

# 中国代际收入流动：城乡差异及其时间趋势\*

胡霞 李文杰

**摘要：**本文利用 CHARLS2018 和 CHIP（1988、1995）的微观数据，采用双样本两阶段最小二乘法（TS2SLS），基于出生于 1970—1995 年的非迁移城镇子辈、非迁移农村子辈与城乡迁移子辈的代际组合样本，分析了中国代际收入流动性的城乡差异及其时间趋势，并进一步探究了不同城乡含义下中国代际收入流动性城乡差异的结构特征。研究结果显示，第一，从静态总体水平看，中国非迁移农村子辈的代际收入流动性显著低于非迁移城镇子辈，户籍城乡与区域城乡含义下的中国农村代际收入流动性都显著低于城镇，但存在程度差异。第二，从动态时间趋势看，出生于 1975 年之前的非迁移农村子辈的代际收入流动性高于出生于 1975 年之前的非迁移城镇子辈，但之后随着子辈出生年份的推移，非迁移农村子辈的代际收入流动性逐渐低于非迁移城镇子辈，并且两者之间的差异总体呈扩大趋势。第三，虽然同是初始户口为农村户口的子辈，但出生于 1973 年以后的城乡迁移子辈的代际收入流动性高于出生于 1973 年以后的非迁移农村子辈，并且，随着子辈出生年份的推移，两者的代际收入流动性差异呈现出先增大后缓慢减小而后又缓慢增大的趋势。本文的研究结果表明，虽然中国城乡内部都存在着一定程度的机会不平等，但农村的机会不平等程度相对更高，并且城乡之间的这种相对机会不平等差异在 1970—1995 世代中随世代的推移呈扩大趋势。

**关键词：**代际收入流动性 城乡迁移 城乡差异 时间趋势

**中图分类号：**F014.4 **文献标识码：**A

## 一、引言

随着精准扶贫目标的实现，绝对贫困的消除，在发展不平衡不充分的社会经济背景下，如何防止穷者越穷、富者越富，已成为中国实现共同富裕需要关注的重要问题。在代内视角下，该问题常被具体化为长期收入均等化问题（胡霞和李文杰，2016），而在代际视角下，该问题则可被具体化为代际收入流动性问题。代际收入流动性，其直观含义是子辈收入地位相对于父辈收入地位的变动程度，实际揭示的则是父辈收入或与之相关的家庭背景对子辈经济成就（economic achievement）的影响。由于家庭背景是子辈自身无法控制同时又对其经济成就具有重要影响的因素，所以，代际收入流动性也常

---

\*本文研究得到北京高校“双一流”建设资金的支持。作者感谢匿名审稿专家提出的宝贵意见，但文责自负。本文通讯作者：李文杰。

被视为衡量一个社会机会不平等程度的重要标志（Mayer and Lopoo, 2005）。

现有文献对中国代际收入流动性的研究主要集中于两个方面，一是对中国代际收入流动性总体水平、结构特征与时间趋势的估计，二是对中国代际收入流动内在机制的理论解释与实证检验。从政策含义角度看，前者有助于为“是否需要干预”提供判断依据，后者则有助于为“如何进行干预”提供理论指导，前者是后者的基础。然而，就第一方面而言，目前现有相关研究虽已取得一定成果，但尚未形成一致性结论（陈杰等，2016）。此外，与总体水平相比，代际收入流动性的结构特征与时间趋势更侧重于揭示不同群体或不同世代之间的机会不平等差异，因而也更有助于提高政策的针对性。基于此，本文拟主要关注中国代际收入流动性的城乡差异及其时间趋势。

关于中国代际收入流动性的城乡差异，从研究数据来源看，现有相关研究主要采用 CHIP（Chinese Household Income Project）与 CHNS（China Health and Nutrition Survey）数据，这两个数据存在一个比较明显的特征，即 CHIP 数据对被访住户成员的界定决定了其只包含同住子辈样本，CHNS 数据的追踪调查虽然包含了不同住子辈样本，但也只包含之前与父辈同住一户而后与父辈居住于同一村或同一社区的不同住子辈样本（江求川，2017），因此，利用 CHIP 与 CHNS 数据所提供的代际组合（parent-child pairs）样本估计得到的中国代际收入流动性的城乡差异，可能存在同住偏误。从研究结果的一致性看，陈琳和袁志刚（2012）、徐晓红（2015）基于 CHIP 数据研究发现，中国农村的代际收入流动性高于城镇；韩军辉（2010）、胡洪曙和元寿伟（2014）、江求川（2017）基于 CHNS 数据研究发现，中国农村的代际收入流动性低于城镇。但是，同样基于 CHNS 数据，杨沫和王岩（2020）以子辈的初始户口类型为依据划分城乡样本，其估计结果显示，中国农村的代际收入流动性高于城镇。关于上述研究结果的争议，从研究视角看，由于 CHIP 数据以户主的户口与居住地类型为依据划分城乡样本<sup>①</sup>，CHNS 数据以被访住户的所在地类型为依据划分城乡样本，所以，与以子辈初始户口类型为依据划分的城镇样本相比，CHIP 数据中的城镇样本不仅包含始终持有城镇户口且居住于城镇的群体<sup>②</sup>，还包含农村户口转为城镇户口且居住于城镇的城乡迁移群体<sup>③</sup>，而 CHNS 数据中的城镇样本不仅包含始终持有城镇户口且居住于城镇的群体，还包含由农村迁往城镇的所有城乡迁移群体（无论是否获得城镇户口，只需居住于城镇），因此，上述研究结果之间的争议可能源于城乡含义的差异。从中国的现实特征看，城乡二元户籍制度、城乡地理空间差异、城镇化过程中的城乡迁移以及个体户口类型的改变，四个因素相互叠加，使得中国的城乡含义并不唯一，初始户口类型、当前户口类型以及所在地的城乡类型，都可以成为界定城乡群体的依据。在不同的城乡含义下，中国代际收入流动性的城乡差异会呈现出怎样的结构特征，目前尚未有研究对该问题展开深入分析。

关于中国代际收入流动性城乡差异的时间趋势，总体而言，目前还未有文献对该问题展开深入研

<sup>①</sup>具体内容请参见 CHIP 公布的问卷指导手册。

<sup>②</sup>这里的“始终持有城镇户口”是指初始户口与调查年份的当前户口均为城镇户口。

<sup>③</sup>2002 年及其以后的 CHIP 数据将户主为农业户口且居住于户口所在地（乡镇）之外的样本视为外来务工样本，但现有研究在利用 CHIP 数据分析中国代际收入流动的城乡差异时，都没有提及这类样本的城乡归属。

究，尽管已有部分文献分析了中国城镇与农村各自的代际收入流动性的时间趋势（如陈琳和袁志刚，2012；杨沫和王岩，2020）。从经济含义角度看，城镇或农村代际收入流动性的时间趋势，关注的是城镇或农村内部的机会不平等程度是否随时间推移而改变，而代际收入流动性城乡差异的时间趋势，关注的则是城镇与农村的相对机会不平等是否随时间而改变。此外，从研究视角看，现有相关文献主要侧重于刻画（调查）年份视角下中国城镇与农村代际收入流动性的时间趋势，如陈琳和袁志刚（2012）、杨沫和王岩（2020）分别分析了1988—2005年、1993—2015年间中国城镇与农村代际收入流动性的时间趋势，很少有文献分析世代视角下中国城镇与农村代际收入流动性的时间趋势，而该视角下，中国代际收入流动性城乡差异的时间趋势，也还未有研究对其进行深入分析。相比较而言，世代视角下，代际收入流动性城乡差异的时间趋势，揭示的是出生并成长于不同年代的城乡世代群体相对机会不平等的变动特征。

相比CHIP与CHNS数据，CHARLS（China Health and Retirement Longitudinal Study）数据提供了更为充足的信息，即该数据一方面同时提供了同住与非同住子辈样本，可从数据源头避免同住偏误对中国代际收入流动性城乡差异估计结果的影响；另一方面同时提供了子辈的户口信息及其在调查年份的当前居住地信息，可为构建不同含义的城乡群体提供数据支持。本文拟基于CHARLS2018数据，首先采用双样本两阶段最小二乘法（Two-sample Two-stage Least Squares, TS2SLS）分析不同城乡含义下中国代际收入流动性城乡差异的结构特征，然后以子辈的出生年份为时间线，分析世代视角下中国代际收入流动性城乡差异的时间趋势。需要说明的是，不同世代在同一调查年份的年龄不同，单独分析世代视角下城镇或农村代际收入流动性的时间趋势，可能会得到由生命周期偏误导致的虚拟趋势，但是，如果出生于同一年份的城乡世代群体，在同一年龄时其代际收入流动性估计结果的生命周期偏误相同<sup>①</sup>，那么，世代视角下，代际收入流动性城乡差异时间趋势的估计结果则不存在这一问题。

## 二、模型与方法

本文使用的代际收入流动性测度指标为代际收入弹性（Intergenerational Income Elasticity, IGE），其基础估计方程如下：

$$\ln y_c = \beta \ln y_s + \varepsilon_c \quad (1)$$

（1）式中， $y_c$ 与 $y_s$ 分别为子辈与父辈的持久收入（permanent income）， $\beta$ 表示父辈收入变动1%所引起的子辈收入变动百分比，即代际收入弹性（IGE）。由于该指标反映的是子辈收入对父辈收入的依赖程度，所以， $\beta$ 越大意味着收入地位的代际传递性越高，代际流动性越低。

估计IGE需要个体的持久收入，但由于个体完整职业生涯的持久收入数据很难观测得到，所以，在具体的实证研究中，研究者常将年收入等短期收入作为样本个体持久收入的代理变量（proxy

<sup>①</sup>虽然目前尚未有研究对这一假设进行检验，但现有相关研究对中国代际收入流动性城乡差异的分析，实际上都暗含了这一假设，否则，如果生命周期偏误存在城乡差异，那么，现有研究基于出生年份范围相同的群体样本，估计得到的同一调查年份下的城乡代际收入流动性则无法直接进行比较。

variable)。但是，由于父辈或子辈的短期收入与其持久收入之间的关系，随着个体年龄的推移，可能存在生命周期性变化，父辈短期收入也可能存在测量误差或短期冲击，所以，使用短期收入往往会导致 IGE 的 OLS 估计存在生命周期偏误 (Haider and Solon, 2006; Böhlmark and Lindquist, 2006) 与衰减偏误 (Solon, 1992)。此外，由于有些调查数据只提供了子辈与父辈同住于一户的样本，所以，IGE 的估计可能还存在由样本选择问题导致的同住偏误。基于此，IGE 的准确估计依赖于对衰减偏误、生命周期偏误以及同住偏误的控制。本文选用的 CHARLS2018 数据同时提供了同住与非同住子辈样本的收入及其他个体特征信息，为解决同住偏误问题提供了条件。然而，在 CHARLS2018 中，大约 67% 以上的父辈样本在 2018 年超过了 60 岁，如果将父辈样本在 2018 年的收入作为其持久收入的代理变量，则可能会存在由父辈收入年龄过大而引致的生命周期偏误问题<sup>①</sup>。为解决这一问题，本文选用双样本两阶段最小二乘 (TS2SLS) 方法来构建父辈持久收入的代理变量。Jerrim et al. (2016) 的研究显示，该方法在代际收入流动领域已被广泛使用。

TS2SLS 估计方法的基本步骤如下：首先，构建主样本与辅助样本，其中，主样本包含子辈的收入与真父辈的收入预测变量 ( $X_s^{real}$ )，辅助样本包含伪父辈 (pseudo parent) 的收入 ( $\ln y_s^{pseudo}$ ) 及其收入预测变量 ( $X_s^{pseudo}$ )；然后，基于辅助样本估计如下收入决定方程：

$$\ln y_s^{pseudo} = X_s^{pseudo} \gamma + \mu_s^{pseudo} \quad (2)$$

其次，利用 (2) 式中伪父辈收入预测变量的回归系数 ( $\widehat{\gamma}_{OLS}$ ) 与主样本中真父辈样本的收入预测变量 ( $X_s^{real}$ ) 估计真父辈样本的收入 ( $\ln y_s^{real}$ )，具体估计方程如下：

$$\ln y_s^{real} = X_s^{real} \widehat{\gamma}_{OLS} \quad (3)$$

最后，使用主样本中的子辈收入与真父辈的估计收入 ( $\ln y_s^{real}$ ) 来估计 IGE。

关于  $\beta_{TS2SLS}$  的统计性质，如果将  $\beta_{TS2SLS}$  与  $\beta_{TSIV}$  相比较，两者依概率收敛于相同的极限 (Inoue and Solon, 2010)，但  $\beta_{TS2SLS}$  更渐近有效。如果将  $\beta_{TS2SLS}$  视为  $\beta_{TSIV}$  的一种变形，根据 Björklund and Jäntti (1997) 的研究， $\beta_{TSIV}$  与  $\beta_{IV}$  渐近等价的条件为两个样本来自同一总体。如果将  $\beta_{TS2SLS}$  与  $\beta_{OLS}$  相比较，根据 Nicoletti and Ermisch (2007) 的研究， $\beta_{TS2SLS}$  渐近等价于  $\beta_{OLS}$  的核心条件为  $X_s^{real}$  与  $X_s^{pseudo}$  的选择能够使第一阶段的收入决定方程的  $R^2$  尽可能地接近于 1<sup>②</sup>。由此可见， $\beta_{TS2SLS}$  的一致性 (consistency) 与渐近有效性 (asymptotic efficiency) 要求，TS2SLS 方法的使用至少应该满足以下

<sup>①</sup> “收入年龄”即样本个体获得其被观测到的单年收入的年龄。为了表述简便，本文将其简称为收入年龄。关于由子辈或父辈的收入年龄过小或过大引致的生命周期偏误，具体分析请参见 Haider and Solon (2006)。

<sup>②</sup>  $\beta_{TS2SLS}$  为 TS2SLS 估计量， $\beta_{OLS}$  为 OLS 估计量。 $\beta_{IV}$  为使用传统工具变量 (Instrumental Variables, IV) 时得到的估计量，这里的传统是指 IV 来自同一样本。 $\beta_{TSIV}$  是使用双样本工具变量 (two-sample Instrumental Variables, TSIV) 时得到的估计量，这里的工具变量来自不同样本。关于  $\beta_{TS2SLS}$  与  $\beta_{TSIV}$ 、 $\beta_{TSIV}$  与  $\beta_{IV}$ 、 $\beta_{TS2SLS}$  与  $\beta_{OLS}$  的详细比较，请分别参见 Inoue and Solon (2010)、Björklund and Jäntti (1997) 与 Nicoletti and Ermisch (2007)，本文不再赘述。

两个条件：①主样本与辅助样本来自同一总体；②收入决定方程的预测效果相对较强。基于此，本文将主要依据这两个条件来构建辅助样本，选择收入预测变量。

由于本文利用 TS2SLS 方法估计得到的真父辈收入只包含其收入中的持久收入部分（permanent component），所以，该方法在一定程度上也控制了由父辈短期收入的测量误差或短期波动而引致的衰减偏误。此外，关于生命周期偏误，现有实证研究主要通过选用单年收入与持久收入较为接近的年龄阶段的样本，同时在 IGE 估计方程中控制子辈与父辈的收入年龄，对生命周期偏误进行控制，本文也将采用上述方法。其中，在 TS2SLS 方法下，控制了子辈与父辈收入年龄的 IGE 估计方程如下：

$$\ln y_{ct} = \alpha + \beta \ln y_{sd}^{real} + \varphi_{c1} age_{ct} + \varphi_{c2} age_{ct}^2 + \varphi_{s1} age_{sd} + \varphi_{s2} age_{sd}^2 + \varepsilon_{ct} \quad (4)$$

(4) 式中， $y_{ct}$ 、 $age_{ct}$  分别为子辈在年份  $t$  的收入（作为子辈的持久收入代理变量）与年龄， $\ln y_{sd}^{real}$ 、 $age_{sd}$  分别为父辈在年份  $d$  的估计收入（作为父辈的持久收入代理变量）与年龄。

### 三、数据来源、样本构建与核心变量

#### （一）数据来源

本文数据来源于中国健康与养老追踪调查（CHARLS）和中国家庭收入调查（CHIP）。CHARLS 是北京大学主持开展的入户调查项目，其全国基线调查始于 2011 年，覆盖了全国 28 个省级行政区下辖的 150 个县 450 个社区或村庄，目前共公布了 2011 年、2013 年、2014 年、2015 年与 2018 年共 5 个年份的调查数据。从数据可得性看，CHARLS 不仅提供了被访者及其配偶以及两者子女的收入与个体特征信息，还提供了有助于识别子辈是否与其父辈同住的相关信息。此外，CHARLS2014 追踪了被访者及其配偶的户口类型与户口所在地的变迁史、第一份工作到当前工作的就业史、居住地的迁移史等信息，为本文构建父辈持久收入代理变量提供了数据支持。CHIP 是先后由国家统计局与中国社会科学院经济研究所、北京师范大学以及国外高校联合开展的入户调查项目，到目前为止，已相继公布了 1988 年、1995 年、2002 年、2007 年、2008 年和 2013 年共 6 个年份的调查数据。该数据提供了每个调查年份样本住户的家庭成员关系、家庭成员的个体特征以及上一年份的个体收入信息，为本文构建辅助样本提供了条件。

#### （二）样本构建

本文的主样本来自 CHARLS 2018，其中，父辈样本为被访者及其配偶，子辈样本为被访者与配偶的子女。首先，为控制由子辈引致的生命周期偏误，本文排除了 2018 年子辈小于 23 岁或大于 48 岁的代际组合样本；其次，为避免异常代际组合，代际年龄差不足 18 岁的样本也被排除在外。在上述约束条件下，本文得到的子辈样本出生于 1970—1995 年，父辈样本出生于 1922—1976 年。

在上述样本中，大部分父辈样本在 2018 年超过了 60 岁，为避免由父辈引致的生命周期偏误，本文将采用 TS2SLS 方法估计父辈样本在 23~50 岁期间的收入，并将其作为父辈持久收入的代理变量。为了便于区分与表述，下文将来自 CHARLS2018 的父辈样本称为真父辈样本。由于部分真父辈样本

出生年份较早，而在目前可公开获得的微观调查数据中，CHIP 数据的调查年份最早，为 1988 年，因此，为尽可能保留出生年份较早的父辈样本，本文将利用 CHIP1988 来构造伪父辈样本，估计父辈收入决定方程，以据此估计真父辈样本在 23~50 岁期间的收入。但是，在 1988 年，出生于 1989—1995 年的子辈样本还尚未出生，所以，本文也将同时利用 CHIP1995 来构造伪父辈样本，估计父辈收入决定方程，同时也据此估计真父辈样本在 23~50 岁期间的收入。此外，由于在利用伪父辈样本估计父辈收入决定方程时，一方面需要避免因子辈参与家庭经营而导致的家庭经营性收入无法分割的问题，另一方面需要避免因子辈赡养父辈而给予父辈收入的问题，所以，本文对伪父辈样本施加了如下约束条件，即伪父辈样本的子辈在伪父辈的收入年份（1988 年或 1995 年）不得超过 18 岁。由于子辈出生年份范围相同是现有文献为保证伪父辈样本与真父辈样本来自同一总体所使用的约束条件之一（Björklund and Jäntti, 1997），所以，本文也排除了子辈在 1988 年或 1995 年超过 18 岁的真父辈样本。基于此，本文对真父辈样本施加了如下约束条件：①子辈出生于 1970—1988 年的父辈样本在 1988 年的收入年龄为 23~50 岁；②子辈出生于 1977—1995 年的父辈样本在 1995 年的收入年龄为 23~50 岁。

此外，由于本文的研究对象包含农村样本，而农村样本中的家庭经营性收入存在难以分割的问题，所以，本文中的父辈一代为父母，相对应的父辈收入为父母总收入。为避免父辈婚姻状况差异所导致的影响，本文追踪了父辈样本的婚姻史，并进一步选取了 1988 年或 1995 年至少至 2011 年父辈均为已婚且最近一次结婚的年份不晚于 1988 年或 1995 年的真父辈样本<sup>①</sup>。在上述约束条件下，本文得到的真父辈样本如下：子辈出生于 1970—1988 年的父辈样本出生于 1938—1965 年，子辈出生于 1977—1995 年的父辈样本出生于 1945—1972 年。

关于伪父辈样本的构建，在具体的实证研究中，为尽可能地使伪父辈样本与真父辈样本来自同一总体，使用 TS2SLS 方法的研究者构建伪父辈样本会采用一定的约束条件，如表 1 所示。

表 1 TS2SLS 伪父辈样本的构建条件

相关文献	辅助样本的核心约束条件
Björklund and Jäntti (1997)	伪父亲样本的孩子与真父亲样本的孩子的出生年份范围相同
Lefranc and Trannoy (2005)	伪父亲样本的出生年份范围与真父亲样本的出生年份范围相同
Piraino (2007)	伪父亲样本的出生年份范围与真父亲样本的出生年份范围相同
Nicoletti and Ermisch (2007)	伪父亲样本的出生年份范围是真父亲样本的出生年份范围的子集

在借鉴现有相关研究方法的基础上，本文对伪父辈样本施加了如下三个约束条件：①伪父辈样本

<sup>①</sup>对于 1988（或 1995）年父亲已婚而（现）母亲单身以及 1988（或 1995）年母亲已婚而（现）父亲单身的样本，由于无法追踪到 1988（或 1995）年已婚父亲或母亲的配偶信息，所以，本文不得不选择 1988（或 1995）年至少至 2011 年均处于已婚状态且最近一次结婚的年份不晚于 1988（或 1995）年的父辈样本。这里的 2011 年是 CHARLS 最早公开的调查数据的年份。虽然本文使用的是 2018 年的数据，但由于 CHARLS 是追踪数据，父辈样本在 1988（或 1995）年的婚姻状况实际上是基于 CHARLS2011—2018 的追踪数据而整理得到。

与真父辈样本的出生年份相同；②伪父辈样本与真父辈样本的子辈出生年份相同；③伪父辈样本与真父辈样本的平均年龄相差不大。基于此，本文首先选取了CHIP1988中出生于1938—1965年的户主与配偶来构造1988年伪父辈样本，选取了CHIP1995中出生于1945—1972年的户主与配偶来构造1995年伪父辈样本；然后在此基础上，进一步选取子辈出生于1970—1988年且代际年龄差不小于18岁的1988年伪父辈样本，子辈出生于1977—1995年且代际年龄差不小于18岁的1995年伪父辈样本；最后，由于1988年伪父辈样本的平均年龄与真父辈样本的平均年龄相差较大（如表2所示），本文以真父亲样本的年龄分布中各年龄段的样本比例与伪父亲样本的年龄分布中相对应的各年龄段的样本比例的比值为权重，对1988年伪父辈样本进行加权调整，以尽可能保证1988年伪父辈样本与真父辈样本的平均年龄相差不大。表2汇报了真父辈样本与伪父辈样本的年龄特征。

表2 真父辈样本与伪父辈样本的年龄特征

		真父辈	伪父辈			真父辈	伪父辈
1988	父亲年龄范围	23~50	23~50	1995	父亲年龄范围	23~50	23~50
农村	母亲年龄范围	23~50	23~50	农村	母亲年龄范围	23~50	23~50
父辈	父母平均年龄	33.57	37.27	父辈	父母平均年龄	37.31	37.60
样本	加权后父母平均年龄		33.71	样本	加权后父母平均年龄		
1988	父亲年龄范围	23~50	23~50	1995	父亲年龄范围	23~50	23~50
城镇	母亲年龄范围	23~48	23~48	城镇	母亲年龄范围	23~50	23~50
父辈	父母平均年龄	33.85	36.95	父辈	父母平均年龄	37.17	38.20
样本	加权后父母平均年龄		34.06	样本	加权后父母平均年龄		

真父辈样本与伪父辈样本具有相同的收入预测变量（predictor variables）是使用TS2SLS方法的基础，所以，在构建两者所共同拥有的收入预测变量时，本文对主样本与辅助样本还施加了进一步的约束条件。基于此，本文最终使用的主样本如下：①子辈样本出生于1970—1988年的代际组合样本共6344个<sup>①</sup>，其中，父辈样本出生于1938—1965年；②子辈样本出生于1977—1995年的代际组合样本共5134个，其中，父辈样本出生于1945—1972年<sup>②</sup>。本文最终使用的辅助样本包含10182个1988年伪父辈样本与8302个1995年伪父辈样本。

### （三）核心变量

1. 父辈收入。本文构建的父辈持久收入代理变量为基于TS2SLS方法估计得到的父辈在1988年或

<sup>①</sup>对于多子家庭，本文采用了陈杰等（2016）的处理方法，将来自多子家庭的多个“子辈—父辈”代际组合视为多个观测样本。为控制由子辈来自同一家庭所导致的干扰项相关问题，本文采用了聚类标准误（cluster standard error）。

<sup>②</sup>CHARLS是一套旨在收集代表中国45岁及以上中老年人家庭和个人情况的微观调查数据，其受访者在2018年的年龄都不低于45岁。该数据特征意味着，基于CHARLS数据构建样本时，还需考虑样本的代表性。本文为控制由子辈、父辈引致的生命周期偏误以及父辈与子辈的代际年龄差所施加的样本约束条件（sample constraints），在很大程度上确保了样本的代表性。在所述约束条件下，本文最终使用的父辈样本，即CHARLS2018中出生于1938—1972年的受访者及其配偶，在2018年的年龄都不小于45岁。

1995年的收入。关于父辈的收入预测变量，首先，根据人力资本理论，受教育水平与工作经验是个体收入的重要影响因素；其次，根据徐林清（2006）、许庆红（2017）的研究，中国劳动力市场存在着部门（单位所有制性质）、职业、行业、地区等多重分割，由于劳动力市场的不同分割领域存在着不同的收入决定机制，所以，在中国，就业单位、职业、行业以及所在地区也是个体收入的重要影响因素。此外，刘精明（2006）的研究结果显示，个体的党员身份，作为其政治资本的测度变量，在中国劳动力市场上也具有市场回报。基于此，在综合考虑 CHARLS2018 中真父辈样本与 CHIP1988、CHIP1995 中伪父辈样本收入预测变量可得性的基础上<sup>①</sup>，本文选用的父辈收入预测变量为父辈的年龄及其平方（作为其工作经验的测度变量）、教育程度、党员身份、民族身份以及父辈在 1988 年或 1995 年的单位、职业与所在省份。表 3 和表 4 汇报了父辈收入决定方程的估计结果，其中，父辈收入为父亲和母亲在 1988 年或 1995 年的总收入。

表 3 农村父辈收入决定方程估计结果

关键变量	CHIP1988 农村	关键变量	CHIP1995 农村
年龄	0.005 (0.014)	年龄	0.044*** (0.015)
年龄平方	0.0002 (0.0002)	年龄平方	-0.0003 (0.0002)
至少 1 个父辈为党员	0.074*** (0.022)	至少 1 个父辈为党员	0.064** (0.026)
至少 1 个父辈为少数民族	-0.194*** (0.035)	至少 1 个父辈为少数民族	-0.192*** (0.036)
受教育程度（参照组：小学以下）		受教育程度（参照组：小学以下）	
小学	0.104*** (0.026)	小学	0.037 (0.041)
初中	0.127*** (0.026)	初中	0.051 (0.040)
高中及以上	0.172*** (0.032)	高中	0.055 (0.043)
		大专及以上学历	0.265*** (0.092)
职业与单位（参照组：父辈都是农民）		职业与单位（参照组：父辈都是农民）	
都非农就业	0.398*** (0.045)	都受雇	0.278*** (0.072)
公有单位负责人+农民	0.169*** (0.062)	受雇者+非农个体户主	0.412*** (0.106)
国有单位专业技术人员+农民	0.071 (0.054)	都是非农个体户主	0.349*** (0.087)
集体单位专业技术人员+农民	0.140** (0.064)	非农就业者+其他	0.174* (0.092)
私营个体专业技术人员+农民	0.178*** (0.069)	国有单位负责人或专业技术人员+农民	0.226*** (0.061)
其他单位专业技术人员+农民	0.163* (0.091)	集体单位负责人或专业技术人员+农民	0.140 (0.096)
国有单位普通员工+农民	0.203*** (0.066)	公有单位普通员工+农民	0.139*** (0.035)
集体单位普通员工+农民	0.158*** (0.045)	私营单位受雇者+农民	0.055 (0.051)
私营个体普通员工+农民	0.122*** (0.042)	其他单位受雇者+农民	0.113 (0.092)
其他单位普通员工+农民	0.113 (0.076)	村干部+农民	0.074 (0.046)
村干部+农民	0.149*** (0.045)	非农就业者+村干部	0.287*** (0.088)
非农就业者+村干部	0.447*** (0.065)	非农个体户主+农民	0.193*** (0.038)
		农民+其他	0.003 (0.049)

<sup>①</sup>关于真父辈样本与伪父辈样本可获得的收入预测变量，限于篇幅未列出详细信息，感兴趣的读者可向作者索取。



中国代际收入流动：城乡差异及其时间趋势

		都是其他	0.167 (0.127)
常数项	6.812*** (0.266)	常数项	6.808*** (0.286)
被访时居住地所在省份	虚拟变量	被访时居住地所在省份	虚拟变量
样本容量	5478	样本容量	4676
调整 R <sup>2</sup>	0.279	调整 R <sup>2</sup>	0.300
复相关系数	0.528	复相关系数	0.548

注：\*\*\*、\*\*、\*分别表示 1%、5%、10%的显著性水平；括号内的数字为标准误。

表 4 城镇父辈收入决定方程估计结果

关键变量	CHIP1988 城镇	关键变量	CHIP1995 城镇
年龄	0.027** (0.011)	年龄	0.048*** (0.013)
年龄平方	-0.0002 (0.0001)	年龄平方	-0.0005*** (0.0002)
至少 1 个父辈为党员	0.015 (0.010)	至少 1 个父辈为党员	0.054*** (0.012)
至少 1 个父辈为少数民族	-0.028* (0.016)	至少 1 个父辈为少数民族	-0.020 (0.025)
受教育程度（参照组：小学及以下）		受教育程度（参照组：小学及以下）	
初中	0.039** (0.020)	初中	0.077 (0.054)
高中	0.053*** (0.020)	高中	0.148*** (0.054)
大专及以上学历	0.068*** (0.022)	大专	0.191*** (0.056)
		本科及以上	0.286*** (0.058)
职业（参照组：父辈都是普通员工）		职业（参照组：父辈都是普通员工）	
都是负责人	0.088*** (0.031)	都是负责人	0.162** (0.079)
负责人+专业技术人员	0.066*** (0.025)	负责人+专业技术人员	0.145*** (0.039)
都是专业技术人员	0.056*** (0.015)	都是专业技术人员	0.155*** (0.018)
负责人+普通员工	0.044*** (0.014)	负责人或专业技术人员+普通员工	0.072*** (0.015)
专业技术人员+普通员工	0.028*** (0.010)	至少 1 个是私营或个体户主	0.032 (0.056)
至少 1 个为私营或个体户主	0.029 (0.048)	负责人或专业技术人员+其他	0.036 (0.037)
受雇者+其他	-0.073 (0.057)	普通员工+其他	-0.033 (0.026)
		都是其他	0.045 (0.031)
单位（参照组：父辈都在集体单位）		单位（参照组：父辈都在集体单位）	
都在国有单位	0.097*** (0.019)	都在国有单位	0.165*** (0.023)
国有单位+集体单位	0.037* (0.020)	国有单位+集体单位	0.101*** (0.025)
公有单位+其他单位	-0.140** (0.055)	至少 1 个在外资或合资单位	0.194*** (0.055)
至少 1 个在私营或个体户	0.149 (0.103)	公有单位+私营或个体户	0.080** (0.040)
		都在私营或个体户	0.028 (0.063)
		至少 1 个在其他单位	0.011 (0.049)
常数项	7.256*** (0.204)	常数项	7.683*** (0.240)
被访时居住地所在省份	虚拟变量	被访时居住地所在省份	虚拟变量
样本容量	4704	样本容量	3626
调整 R <sup>2</sup>	0.375	调整 R <sup>2</sup>	0.428

复相关系数	0.612	复相关系数	0.654
-------	-------	-------	-------

注：\*\*\*、\*\*、\*分别表示 1%、5%、10% 的显著性水平；括号内的数字为标准误。

由于收入决定方程的预测效果相对较强是 TS2SLS 估计量满足一致性与渐近有效性的重要条件之一，所以，本文主要关注收入决定方程对父辈收入的整体预测效果，而不是单个预测变量的因果效应。虽然回归模型的（调整） $R^2$  是衡量收入决定方程预测效果的重要指标，但由于缺乏判断标准，同时，回归直线的预测功能又与变量的相关关系密切相关（穆尔和诺茨，2017），所以，本文检验了上述收入决定方程中父辈收入预测变量与父辈收入的复相关系数（multi-correlation coefficient）。相比较而言，相关关系的强弱存在一个经验标准，即相关系数小于 0.5 时常被视为弱相关。基于此，由表 3 和表 4 可知，本文选用的父辈收入预测变量与父辈收入的复相关系数都在 0.5 以上，意味着，本文估计得到的 1988 年与 1995 年城镇与农村的父辈收入决定方程，具有相对较强的预测效果。

2. 子辈收入。本文将子辈 2017 年的年收入作为其持久收入的代理变量。由于 CHARLS2018 提供的子辈收入数据为区间形式<sup>①</sup>，所以，本文利用区间回归（interval regression）方法估计得到子辈样本在 2017 年的年收入（单位：元，自然对数形式）。区间回归，本质上将收入区间视为一个定序分类变量，但又与经典的定序回归（ordered regression）不同。定序回归估计得到的是个体收入落入某一收入区间的概率，而区间回归通过不断迭代估计得到的是个体收入估计值的均值（the predicted mean），并且区间回归中解释变量的回归系数的经济含义也是解释变量对被解释变量估计值（如个体收入估计值）的边际效应。关于估计子辈年收入的解释变量，根据人力资本理论以及中国劳动力市场的分割特征，在综合考虑 CHARLS2018 中子辈个体特征信息可获得性的基础上，本文选用的解释变量包括子辈的年龄及其平方、性别、婚姻状况、教育程度以及子辈在调查年份的职业、户口类型与居住位置。表 5 汇报了子辈年收入的区间回归估计结果。

表 5 子辈年收入估计（区间回归）

城镇子辈	回归系数	农村子辈	回归系数
年龄	0.105*** (0.024)	年龄	0.105*** (0.024)
年龄平方	-0.001*** (0.0003)	年龄平方	-0.001*** (0.0003)
性别与婚姻（参照组：单身女）		性别与婚姻（参照组：单身女）	
单身男	0.101 (0.063)	单身男	0.058 (0.101)
已婚男	0.617*** (0.053)	已婚男	0.510*** (0.092)
已婚女	0.660*** (0.053)	已婚女	0.617*** (0.093)
受教育程度（参照组：小学及以下）		受教育程度（参照组：小学以下）	
初中	0.148*** (0.037)	小学	0.145*** (0.042)
高中	0.333*** (0.053)	初中	0.290*** (0.042)

<sup>①</sup>子辈的年收入（单位：元）区间为：0、(0, 2000)、[2000, 5000)、[5000, 10000)、[10000, 20000)、[20000, 30000)、[30000, 50000)、[50000, 100000)、[100000, 150000)、[150000, 200000)、[200000, 300000)、[300000, +∞)。

中国代际收入流动：城乡差异及其时间趋势

中专	0.362*** (0.050)	高中	0.352*** (0.064)
大专	0.519*** (0.049)	中专	0.396*** (0.071)
本科及以上	0.755*** (0.048)	大专	0.567*** (0.086)
		本科及以上	0.741*** (0.103)
职业(参照组:农、林、牧、副、渔、水利从业者)		职业(参照组:农、林、牧、副、渔、水利从业者)	
各类单位负责人	0.600*** (0.085)	各类单位负责人	1.073*** (0.103)
专业技术人员	0.383*** (0.074)	专业技术人员	0.574*** (0.044)
办事员及相关人员	0.270*** (0.082)	办事员及相关人员	0.488*** (0.090)
商业、服务业人员	0.330*** (0.071)	商业、服务业人员	0.604*** (0.039)
生产与交运工人	0.160** (0.074)	生产与交运工人	0.522*** (0.037)
其他	0.118 (0.083)	其他	0.296*** (0.053)
当前户口(农村户口=0)	0.042 (0.029)	当前户口(农村户口=0)	0.099 (0.073)
当前居住位置(参照组:与父辈同住经济不独立)		当前居住位置(参照组:与父辈同住经济不独立)	
与父辈同住但经济独立	0.141* (0.082)	与父辈同住但经济独立	0.223*** (0.051)
与父辈邻院或邻公寓	0.178** (0.085)	与父辈邻院或邻公寓	0.182*** (0.052)
与父辈不相邻	0.153** (0.067)	与父辈不相邻	0.195*** (0.046)
当前居住地类型(参照组:城乡或镇乡接合区)			
城或镇中心区	0.122*** (0.028)		
特殊区域	0.156* (0.088)		
常数项	7.385*** (0.430)	常数项	7.036*** (0.455)
样本量	3910	样本量	3986

注:\*\*\*、\*\*、\*分别表示1%、5%、10%的显著性水平;括号内的数字为标准误。

表6汇报了父辈样本与子辈样本的基本统计特征,其中,子辈出生于1977—1988年的父辈样本的持久收入代理变量为父辈1988年与1995年估计收入的均值<sup>①</sup>,所有收入都已调整为2017年的价格。根据表6所示,父辈样本在1988年或1995年的平均年龄约为36岁,子辈样本在2018年的平均年龄约为36.5岁,父辈样本在1988年或1995年的平均收入低于子辈样本在2017年的平均收入,反映了中国经济快速增长所带来的居民收入的增加。此外,对于年龄介于23~48岁的子辈而言,尚未结婚或与父辈同住的子辈样本相对较少,与中国的现实状况也比较一致。

表6 父辈与子辈的基本统计特征

变量	均值	标准差	最小值	最大值
父辈1988年或1995年的年龄(父母平均年龄)	35.92	5.23	23	50
父辈1988年或1995年的收入(自然对数形式)	9.02	0.33	8.10	10.34
子辈2018年的年龄	36.45	6.20	23	48
子辈2017年的收入(自然对数形式)	10.47	0.90	7.22	12.98

<sup>①</sup>由于部分父辈样本在1988年或1995年的追溯信息缺失,在子辈出生于1977—1988年的代际组合样本中,大约26%的父辈样本的持久收入代理变量为父辈在1988年或1995年的单年估计收入。

子辈性别（男=1，女=0）	0.53	0.50	0	1
子辈婚姻（单身=1，已婚=0）	0.12	0.33	0	1
子辈同住（同住=1，不同住=0）	0.18	0.39	0	1
样本量	7896	7896	7896	7896

#### 四、估计结果

##### （一）不同城乡含义下代际收入流动性的城乡差异

1.基础模型。在现有实证研究中，IGE的一般估计方程如公式（4）所示，但由于本文采用的子辈样本同时包含与父辈同住和不同住的单身以及已婚的儿子与女儿，为避免子辈性别、婚姻与同住特征异质性对代际收入流动性城乡差异估计结果的影响，本文采用的基础模型如下：

$$\ln y_{ct} = \alpha + \beta \ln y_{sd}^{real} + \varphi_{c1}age_{ct} + \varphi_{c2}age_{ct}^2 + \varphi_{s1}age_{sd} + \varphi_{s2}age_{sd}^2 + \varphi_{c3}gender_{ct} + \varphi_{c4}single_{ct} + \varphi_{c5}coreside_{ct} + \varepsilon_{ct} \quad (5)$$

（5）式中， $gender_{ct}$ 、 $single_{ct}$ 、 $coreside_{ct}$ 分别为子辈在年份 $t$ 是否为男性、是否单身、是否与父辈同住的虚拟变量，其他变量的含义与公式（4）相同。

2.城乡划分。鉴于中国城乡二元户籍制度与城乡地理空间的相互叠加、个体的城乡迁移行为以及户口类型的改变、父辈与子辈的户口类型与居住地类型可能存在差异等现实特征，在分析中国代际收入流动性的城乡差异时，如何划分城乡样本至少应该明确以下信息：①“子辈—父辈”代际组合的城乡归属界定是以子辈还是父辈为主体；②主体城乡归属的界定依据是其出生时的户口类型还是当前时期的户口类型，抑或是其出生地类型、成长时期的居住地类型还是其当前的居住地类型；③由于无论依据户口类型还是居住地类型界定主体的城乡归属，都存在着由农村迁往城镇的迁移者，所以，主体的城乡归属界定还需考虑其由乡向城的迁移行为。

基于此，由于本文测度的是子辈一代的代际收入流动性，所以，本文将依据子辈的城乡归属来界定“子辈—父辈”代际组合的城乡归属。关于子辈的城乡归属，结合 CHARLS2018 所提供的子辈样本的户口与居住地信息<sup>①</sup>，本文对子辈城乡归属的基础界定如下：①如果子辈的初始户口与调查年份（2018年）的当前户口均为农村户口且子辈在调查年份居住于农村，则该子辈被视为非迁移农村子辈；②如果子辈的初始户口与调查年份的当前户口均为城镇户口且子辈在调查年份居住于城镇，则该

<sup>①</sup>CHARLS2018 只提供了子辈样本在 2018 年的当前户口类型与当前居住地类型，子辈样本的初始户口类型主要通过 CHARLS2011—2013 的追踪数据得到，CHARLS 没有提供子辈样本出生地与成长时期所在地的城乡类型，所以，在界定子辈样本的城乡归属时，本文只参考了子辈的初始与当前户口类型及其当前居住地的城乡类型。此外，由城镇户口变为农村户口、由农村户口变为城镇户口但居住于农村、始终持有城镇户口但居住于农村的子辈样本分别占总样本的 1.19%、1.28% 与 0.53%，这些类别的样本量相对较少，单独作为一类子样本可能会影响估计结果的稳健性，所以被本文排除。因此，本文最终用于分析中国代际收入流动性城乡差异的样本实际只包含 7659 个代际组合。

类子辈被视为非迁移城镇子辈；③如果子辈的初始户口与调查年份的当前户口均为农村户口，但子辈在调查年份居住于城镇，则该子辈被视为户口未变式城乡迁移子辈；④如果子辈的户口由农村户口转为城镇户口，并且子辈在调查年份居住于城镇，则该子辈被视为户口改变式城乡迁移子辈。此外，由于本文分析的是中国代际收入流动性的城乡差异，所以，本文只关注由农村向城镇的迁移，没有考虑不同城市之间或不同农村之间或不同距离的迁移。

基于上述子辈城乡归属的基础界定，本文进一步分析了户籍城乡与区域城乡含义下中国代际收入流动性的城乡差异。户籍城乡，根据子辈户籍类型划分的城乡，其本质是一种社会身份意义上的城乡属性。在中国户籍制度的相关安排下，户籍与一系列的发展机会、权利等密切相关，所以，城乡户籍身份常被视为中国社会身份结构的一个重要维度（李路路，2019）。因此，户籍城乡含义下的代际收入流动性城乡差异侧重于揭示中国机会不平等的社会结构性差异。区域城乡，根据子辈成年后的居住地类型划分的城乡，其本质是一种地理空间意义上的城乡属性，该含义下的代际收入流动性城乡差异侧重于揭示中国机会不平等的空间结构性差异。此外，由于子辈初始户籍身份在其出生时就已由其父辈的户籍身份所决定，而子辈成年后的户籍身份，虽与子辈的初始户籍身份及其家庭背景密切相关，但也深受其个体努力的影响，所以，子辈成年后的户籍身份与其初始户籍性质的性质又有所不同。基于此，本文进一步分析了子辈初始户籍身份与其成年后的当前户籍身份所界定的不同户籍城乡含义下的中国代际收入流动性的城乡差异。从样本的群体构成看，初始户籍城乡、当前户籍城乡、成年后居住地城乡含义下的城镇与农村样本的区别，主要表现为城乡迁移子辈群体的城乡归属差异。

3. 估计结果。本文首先基于公式（5）与分组回归方法分别估计了非迁移农村子辈、非迁移城镇子辈、户口未变式城乡迁移子辈以及户口改变式城乡迁移子辈的 IGE，估计结果如表 7 所示<sup>①</sup>。关于 IGE 组间差异显著性的检验，根据连玉君和廖俊平（2017）对邹检验（Chow Test）、基于似不相关模型的检验（SUEST）以及费舍尔组合检验（Fisher's Permutation Test）这三种常用的分组回归组间系数差异显著性检验方法的比较分析，邹检验（Chow Test）依赖于较为严格的假设条件，即除目标变量外，其他控制变量的系数不存在组间差异并且两者的扰动项同分布；基于似不相关模型的检验（SUEST）的假设条件相对比较宽松，既允许两组中所有变量的系数都存在差异，也允许两组的干扰项具有不同的分布且彼此相关；费舍尔组合检验（Fisher's Permutation Test）的假设条件虽然最为宽松，但其检验结果可能会因自举抽样次数而波动。基于此，本文选用基于似不相关模型的检验（SUEST）方法检验了 IGE 组间差异的显著性，结果如表 8 所示。

表 7 非迁移农村子辈、非迁移城镇子辈与城乡迁移子辈的代际收入流动性差异估计结果

	子辈城乡归属				城乡与迁移	
	非迁移 农村子辈	非迁移 城镇子辈	户口未变式 城乡迁移子辈	户口改变式 城乡迁移子辈	非迁移 城乡子辈	城乡 迁移子辈
IGE	0.719*** (0.065)	0.445*** (0.106)	0.523*** (0.075)	0.473*** (0.108)	0.667*** (0.057)	0.502*** (0.064)

<sup>①</sup>限于篇幅未列出详细的估计结果，感兴趣的读者可向作者索取。

样本量	3802	776	2345	736	4578	3081
-----	------	-----	------	-----	------	------

注：\*\*\*表示 1% 的显著性水平；括号内的数字为聚类标准误。

表 8 IGE 组间差异显著性的检验结果

IGE 组间差异显著性检验的原假设	SUEST P 值
非迁移农村子辈的 IGE=非迁移城镇子辈的 IGE	0.028
非迁移城乡子辈的 IGE=城乡迁移子辈的 IGE	0.044
非迁移农村子辈的 IGE=户口未变式城乡迁移子辈的 IGE	0.041
非迁移农村子辈的 IGE=户口改变式城乡迁移子辈的 IGE	0.048
非迁移农村子辈的 IGE=城乡迁移子辈的 IGE	0.013
非迁移城镇子辈的 IGE=户口未变式城乡迁移子辈的 IGE	0.547
非迁移城镇子辈的 IGE=户口改变式城乡迁移子辈的 IGE	0.855
非迁移城镇子辈的 IGE=城乡迁移子辈的 IGE	0.646
户口未变式城乡迁移子辈的 IGE=户口改变式城乡迁移子辈的 IGE	0.689

根据表 7 和表 8 所示的分析结果，在非迁移农村子辈、非迁移城镇子辈与城乡迁移子辈这一城乡划分下，本文基于出生于 1970—1995 年的世代群体样本估计得到的中国代际收入流动性城乡差异，具有如下特征：①如果以非迁移农村子辈与非迁移城镇子辈分别代表中国的农村与城镇，那么，本文的估计结果显示，中国农村的代际收入流动性在 5% 显著性水平上低于城镇；②非迁移农村子辈的代际收入流动性在 5% 显著性水平上低于城乡迁移子辈，而非迁移城镇子辈的代际收入流动性与城乡迁移子辈则无显著差异，无论后者是否获得了城镇户口；③非迁移城乡子辈与城乡迁移子辈之间的代际收入流动性的显著差异（在 5% 显著性水平上），主要源于非迁移农村子辈与城乡迁移子辈之间的代际收入流动性差异。

关于城乡迁移子辈的代际收入流动性显著高于非迁移农村子辈的原因，一方面可能源于城镇地区与农村地区的机会结构差异，个体由农村迁移到城镇面临着相对更好的职业机会，因而城乡迁移子辈可能会获得相对较高的代际收入流动性；另一方面也可能是个体由农村向城镇的迁移，使其可以借助城镇地区人力资本外部性产生的学习效应或通过“干中学”，提高自身的技能或非认知能力（魏东霞和陆铭，2021），或增加新的社会关系，从而改变其原有的代际收入传递机制（孙三百等，2012），进而提高了其代际收入流动性。需要注意的是，非迁移农村子辈与城乡迁移子辈的代际收入流动性差异，不能被视为城乡机会结构差异或城乡迁移行为所带来的人力资本与社会资本增加对代际收入流动性的因果效应，因为个体的城乡迁移行为可能具有自选择性，如果具有相对较高代际流动性的个体更有可能由农村迁往城镇，那么，城乡迁移子辈可能本身就具有相对较高的代际收入流动性。关于非迁移城镇子辈与城乡迁移子辈的代际收入流动性无显著差异的原因，可能是因为由农村迁往城镇的城乡迁移者占据了城镇地区较低层级的收入机会，因而也增加了非迁移城镇子辈的向上流动机会，所以，虽然城乡迁移子辈的代际收入流动性显著高于非迁移农村子辈，但与非迁移城镇子辈无显著差异。

本文同样利用分组回归方法分别估计了不同城乡含义下的城乡 IGE<sup>①</sup>，如表 9 所示。根据表 9 所示的估计结果，不同城乡含义下中国代际收入流动性的城乡差异呈现出以下几个方面的特征：①无论在户籍城乡含义下还是在区域城乡含义下，中国农村的代际收入流动性都至少在 10%的水平下显著低于城镇；②不同城乡含义下中国代际收入流动性的城乡差异程度不同，总体而言，户籍城乡含义下中国代际收入流动性的城乡差异大于区域城乡含义下中国代际收入流动性的城乡差异；③表 7 的估计结果显示，非迁移农村子辈与非迁移城镇子辈之间的 IGE 差异约为 0.274，根据表 9 的估计结果，随着城乡迁移子辈被归属于城镇或农村，代际收入流动性城乡差异的程度也发生改变（变小）。基于此，考虑到无论户籍城乡含义还是区域城乡含义下，样本中非迁移子辈与城乡迁移子辈的相对占比在一定程度上会受样本可得性的影响，因而这两类城乡含义下的代际收入流动性城乡差异的估计结果，可能会因样本差异而不同。因此，本文拟在非迁移农村子辈（代表农村）、非迁移城镇子辈（代表城镇）与城乡迁移子辈这一城乡划分下分析中国代际收入流动性城乡差异的时间趋势。

表 9 不同城乡含义下代际收入流动城乡差异的估计结果

城乡划分依据	城乡	城乡群体含义	IGE	IGE 城乡差异	SUEST P 值
初始户口	农村	非迁移农村子辈+所有城乡迁移子辈	0.649*** (0.047)	0.204	0.077
	城镇	非迁移城镇子辈	0.445*** (0.106)		
当前户口	农村	非迁移农村子辈+户口未变式城乡迁移子辈	0.645*** (0.051)	0.188	0.036
	城镇	非迁移城镇子辈+户口改变式城乡迁移子辈	0.457*** (0.077)		
当前居住地	农村	非迁移农村子辈	0.719*** (0.065)	0.176	0.033
	城镇	非迁移城镇子辈+所有城乡迁移子辈	0.543*** (0.055)		

注：\*\*\*表示 1%的显著性水平；括号内的数字为聚类标准误。

## （二）代际收入流动性城乡差异的时间趋势

为分析世代视角下中国代际收入流动性城乡差异的长期变动趋势，本文借鉴了滚动回归选取子样本的方法<sup>②</sup>，首先将出生于 1970—1973 年的子辈样本划分为一组，然后将其作为固定起始点<sup>③</sup>，之后大致按照逐年递增的方式对子辈样本进行如下分组：1970—1973 年、1970—1974 年、1970—1975 年……

<sup>①</sup>本文还估计了代际收入排序关联系数（Intergenerational Rank Association, IRA）这一测度指标，该指标与 IGE 一样，取值越大，意味着代际收入流动性越低。本文基于 IRA 得到的中国代际收入流动性城乡差异的结构特征，与 IGE 基本一致。关于 IRA 的估计结果，限于篇幅未列出详细信息，感兴趣的读者可向作者索取。

<sup>②</sup>严格的滚动回归（rolling regression）方法适用于时间序列或面板数据，不适用于本文的数据。本文只是借鉴了滚动回归通过滚动方式选取子样本的方法，即固定起始值，然后按照某一规律来递增子样本的观测值个数。当然，滚动回归还可以通过其他滚动方式来选取子样本，由于与本研究无关，本文不再赘述。

<sup>③</sup>由于出生于每一年份的非迁移城镇子辈的样本量相对较少，出生于 1970—1972 年的子辈样本与其父辈的代际收入传递性（IGE）不显著，所以，本文将出生于 1970—1973 年的子辈样本划分为一组。事实上，也正是基于同样的原因，本文选择固定起始点的滚动方法来分析中国代际收入流动性城乡差异的时间趋势。

1970—1989年、1970—1990年、1970—1991年、1970—1995年<sup>①</sup>。同时，为了表述方便，本文将上述世代组分别依次记为1~20。此外，根据前文表8的检验结果，本文将主要关注非迁移农村子辈与非迁移城镇子辈、非迁移农村子辈与城乡迁移子辈之间的代际收入流动性差异随世代推移的变动趋势。基于此，本文利用公式(5)估计了每一世代组中非迁移农村子辈、非迁移城镇子辈与城乡迁移子辈的IGE，并据此计算了同一世代组内不同子辈群体之间的IGE差异。图1、图2汇报了世代视角下中国代际收入流动性城乡差异变动趋势的估计结果。

关于非迁移农村子辈与非迁移城镇子辈的IGE差异的变动趋势，如图1所示，出生于1975年之前的非迁移农村子辈的IGE小于出生于1975年之前的非迁移城镇子辈，随着后出生子辈的加入，非迁移农村子辈的IGE都大于非迁移城镇子辈，并且两者的IGE差异总体呈增大趋势。这一结果意味着，随着世代或子辈出生年份的推移，出生于1975年以后的非迁移农村子辈与非迁移城镇子辈之间的代际收入流动差异总体呈扩大趋势。关于非迁移农村子辈与城乡迁移子辈的IGE差异的变动趋势，如图2所示，出生于1973年之前的非迁移农村子辈的IGE小于出生于1973年之前的城乡迁移子辈，随着后出生子辈的加入，非迁移农村子辈的IGE都大于城乡迁移子辈，并且两者的IGE差异呈现出先增大（出生于1973—1983年的子辈）后缓慢下降（出生于1983—1987年的子辈）又缓慢上升（出生于1987—1995年的子辈）趋势。

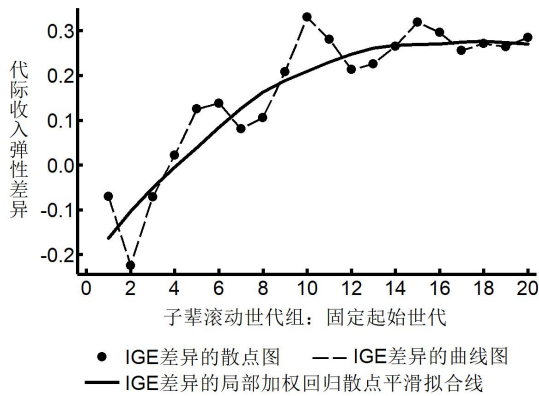


图1 非迁移农村子辈与非迁移城镇子辈的IGE差异

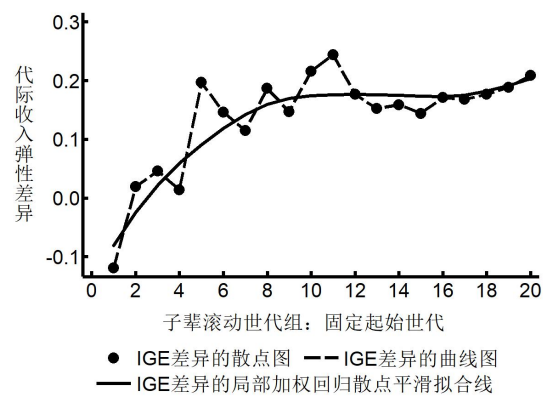


图2 非迁移农村子辈与城乡迁移子辈的IGE差异

由于生命周期偏误的存在，不同世代在同一年份的收入数据，无法用来分析世代视角下非迁移农村子辈、非迁移城镇子辈以及城乡迁移子辈的变动趋势，进而也无法进一步探究非迁移农村子辈与非迁移城镇子辈、非迁移农村子辈与城乡迁移子辈的代际收入流动性差异的变动趋势的来源，所以，本文只能尝试提供一个可能的逻辑解释。根据现代化—理性逻辑，工业化进程一方面可以促进劳动力从第一产业向第二、三产业转移，从而冲击子承父业的社会传承模式，进而引致代际流动性的提高；另一方面还可以推进以绩效原则为基础的社会筛选机制的运行，使得自致性因素成为影响个体经济成就

<sup>①</sup>因为出生于1992—1995年的子辈样本相对较少，为避免样本量过少而引致的估计结果不稳健，本文将出生于这几个年份的子辈样本进行合并处理。



的主要因素，从而提高社会的代际流动性。从现实背景看，随着中国工业化与市场化的发展，城镇地区与农村地区的经济结构差异不断扩大，以绩效原则为基础的社会筛选机制也主要存在于城镇现代化部门，出生年份越晚的世代所面临的城乡产业结构差异与社会筛选机制差异也就越大。由于非迁移农村子辈位于农村地区，而非迁移城镇子辈与城乡迁移子辈都位于城镇地区，所以，随着子辈出生年份的推移，非迁移农村子辈与后两者的代际收入流动性差异总体呈扩大趋势。

## 五、结论与思考

本文利用 CHARLS2018、CHIP1988 与 CHIP1995 数据以及双样本两阶段最小二乘 (TS2SLS) 方法，在控制同住偏误、衰减偏误与生命周期偏误的基础上，基于子辈出生于 1970—1995 年的代际组合样本，分析了中国代际收入流动性的城乡差异及其时间趋势，基本结论如下：①如果以非迁移农村子辈与非迁移城镇子辈分别代表中国的农村与城镇，那么，从静态视角看，中国农村的代际收入流动性显著低于城镇；从动态视角看，出生于 1975 年以前的农村子辈的代际收入流动性高于城镇子辈，但随着子辈出生年份的推移，农村子辈的代际收入流动性逐渐低于城镇子辈，并且，两者的差异总体呈扩大趋势；②虽然初始户口同为农村户口，但从静态视角看，非迁移农村子辈的代际收入流动性显著低于城乡迁移子辈，并且从动态视角看，随着子辈出生年份的推移，两者之间的差异总体呈现出先扩大后缓慢下降而后又缓慢上升的趋势；③当分别以子辈的户口类型与居住地类型为依据界定城乡含义时，户籍城乡与区域城乡含义下的中国农村代际收入流动性都显著低于城镇，但两类城乡含义下的代际收入流动性城乡差异的程度并不相同，并且，从方法方面来看，该差异程度的估计结果与给定城乡含义下的农村或城镇样本中的非迁移子辈样本与城乡迁移子辈样本的相对占比密切相关。

基于本文的研究结果，改善代际收入流动性的社会经济政策可以考虑以下两个方面。第一，推动农村地区的产业转型或升级，改善农村地区的职业结构，增加后出生的非迁移农村子辈所面临的收入机会。第二，完善城乡迁移的相关制度安排，促进后出生的农村户口子辈的迁移与流动。此外，本文关于中国代际收入流动性城乡差异及其时间趋势的相关结论，基于出生于 1970—1995 年的子辈样本估计得到，如果由样本推及总体，原则上该结论只适用于中国出生于 1970—1995 年的群体。最后，本文对中国代际收入流动性城乡差异的分析侧重于对其相关特征的解读，基本上没有涉及内在机制的解释，这也是本文最大的不足之处。

### 参考文献

- 1.陈杰、苏群、周宁，2016：《农村居民代际收入流动性及传递机制分析》，《中国农村经济》第3期，第36-53页。
- 2.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，第115-131页。
- 3.穆尔、诺茨，2017：《统计学的世界》，郑磊译，北京：中信出版社，第543页。
- 4.韩军辉，2010：《机会不等与收入不均：城乡家庭收入的代际流动》，《华南农业大学学报（社会科学版）》第3期，第72-78页、第142页。
- 5.胡洪曙、元寿伟，2014：《中国居民家庭收入分配的收入代际流动性》，《中南财经政法大学学报》第2期，第20-29页。

- 6.胡霞、李文杰, 2016: 《中国城乡收入流动性与收入不平等——基于CHNS 九省调查数据的经验分析》, 《中国农村经济》第8期, 第15-27页。
- 7.江求川, 2017: 《中国代际收入流动性估计: 基于随机系数模型》, 《南方经济》第5期, 第66-82页。
- 8.李路路, 2019: 《社会分层与社会流动》, 北京: 中国人民大学出版社, 第98-101页。
- 9.连玉君、廖俊平, 2017: 《如何检验分组回归后的组间系数差异?》, 《郑州航空工业管理学院学报》第6期, 第97-109页。
- 10.刘精明, 2006: 《市场化与国家规制——转型期城镇劳动力市场中的收入分配》, 《中国社会科学》第5期, 第110-124页、第206页。
- 11.孙三百、黄薇、洪俊杰, 2012: 《劳动力自由迁移为何如此重要? ——基于代际收入流动的视角》, 《经济研究》第5期, 第147-159页。
- 12.魏东霞、陆铭, 2021: 《早进城的回报: 农村移民的城市经历和就业表现》, 《经济研究》第12期, 第168-186页。
- 13.徐林清, 2006: 《中国劳动力市场分割问题研究》, 北京: 经济科学出版社, 第4页。
- 14.许庆红, 2017: 《中国城市劳动力市场分割与代际流动(1978-2010)》, 北京: 社会科学文献出版社, 第153页。
- 15.徐晓红, 2015: 《中国城乡居民收入差距代际传递变动趋势: 2002-2012》, 《中国工业经济》第3期, 第5-17页。
- 16.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, 第60-76页。
- 17.Björklund, A. and M. Jäntti ,1997, “Intergenerational Income Mobility in Sweden Compared to the United States”, *American Economic Review*, 87(5): 1009-1018.
- 18.Böhlmark, A. and M. 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): 879-896.
- 19.Haider, S., and G. Solon ,2006, “Life-Cycle Variation in the Association between Current and Lifetime Earnings”, *American Economic Review*, 96(4): 1308-1320.
- 20.Inoue, A., and G. Solon ,2010, “Two-sample instrumental variables estimators”, *The Review of Economics and Statistics*, 92(3): 557-561.
- 21.Jerrim, J., A. Choi, and R.S. Rodriguez, 2016, “Two-Sample Two-Stage Least Squares (TSTSLS) estimates of earnings mobility: how consistent are they?” *Survey Research Methods*, 10(2): 85-102.
22. Lefranc, A. and A. Trannoy, 2005, “Intergenerational Earnings Mobility in France: Is France More Mobile than the US?” *Annals of Economics and Statistics*, 2005(78): 57-77.
- 23.Mayer, S.E. and L.M. Lopoo,2005, “Has the Intergenerational Transmission of Economic Status Changed?”, *Journal of Human Resources*, 40(1): 169-185.
- 24.Nicoletti, C. and J.F. Ermisch ,2007, “Intergenerational Earnings Mobility: Changes across Cohorts in Britain”, *B.E. Journal of Economic Analysis and Policy*,7(2): 1-36(Article 9).
- 25.Piraino, P.,2007, “Comparable estimates of intergenerational income mobility in Italy”, *B.E. Journal of Economic Analysis and Policy*,7(2):1-25(Article 1).

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

(作者单位：中国人民大学经济学院)

(责任编辑：陈静怡)

## **Intergenerational Income Mobility in China: Urban-rural Differences and Their Time Trends**

HU Xia LI Wenjie

**Abstract:** This article analyzes the urban-rural differences in China's intergenerational income mobility and their time trends as well as its structural characteristics under different definitions of urban and rural China with the micro data of CHARLS2018, CHIP1988 and 1995 and the Two-sample Two-stage Least Squares (TS2SLS) method. The samples used in this study are the non-migrant urban children group, non-migrant rural children group, and rural-urban migration children group, who were all born between 1970 and 1995. The results of this study show that, firstly, from the perspective of static overall level, the intergenerational income mobility of China's non migrant rural children is significantly lower than that of non-migrant urban children. The intergenerational income mobility of China's rural children under the meaning of registered residence urban and rural areas and regional urban and rural areas is significantly lower than that of cities and towns, but there are differences in degree. Secondly, from the perspective of the dynamic time trend, the intergenerational income mobility of non-migrant rural children born before 1975 is higher than that of non-migrant urban children. However, with the passage of their birth years, the intergenerational income mobility of non-migrant rural children is gradually lower than that of non-migrant urban children, and the difference between the two is generally expanding. Thirdly, although non-migrant rural children and rural-urban migration children were both born with rural Hukou, the intergenerational income mobility of the latter born after 1973 is significantly higher than that of the former. With the passage of their birth years, the intergenerational income mobility difference between the two shows a trend of first increasing, then slowly decreasing, and then slowly increasing. These results imply that although there is a certain degree of opportunity inequality in both urban and rural areas in China, the degree of opportunity inequality in rural areas is relatively higher, and this relative opportunity inequality between urban and rural areas tends to expand with the passage of generations from 1970 to 1995.

**Keywords:** Intergenerational Income Mobility; Rural-urban Migration; Rural-urban Difference; Time Trend