北京师范大学经济与工商管理学院，邮政编码：100875，电子邮箱：yuanqq@ mail.bnu.edu.cn；刘泽云，北京师范大学经济与工商管理学院/首都教育经济研究基地，邮政编码：100875，电子邮箱：zeyun\_liu@bnu.edu.cn。本文受国家社会科学基金重大项目“中国农村家庭数据库建设及其应用研究”（编号18ZDA080）资助。感谢匿名审稿专家的宝贵意见，文责自负。 [↑](#footnote-ref-2)
2. 汪小芹（2018）和李任玉等（2017）分别研究了社会经济地位指数代际流动性和教育代际流动性的出生队列趋势，由于本文关注收入的代际流动性，故不再做综述比较。 [↑](#footnote-ref-3)
3. 关于代际收入流动性指标的详细分类和定义见文献Deutscher & Mazumder（2021）。这里再次感谢匿审专家提供的这一重要参考文献。 [↑](#footnote-ref-4)
4. 国内文献对这一方法的称呼略有不同，如“代际次序相关性”（王伟同等，2019）、“代际收入秩关联系数”（杨沫、王岩，2020）、“代际收入百分位排序关联系数”（刘琳、赵建梅，2020）等。而国外常见的用法有Rank-Rank Correlations/Slope（Chetty et al，2014b）、Rank-Rank Regressions（Simard-Duplain & St-Denis，2020）和Intergenerational Rank Association（Harding & Munk，2020）等。 [↑](#footnote-ref-5)
5. 关于个人农业经营净收入指标获得的合理性问题，通过考察CFPS数据和CHNS数据对个人农业收入的统计办法，本文认为基于劳动时间分摊获得的收入数据也是合理的。原因在于个人农业收入的报告也是基于家庭农业经营总收入以及其所做贡献估算得出，因此两种方法具有共通性。而关于这两种方法获得收入均可能存在的测量偏差问题，在下文计量问题处理部分做了考虑。当然，也有研究运用CHIP数据时，使用“人均家庭农业经营性收入”衡量个人农业经营收入，即“家庭农业经营性收入÷家庭劳动力人数”（汪德华等，2019），但考虑到指标精确性，本文采用了劳动时间分摊的办法。 [↑](#footnote-ref-6)
6. 关于样本中男性比例较高的原因，主要在于男性相对于女性与父母共同居住的比例更高，并且在使用户主及其配偶作为子代的情形中，男性的比例会更高。关于这一同住样本选择偏差问题，一方面通过Heckman两步法予以纠正，另一方面在稳健性检验部分仅考虑户主子女作为子代的样本进行估计。 [↑](#footnote-ref-7)
7. 虽然本文流动人口样本占比很小（2.6%），遗漏可能对结论没有影响，但从研究设计来讲，仍需要给予考虑。与本文不同，另一类由Fan et al（2021）在运用CFPS数据研究时提出的偏误类型，即同住家庭成员因流动而未调查收入信息导致样本缺失的情况，在CHIP数据中不存在这一问题。 [↑](#footnote-ref-8)
8. 对于因出国而未调查到的样本选择偏差，一方面可以归入同住偏差问题进行纠正；另一方面考虑到出国人员在全国的占比很低，所以对样本的代表性影响较小，可以忽略不计。但随着出国人数增加，未来在对大学和研究生群体的相关研究中可能需要关注。感谢匿审专家指出这一重要问题。 [↑](#footnote-ref-9)
9. 同时，本文也考虑限定3个出生队列的样本分别来自3个调查年份，并估计了3个调查年份均有的23-28岁、24-28岁、25-28岁等不同年龄范围的样本，也得到了稳健结论（限于篇幅，未在正文展示）。另外，这一分析也避免了同一队列样本来自不同调查年份带来的不可比性。 [↑](#footnote-ref-10)
10. 需要说明的是，由于研究代际收入流动性时，关注的核心是父子两代人持久收入的相关性，因此，这里的代表性年龄区间是确定并可比的，即代际收入流动性研究不涉及年龄效应分解问题。 [↑](#footnote-ref-11)
11. 考虑到估计的准确性，采用逐年回归和较小区间样本回归时，对样本量的要求相对较高。但在受样本量约束的情况下，也可以比较特定出生队列代际收入流动性的平均差异，如前文对“70后”、“80后”和“90后”的比较。 [↑](#footnote-ref-12)
12. 由于户籍制度的存在，城乡差异可以从户口和居住地两个视角进行考察。但在这两类分析中，均要对样本进行处理，即需要将户口和居住地还原为初始状态，否则分样本估计会出现选择偏差问题。由于还原发生过户口转换和居住地变化的样本时，户口划分的概念较城乡划分更准确和明晰（CHIP数据有个人户口转换信息），所以本文选择“分户口”的研究。另外，本文也估计了还原居住地后的城乡样本（即将城镇住户调查中户口经历过“农转非”的样本归入农村，将流动人口样本也归入农村），发现与分户口考察的结论基本一致，限于篇幅，未在正文展示。 [↑](#footnote-ref-13)