

# 机会公平视角的共同富裕\*

——来自低收入群体的实证研究

史新杰 李实 陈天之 方师乐

内容提要: 共同富裕是社会主义的本质要求,而机会不平等是实现全体人民共同富裕的主要障碍之一。本文利用CHIP 2018年的最新数据,通过引入机会不平等的分析框架并加以拓展,从学理上对机会的缺失如何影响低收入群体收入跃升进行系统分析。研究发现,个体是否陷入低收入陷阱很大程度上取决于个体出生即很难改变的机会因素。基于参数估计方法和机器学习方法,本文测算发现与个体低收入状态相关的机会不平等系数为0.310(参数估计结果)至0.336(机器学习结果),这意味着超过30%的低收入状态与机会因素差异有关。本文的渠道分析进一步解释了其中的传导机制,发现机会因素会通过影响个体教育和就业选择来影响低收入陷阱的发生。本文为如何从机会公平的视角认识共同富裕问题提供了重要依据。

关键词: 机会公平 CHIP 2018 共同富裕 机器学习

## 一、引言

贫困是全球性的发展议题,中国曾经的贫困人口基数庞大,1978年贫困发生率达到97.5%。但是经过40多年的改革开放,中国在实现经济飞速发展的同时,绝对贫困人口数量也大幅减少,成为世界范围内减贫的主要推动力(汪三贵和胡骏,2020)。随着2020年习近平总书记在全国脱贫攻坚总结表彰大会上宣布“我国脱贫攻坚战取得了全面胜利”,党的十九届五中全会进一步强调要“扎实推动共同富裕”,并将“全体人民共同富裕取得更为明显的实质性进展”作为2035年基本实现社会主义现代化的重要目标之一。

目前针对共同富裕问题的相关研究主要着眼于理论层面,重点关注两个问题:第一,共同富裕的内涵和度量方法。学界对于其内涵基本达成两个共识,首先,共同富裕应该包括“共享与发展”两个维度,一方面要继续发展,做大“蛋糕”,即通过经济社会高质量发展,达到更高层次的总体富裕;另一方面要实现更大程度的共享,分好“蛋糕”,包括发展成果的共享,发展机会的共享,各种公共服务的共享等(李实和朱梦冰,2022)。其次,共享并不是一味的均等,共同富裕应该是普遍富裕基础上的差别富裕,不是均等富裕、同步富裕(李实,2021;陈丽君等,2021;刘培林等,2021)。基于这两个维度,万海远和陈基平(2021)设计了一套具体的量化方法;而陈丽君等(2021)则增加了对于收入之外的其他指标的度量。第二,共同富裕实现路径。目前对于这个问题的认识,主要都是基于收入分配制度的讨论展开(刘国光,2011;万广华和吴万宗,2019;李实和朱梦冰,2022),其中低

\* 史新杰,浙江大学中国农村发展研究院、公共管理学院,邮政编码:310058,电子信箱:xjshi1990@zju.edu.cn;李实(通讯作者),浙江大学共享与发展研究院、公共管理学院,邮政编码:310058,电子信箱:lishi9@zju.edu.cn;陈天之,澳大利亚国立大学,邮政编码:2601,电子信箱:chtz1993@gmail.com;方师乐,浙江工商大学经济学院,邮政编码:310018,电子信箱:fangshile@zjgsu.edu.cn。本研究得到国家自然科学基金项目(72003170,71803031)、国家社会科学基金重大项目(18ZDA080,21&ZD091)和中央高校基本科研业务费专项资金资助。作者感谢匿名审稿人的宝贵意见,文责自负。

收入群体的收入跃升是关注的重点,比如厉以宁(2002)和李实(2021)等都认为提高低收入者收入水平是实现共同富裕的重要路径。由此可见,要实现更高质量的发展(即进一步做大“蛋糕”),如何激发低收入群体的内在潜力促进其收入提升是一个重要路径;而要实现共享(分好“蛋糕”),如何通过制度设计让低收入群体拥有更加公平的发展机会是一个重要维度。基于此,本文以低收入群体收入跃升作为切入点,从机会公平的视角来探讨共同富裕中的一个关键问题:个体陷入低收入状态是否以及在多大程度上受到“机会”(那些出生即决定,在一定程度上个体无法掌控的因素)差异的影响。

本文可能的边际贡献体现在以下两点:第一,从研究视角来看,已有关于共同富裕的文献大多认为提高低收入者收入水平是实现共同富裕的重要路径,但鲜有研究进一步分析如何提高低收入者收入水平。范从来(2017)提出的“益贫式增长”认为机会平等是关键,实现共同富裕要关注机会平等,强调经济增长给穷人带来的收入增长率要大于平均增长率。但对于什么是机会(不)平等,机会差异在多大程度上会影响低收入者收入水平和共同富裕的实现等问题还鲜有关注。本文在这些理论基础上,进一步从实证层面揭示要实现低收入群体收入跃升,机会均等是一个重要前提。本文发现,机会缺失是个体陷入低收入陷阱的根源之一,超过30%的低收入状态与机会因素差异有关。由此本文认为,机会均等是共同富裕的一个重要内涵,是衡量“共享”的一个重要维度,是在长期范围内从根本上实现共同富裕的重要挑战。除了关注收入的总体不平等(基尼系数),更应该关注收入的机会不平等。本文的相关实证研究为如何通过机会公平政策促进实现共同富裕提供了科学证据,也是对现有文献的有力补充。第二,从研究方法上来看,本文在机会不平等的测算框架上做了一些拓展,探讨了当结果变量为二元变量时如何利用参数估计和机器学习等方法进行测度,在已有文献基础上系统地比较了不同方法的有效性,以及不同维度结果变量(二元变量和连续变量)的差异性,丰富了机会不平等领域相关研究的工具体系。对二元结果变量进行机会不平等测度在国内文献中讨论还很少,但是在数据相对有限的情况下其应用前景非常广泛。本文同时引入了机器学习方法,该方法的分析对象与参数估计相同,基于机会变量将个体划分为若干组别,进而估计出结果变量(本文为低收入状态)的反事实分布,从而计算机会不平等系数。不同之处在于机器学习方法可以基于算法来进行决策,从而选择最优的机会变量和类别划分方式,在一定程度上能够解决遗漏机会(环境)变量的问题,并且缓解机会变量之间的相互关系可能影响模型估计的问题。从经济学意义来看,机器学习侧重于预测效果的准确性,利用机器学习方法来估计机会不平等系数在一定程度上可以减少上述两种偏误,同时对于将因果关系的探讨纳入机会不平等分析框架具有一定的启示意义。但是基于回归的参数估计方法也具有一些比较优势,除了计算过程较为简便外,还可以进一步进行渠道分析。不少研究指出,在经济学分析中应该根据具体问题尝试将机器学习和计量方法结合使用(黄乃静和于明哲,2018;王芳等,2020)。本文综合使用两种方法进行对比分析,提供了一个较好的研究范式。本文的探索也从侧面反映了在经济学研究中应该警惕“唯定量倾向”和“唯方法论”,<sup>①</sup>而是要扎根中国经济学研究体系,以问题为导向选择合适的方法。

下文安排如下:第二部分回顾了机会不平等的相关研究,厘清机会不平等的定义,并从学理上分析了机会因素可能对经济社会结果产生影响的渠道;第三部分进一步介绍低收入群体机会不平等的测算方法,包括如何基于相关机会因素,运用参数估计和机器学习方法测算机会不平等系数以及如何从实证上分析机会不平等的传递渠道;第四部分为数据和描述性统计;第五部分展示了相应的实证结果并进行了延伸探讨;最后一部分总结全文并提出相应的政策启示。

<sup>①</sup> 《破除“唯定量倾向”为构建中国特色经济学而共同努力——〈经济研究〉关于稿件写作要求的几点说明》,《经济研究》第4期封底。

## 二、文献回顾

### (一) 机会不平等: 概念与应用

机会不平等的概念最早在哲学领域开始被关注, Roemer(1998) 将其正式引入经济学分析范式中。他将影响个人经济结果(比如收入)的因素分为两大类:一类是个人的“努力”(选择),诸如一个人付出多少精力在工作和学习上,这类因素是个人可以控制的,由此造成的收入不平等被称为“努力不平等”;另一类因素是从出生就无法掌控的“环境”因素,诸如性别、家庭背景和出生地等,由此造成的收入不平等则被称为“机会不平等”。后者更多地成为政策关注的焦点。

值得注意的是,这里所提及的“机会”与一般意义上所认为的“机会”有所差异,前者更多地是指“机会因素”,而后者更多地是指由这些“机会因素”导致的“机会事实”,比如受教育机会、获得高收入的机会等。由此可以衍生出两类计算机会不平等系数的方法:一类是事前法,主要基于机会因素的信息进行测度;另一类是事后法,需要借助相关努力因素的信息来进行测度,目前大部分的文献主要以事前法测度为主。在事前法的框架下,又可以使用参数估计和非参数测度,其中参数估计是目前文献中主要使用的方法。典型的研究如 Bourguignon et al. (2007) 通过估计收入关于机会和努力变量的方程,基于5个机会变量的反事实收入分布,计算了巴西的收入机会不平等。

基于以上概念界定和方法,目前国际上已经有大量相关研究,这些研究几乎涉及到全球的大部分地区,既包括发展中国家,也包括大部分发达国家。从研究的焦点来看,主要涉及收入和消费、教育以及健康三方面,其中以收入和消费机会不平等的研究最多。目前关于中国机会不平等的研究相对较少,但是近年来关注不断增多。从国际文献来看,只有少数论文有所涉及,主要集中在收入和消费领域。与 Golley et al. (2019) 和 Zhang & Eriksson(2010) 关注收入领域不同, Shi(2019) 使用中国综合社会调查数据(CGSS) 利用参数估计方法探讨了中国能源消费的机会不平等。另外, Golley & Kong(2018) 使用中国家庭追踪数据(CFPS), 讨论教育的机会不平等问题, 研究发现机会差异也会显著影响个体受教育水平。

从国内文献来看,也有相关研究开始使用经济学框架探讨中国收入分配中的机会不平等问题。比如江求川等(2014) 较早对城市居民的机会不平等进行了探讨,刘波等(2015) 较早对中国整体机会不平等进行了测算,陈东和黄旭锋(2015) 从代际转移视角分析了机会不平等在多大程度上影响了收入不平等。最新的测算包括宋扬(2017) 以及李莹和吕光明(2019) 等。其中有两篇文章具有代表性,龚锋等(2017) 探讨了努力的差异如何影响中国机会不平等的估计,发现当全体居民的努力提高之后,由外部环境差异导致的收入分配机会不平等显著降低;史新杰等(2018) 利用CGSS数据系统估算了中国目前的机会不平等系数,发现系数为35.7%,其中年轻个体面临的机会不平等小于年老个体,农村地区高于城市地区。其次,除少数研究外(刘波等,2020),国内文献对于教育和健康机会不平等系数的测算还缺少关注。

### (二) 机会因素及其传递渠道

目前已有较多文献关注贫困或者低收入人群问题(沈扬扬等,2018;陈宗胜等,2013;章元等,2013),但较少探讨“机会因素”在其中的重要作用,也鲜有研究厘清“机会因素”如何通过人力资本的传导进而导致个体陷入低收入状态。要分析这一问题,需要基于相关文献,结合中国研究的情境,明确哪些机会因素是重要的。

首先,户口是重点需要考虑的“机会因素”之一。较多文献探讨了户口对个体收入可能产生的影响,比如吴晓刚和张卓妮(2014) 比较了农民工和城镇职工的同工不同酬现象,发现农民工收入偏低的主要原因在于以户口为基础的职业隔离,而非劳动力市场中的直接歧视。郑冰岛和吴晓刚

(2013) 则划分了两种户口流动方式,即高度选择性的流动(比如城市升学、就业等)和政策安置性的流动(比如土地征用),发现非农户口对前者的收入效应远远大于后者。这说明了户口一定程度上作为一种机会因素,与个体的收入状态息息相关。在史新杰等(2018)、李莹和吕光明(2019)、宋扬(2017)等研究中,也大多将户口作为影响结果差异的重要机会因素。

更为重要的是,户口的差异同样也会影响个体人力资本的积累,进而可能作为一个传导机制影响低收入状态。比如,胡安宁(2014)在研究教育对居民健康影响时发现由于户口隔离造成的城乡义务教育水平的差异导致了义务教育对农民居民的健康没有显著影响,而对于城市居民却有显著影响。这种异质性表明,尽管自1986年起我国就开始推行义务教育,但是农村居民需要更好的接受高质量义务教育的机会才能发挥其健康促进效应。除了义务教育外,李春玲(2014)发现,针对“80后”群体而言,城市居民接受高级中等教育的机会和上大学的机会分别是农村居民的4.7倍和4倍,这说明随着1999年高考扩招,城乡之间的教育不平等反而上升了。马超等(2018)进一步揭示了由于户籍隔阂,农民工群体面临更大的医疗保健方面的机会不平等,同时他们的事后估计方法证实了城乡医保统筹政策并没有缓解这种机会不平等。另外,相关研究也发现,户口带来的人力资本的差异,会进一步影响收入差距,比如赵西亮(2017)通过把有过“农转非”经历的个体从非农户口组中剔除来重新构建农业户口组,发现教育对收入提升效应的城乡差距显著缩小。由此可见,户口是农村居民无法有效利用其所受教育提升收入的重要因素。

其次,有关家庭背景作为机会要素的研究,大多探讨了父母教育和职业状态等因素带来的个体结果差异。李春玲(2014)揭示了家庭背景对于个体升学状态的影响,具体而言父亲多接受1年教育,其子女上大学的机会提升22%;同时对于部分群体而言,管理人员子女上大学的机会是农民子女的2.4倍,这说明来自优势家庭的子女有更多机会获得更多的教育。同时,也有研究探讨了家庭背景对于子女职业状态的影响,比如Jia et al.(2021)利用中国综合社会调查(CGSS)数据发现,父母是企业家或者政府官员的人更可能拥有注册企业,且这一现象在政府干预度更高的地区更为明显。这些研究表明家庭背景会影响个体的教育和职业状态,进而影响个体收入水平以及是否能够实现收入跃升。这一点也在Li et al.(2012)中得到印证,该文利用2010年中国大学生调查的数据进行分析,发现拥有一个干部家长会使得子女收入溢价达到14.5%。

另外,性别和出生地也是机会不平等研究中考虑的重要因素。其中性别作为个体出生即决定的机会因素,在中国劳动力市场中扮演重要角色。李实等(2014)利用中国居民收入调查数据分析了城镇工资的性别差异,发现1995—2007年间工资的性别差异显著扩大,年纪轻、学历低、职业差、行业差的女性职工尤其面临机会弱势;但是后续研究发现这一扩大趋势在2007—2013年间有所下降(罗楚亮等2019)。有关中国机会不平等的研究也大多把出生地或者居住地作为重要的机会要素之一,并且大多揭示了出生在弱势地区的女性在收入水平上会面临更大的机会弱势;部分文献也将年龄和民族等因素纳入到机会不平等的分析框架中(史新杰等2018; Yang et al., 2021)。

这些研究反映了户口、家庭背景、性别和出生地等机会因素对于个体收入等结果的影响,同时也揭示了一种可能的影响个体是否陷入低收入状态的传递渠道,即通过影响个人教育和职业选择发生作用。相关的研究也证实了教育(程名望等2014)和职业选择(周京奎等2020)对个体低收入或者贫困状态的影响。

与已有机会不平等领域研究相比,本文主要有三点不同:第一,现有文献更多关注收入机会不平等本身,而本文从低收入人口收入跃升的视角入手,将机会不平等置于共同富裕的背景下进行探讨,延展了机会不平等研究的理论内涵和实践意义。同时本文不仅仅关注收入机会不平等,也结合教育机会不平等对两者关系进行了深度论证。第二,目前国内对于机会不平等的测算大多以连续

变量为主(比如收入),方法上主要以参数估计为主。然而纵观国内外文献还没有针对低收入群体机会差异的研究,即机会因素如何影响个体是否陷入低收入陷阱。本文则将机会不平等的研究框架拓展至对于二元变量的探讨,以低收入群体作为切入点,系统考察了中国的机会不平等现状,即机会因素在多大程度上导致其陷入低收入状态。同时在参数估计基础上,引入了机器学习方法,为机会不平等相关研究的因果分析提供了可能性。第三,本文探讨了机会差异如何通过教育和职业选择发挥作用进而影响低收入状态。本文将 Palomino et al. (2019) 的机制分析方法引入到机会不平等测算的统一框架中,能够基于教育和职业选择两者的时间先后顺序(一般观点认为教育先于职业发挥作用),清晰识别出其在低收入状态发生过程中起到的作用。

### 三、低收入状态机会不平等测度方法

#### (一) 机会不平等测度

根据 Roemer(1998) 提出的框架,假设有一组有限个体  $i \in \{1, \dots, N\}$  其低收入状态为  $Y = \{y_1, \dots, y_i, \dots, y_N\}$ 。个体  $i$  的低收入状态由两类因素决定,一类是个体一定程度上无法掌控的机会(环境)要素  $\Omega_i = \{C_i^1, \dots, C_i^p, \dots, C_i^k\}$ ,另一类是个体可以掌控的努力要素  $\Theta_i = \{E_i^1, \dots, E_i^p, \dots, E_i^k\}$ 。因此个体是否处于低收入状态的决策方程  $g: \Omega \times \Theta \rightarrow R_+$  可以由下式表达:

$$y_i = g(\Omega_i, \Theta_i) \quad (1)$$

其中,每一个机会变量  $C^k \in \Omega$  有  $X^k$  个可能的取值,每个取值记为  $x^k$ 。据此可以将全部个体划分为若干个机会组别(Type)  $T = (t_1, \dots, t_l, \dots, t_L)$ ,如果  $x_i^k = x_j^k \forall C^k \in \Omega$ ,那么个体  $i$  和  $j$  属于同一机会组别;如果  $\exists C^k \in \Omega: x_i^k \neq x_j^k$ ,则个体  $i$  和个体  $j$  不属于同一机会组别。这意味着属于同一机会组别的任意个体拥有相同的机会要素。因此总共可以划分的机会组别个数为:  $\prod_{k=1}^K X^k$ 。

#### 1. 参数估计方法

根据 Bourguignon et al. (2007) 和 Ferreira & Gignoux(2011) 在研究收入分配中机会不平等的相关思路,本文首先使用参数估计测算机会不平等系数。由于个人的努力程度受到机会因素的影响,(1)式可以采用如下的简化形式:

$$y_i = \sum_{k=1}^K \alpha_k C_i^k + \varepsilon_i \quad (2)$$

其中  $y_i$  代表个体  $i$  的收入水平,  $C_i^k$  代表个体  $i$  的第  $k$  个机会要素,  $\varepsilon_i$  代表残差项。由于每个机会组别内部的个体具有相同的机会要素,基于该回归得到的预测分布  $\tilde{y}_i = \sum_{k=1}^K \hat{\alpha}_k C_i^k$  中每一个机会组别内部的个体拥有相同的收入水平,因此可以视为由原分布平滑后得到。在本文的分析框架中,主要的结果变量是低收入状态二元变量。本文参照 Paes De Barros et al. (2007) 和 Juárez & Soloaga(2014) 的相关思路对(2)式进行了相应的拓展,得到(3)式:

$$Pr[y_i = 1 | C] = \phi\left(\sum_{k=1}^K \beta_k C_i^k + \mu_i\right) \quad (3)$$

在(3)式中  $y_i$  代表个人是否处于低收入状态,方程右侧的机会变量与(2)式一致,基于该式可以通过 Probit 回归得到个体  $i$  陷入低收入状态概率的平滑分布  $\tilde{y}_i$ ,使得每一个机会组别内部的个体拥有相同的低收入发生概率。通过计算该分布的不平等程度(即机会组别之间的不平等程度)即可以得到机会不平等的绝对估计值:

$$IOP = I(\tilde{y}_i) \quad (4)$$

其中  $I$  代表不平等衡量指标,当变量是二元变量时,现有文献主要采用相异指数(Dissimilarity index)。在针对低收入状态的分析中,本文也主要采用相异指数来衡量低收入群体的机会不平等状况,具体可参考 Paes De Barros et al. (2007)。另外作为对比,本文还使用了泰尔指数(Theil index)对主要结果进行衡量(周国富和陈菡彬,2021)。在最后部分针对收入(连续变量)的相关分

析中 根据 Bourguignon et al. (2007) 的做法 采用 MLD 指数 (Mean Log Deviation)。借助以上方法获得整体的机会不平等系数后 还可以通过 Shapley 方法 (Shorrocks 2013) 对其进行分解<sup>①</sup> 获得每一个机会变量的贡献度。其主要优点是每一个机会因素贡献度的计算都考虑了其加入的顺序 同时所有机会因素贡献度加总为 1 具体操作过程可参考 Golley et al. (2019)。

## 2. 机器学习方法

虽然参数估计方法已经成为机会不平等估计中的主流 但也面临诸多问题。为了解决这些问题 Brunori et al. (2018) 提出使用机器学习中回归树的思路来对样本进行预测 从而计算机会不平等系数。运用树方法进行推断得到结果变量  $y$  (低收入状态) 的预测值 是基于一系列的输入变量  $I = \{I^1, \dots, I^k, \dots, I^K\}$  产生的。具体来说 通过集合  $I$  将总样本分为若干组  $G = \{g_1, \dots, g_m, \dots, g_M\}$  每一个组  $g_m$  内部具有同质性 也被称为回归树中的“叶”或者“末端节点”。这里的输入变量  $I$  事实上对应了上述框架中的机会集  $\Omega_i$  组  $g_m$  对应了机会不平等中机会组别 (Type) 的概念 在每一个组别内部的个体具有一样的机会集。由此 每一个个体的预测结果 (低收入状态) 通过个体所在的组别的平均状态计算所得。针对原样本  $\{y_i\}$  对应地可以得到一个预测的样本  $\{\tilde{y}_i\}$  其中:

$$\tilde{y}_i = \frac{1}{N_m} \sum_{i \in g_m} y_i, \forall i \in g_m, \forall g_m \in G \quad (5)$$

在机器学习的设计中 通过选取合适的回归树类型对样本进行分割最为关键 本文基于 Brunori et al. (2018) 的思路选取条件推断树来进行循环操作。<sup>②</sup> 机器学习方法的引入同时为从因果关系解读机会不平等系数提供了可能。机会变量是个体一定程度上无法掌控的因素 大多先于结果变量 (是否低收入状态) 发生 因此基本不存在反向因果关系的影响。但是 在回归过程中无法获得所有的机会变量信息 比如说个人的智商 (Golley & Kong, 2018) 由此可能产生遗漏变量偏差。另外 如果回归模型过度拟合 甚至也会出现估计所得机会不平等系数大于实际值的情况 (Brunori et al., 2019) 而纳入回归中的机会变量之间的相互关系也会使得机会不平等系数的估计变得更为复杂 (Brunori et al., 2018)。机器学习方法在一定程度上可以解决这些问题 从而有助于强化“机会的差异导致低收入状态发生”的结论。<sup>③</sup> 另外 在稳健性检验中 本文进一步使用了条件推断森林 具体操作过程可以参考 Brunori et al. (2018) 以及李金叶和郝雄磊 (2019)。

### (二) 低收入的传递渠道: 教育与职业

本文进一步参考了 Palomino et al. (2019) 对于收入机会差异的分解框架 识别了教育和职业选择作为最重要的两条途径 在低收入状态发生过程中起到的作用。

首先基于 (3) 式得到的预测分布  $\tilde{y}_i$  对个人教育变量  $E$  做回归 得到下式:

$$\tilde{y}_i = \gamma E_i + v_i \quad (6)$$

基于 (6) 式 可以通过 OLS 回归得到一个新的预测分布  $\{\tilde{y}_i^{C,E} = [\hat{\gamma} E_i]\}$  在该分布中 具有相同受教育程度的个体面临着相同的 (基于机会变量得到的) 低收入状态概率。由此 通过计算该分布的不平等系数  $I(\tilde{y}_i^{C,E})$  可以分析机会因素通过影响个人受教育程度进而影响低收入状态的渠道效应。同时 (6) 式的残差部分  $\tilde{y}_i^{C,E} = [v_i]$  可以解释为机会因素影响低收入状态的其他可能渠道。其分解过程可用下式表达:

① 其他分解方法诸如 RIF 分解 可参考徐舒 (2010)。

② 关于具体执行过程和相应的条件推断树 感兴趣的读者可向作者索取。

③ 目前计量经济学中进行因果识别有个思路 第一种是寻找同质样本 比如双重差分和断点回归的方法都基于这个思路; 另一种是工具变量方法 通过寻找外生变量采用两阶段最小二乘法 (2SLS) 来估计局部平均处理效应 (LATE)。而机器学习更加宽松的设定使得其预测性能远远超过传统计量方法 基于这种预测优势构建处理组的反事实一定程度上可以运用于进行因果推断 (王芳等 2020)。

$$I(\tilde{y}_i) = I(\tilde{y}_i^{C,E}) + I(\tilde{y}_i^{C,P}) \quad (7)$$

上述文献回顾中提到 机会要素的差异除了可能对个人的受教育程度造成影响 ,同时也会限制个人的职业选择 ,同时从影响顺序上来说 ,往往前者优先于后者。因此本文基于(7)式 ,进一步分解出在排除教育渠道后职业渠道所起到的作用。其过程与上述对于教育渠道的分解思路类似 ,主要利用了(7)式的残差部分  $\tilde{y}_i^{C,P}$  对个人职业变量  $O$  再次做回归 ,得到(8)式:

$$\tilde{y}_i^{C,P} = \delta O_i + \tau_i \quad (8)$$

由于在(8)式中 ,其因变量已经剔除了教育渠道的影响 ,因此通过 OLS 回归得到的预测分布  $\{\tilde{y}_i^{C,D/E} = [\delta O_i]\}$  代表了机会变量直接通过职业渠道产生的影响 ,<sup>①</sup>而其残差项代表了除了教育和职业渠道之外的其他影响。由此 ,可以进一步分解为以下三个部分:

$$I(\tilde{y}_i) = I(\tilde{y}_i^{C,E}) + I(\tilde{y}_i^{C,D/E}) + I(\tilde{y}_i^{C,others}) \quad (9)$$

由此可计算出教育渠道的贡献度  $IO_E^{Ratio} = \frac{I(\tilde{y}_i^{C,E})}{I(\tilde{y}_i)}$  以及职业渠道的贡献度  $IO_O^{Ratio} = \frac{I(\tilde{y}_i^{C,D/E})}{I(\tilde{y}_i)}$ 。

#### 四、数据及描述性统计

本文的研究主要基于中国家庭收入调查(CHIP)2018年的最新数据。作为国内最早追踪中国收入分配的数据 ,CHIP从1989年到2014年已经进行了五次入户调查 ,分别收集了1988年、1995年、2002年、2007年和2013年的收支信息 ,以及家庭和个人层面的各类相关信息。作为该项目的第六次调查 ,CHIP2018在抽样上进一步优化了其科学性。首先在省样本上 ,与CHIP2013保持严格一致 ,并在其基础上增加了内蒙古地区的样本代替不能如期实施调查的新疆地区 ,由此增加少数民族住户样本。其次在区、县样本上 ,在CHIP2013的基础上进行了调整 ,进一步缩小区县样本与国家统计局常规调查的大样本之间的差距。区县样本决定后 ,区县内国家统计局常规住户调查的所有住户样本 ,均为CHIP2018的住户样本。本文在整合个人和家庭层面的信息基础上 ,选取了处于劳动力市场活跃期的20—60岁的样本 ,剔除了部分缺失信息的个体 ,最终获得39463个有效样本。本文主要的因变量为个体的低收入状态 ,主要通过个体的最低生活保障或社会救济信息来识别 ,在问卷中涉及到的具体问题为:您享有以下哪种最低生活保障或社会救济?选项包括:(1)农村最低生活保障;(2)城镇最低生活保障;(3)五保供养;(4)贫困(扶贫)救助;(5)其他社会救济;(6)没有。基于这个问题 ,本文将个体低收入状态设置为二元虚拟变量 ,当个体享有(1)—(5)至少一种最低生活保障或者社会救济时取值为1 ,否则取值为0。使用这个指标来识别低收入人群的一个好处是 ,对于最低生活保障和社会救济的目标人群的识别往往不仅仅基于收入 ,而是综合考量了其他经济社会状况 ,因此是一个更加综合化的低收入人群指标。但是也有不少研究指出 ,低保等政策在瞄准的精准度上依旧存在问题(朱梦冰和李实 2017) ,因此本文的稳健性检验基于实际收入的划分重新设定了低收入状态。按照上述定义方式 ,样本中4.9%的个体处于低收入状态。样本个体平均年收入为36676元 ,其中低收入样本个体的年收入(16700元)显著低于非低收入样本个体(37705元)。基于现有文献 ,本文主要使用了户口、性别、民族、年龄组、地域、父母受教育程度和父母职业状况等机会因素。户口为二元变量 ,1代表城市 0代表农村 ,样本中53%的个体为城市户口 ,而该比例在低收入样本中仅为23% ,说明目前的低收入人群主要在农村地区。民族也为虚拟变量 ,1代表汉族 0代表少数民族 ,少数民族占比为6%左右 ,在低收入样本和非低收入样本中较为接近。样本的年龄段被划分为20—30岁、31—40岁、41—50岁和51—60四个组 ,各组占比在20%—30%之间 ,但是老年人在低收入群体中的比例明显高于非低收入群体 ,相反年

<sup>①</sup> 由于教育变量比职业变量先引入到模型 ,所以部分通过职业渠道产生的间接影响会体现在教育渠道中。

轻人在非低收入群体中的比例显著更高,这意味着目前的低收入发生存在着老龄化趋势。对于地域变量的选择,最为理想的是选择个体出生地。本文根据 Singh(2012)的做法采用了调查时的居住信息作为替代,并将其划分为东、中、西和东北四个区域,除了东北地区外,东中西部样本较为均匀。另外,家庭背景也是机会不平等及代际流动研究中关注的因素,本文主要使用父母受教育程度和父母职业状况来代理。在具体处理上,主要选取了父母双方受教育程度和职业相对较好一方的结果,前者分为高中以下(取值为0)和高中及以上(取值为1)两类,后者划分为低技能职业(取值为0)和高技能职业(取值为1)两类。在样本中,高教育程度和高技能职业人数显著少于低教育程度和低技能职业人数,在低收入样本中更是如此。同时这两个变量的缺失值相对较多,本文参考了 Jusot et al.(2013)研究健康机会不平等时的做法在后续回归处理上将缺失值也划分为一类进入回归中。<sup>①</sup>

基于这些机会因素的划分,本文进一步比较了农村和城市不同分样本中低收入发生状况。结果显示,农村的低收入发生率(7.92%)明显高于城市(2.17%),这一点与预期相同。另外,低收入状态发生在男性与女性以及少数民族和汉族之间的差异性不大,但是在地域和年龄上具有较为明显的差异,老年群体相对于年轻群体而言低收入发生率更高,东部地区相较于中西部地区低收入发生率更低。另外对于家庭背景好的个体而言,低收入发生率相对也更低。这些差异性为低收入人群机会差异的存在提供了初步证据。

## 五、主要结果

### (一) 机会不平等系数与分解

基于研究方法部分所使用的实证策略,本文一方面使用了参数方法对低收入群体机会不平等系数进行了估计,另一方面使用机器学习中的条件推断树进行了测算,并对两者的结果进行了对比分析。在参数估计中,本文主要基于(3)式使用了 Probit 模型,以低收入状态为因变量对相关的机会因素做回归。相关结果(篇幅所限未列出)与描述性统计传递的信息一致,比如农村地区的低收入发生概率相对更高,性别和民族的影响相对有限,中西部地区的老龄群体相对更容易进入低收入状态,而家庭背景更好的个体收入跃升的可能性更高。从共同富裕“共享”与“发展”的内涵来看,重点在于提高低收入群体的收入(李实和朱梦冰,2022),这些初步结果揭示了哪些因素可能与陷入低收入状态相关。

表1基于该回归进一步测算了相应的机会不平等系数。按照目前文献对于二元变量机会不平等系数的处理方式,本文主要使用了相异指数作为衡量不平等的指标,同时汇报了泰尔指数的结果作为对比。从参数估计方法的结果来看,基于全样本的机会不平等系数为0.310,意味着即便这些个体付出一定程度的努力,由于机会的弱势使其陷入低收入陷阱的概率依旧高达31%,但这些机会的弱势不是个体可以主观选择,而是在一定程度上随着出生而决定的。另外,本文通过机器学习估算的机会差异系数为0.336,略高于使用参数估计得出的结果。目前主流文献认为参数估计方法得到的机会不平等系数由于遗漏机会变量等原因可能会略小于真实值(Ferreira & Gignoux, 2011)。本文的结果为其提供了相应证据,发现机器学习方法能在一定程度上弱化这种偏差,同时机器学习所得到的结果可以有助于从因果视角解读机会不平等系数,即“目前中国低收入状态发生的差异中有33.6%是由于这些个体一定程度无法掌控的机会因素决定的”。

<sup>①</sup> 同时,在稳健性检验中删除了缺失值,对基本结论没有影响。

表 1 低收入状态机会不平等系数

	相异指数 DI		泰尔指数 Theil	
	参数估计方法	机器学习方法	参数估计方法	机器学习方法
全样本	0.310	0.336	0.263	0.276
户籍				
农村	0.182	0.186	0.109	0.105
城市	0.147	0.130	0.065	0.0489
性别				
女性	0.310	0.336	0.263	0.257
男性	0.311	0.335	0.270	0.264
民族				
少数民族	0.282	0.217	0.262	0.162
汉族	0.315	0.341	0.271	0.278
年龄组				
20—30岁	0.327	0.339	0.304	0.280
31—40岁	0.301	0.272	0.242	0.199
41—50岁	0.288	0.319	0.218	0.216
51—60岁	0.273	0.289	0.188	0.195
地域				
东	0.271	0.274	0.214	0.236
中	0.298	0.298	0.227	0.232
西	0.281	0.281	0.193	0.187
东北	0.245	0.151	0.182	0.0587
父母受教育程度				
高中以下	0.333	0.358	0.296	0.297
高中及以上	0.265	0.287	0.219	0.203
缺失	0.277	0.303	0.216	0.223
父母职业				
低技能职业	0.308	0.331	0.255	0.242
高技能职业	0.375	0.244	0.458	0.186
缺失	0.304	0.327	0.256	0.261

注:本表使用的参数估计和机器学习方法可参考第三部分,因变量(个体低收入状态)和相关机会变量的构建可参考第四部分。

基于所采用的机会变量,本文进一步测算了所有子样本的机会不平等系数。表1结果显示,农村机会不平等系数(0.182)高于城市(0.147),一方面农村和城市子样本系数远远低于全样本系数,说明户籍因素很有可能是产生机会差异的重要原因;另一方面除了城乡差异之外,农村内部的机会差异也需要引起重视。从性别角度来看,男女子样本的机会不平等系数差异较小,且非常接近全样本,这说明目前性别的机会差异可能已经很小。进一步关注民族差异性发现,少数民族的机会不平等系数略小于汉族,这可能与我国对于少数民族群体相应的支持政策相关。从年龄组的异质性来看,年轻群体的机会不平等系数明显高于老年群体,这说明目前年轻一代可能更容易因为机会的弱势陷入低收入陷阱,其中的原因很可能与教育和就业等渠道相关。从地域上看,东部地区的机

会不平等系数小于中西部地区,这也在一定程度上解释了为什么东部地区相对来说对年轻人会更有吸引力,其中的原因可能不仅仅与经济发展程度有关,也与个体的发展机会息息相关。最后,本文重点关注来自不同家庭背景的个体面临的机会差异,从父母受教育的视角来看,父母受教育程度较低的个体其机会不平等系数越大,符合预期;但是从父母职业状况的角度来看,参数估计的结果显示父母职业技能高的个体反而更容易陷入低收入陷阱,其中可能的原因是这部分个体的样本数量非常少,在一定程度上弱化了参数估计方法的有效性,而使用机器学习方法的结果显示父母拥有高技能职业的个体其低收入状态机会不平等系数为 0.244,小于父母拥有低技能职业的个体(0.331)。另外使用机器学习方法对其他子样本(大多数样本量较大)估计时,其结果基本与参数估计方法相近,或者略大于参数估计方法。这说明了传统的参数估计方法得出的机会不平等系数虽然可能会小于真实值,但是其有效性相对较高,不过当样本数量特别少的情况下,机器学习方法可能会更占优势。另外,本文也汇报了泰尔指数的结果,整体上略低于相异指数。

表 2 机会不平等分解

	全样本	农村	城市	女性	男性	少数民族	汉族	20—30岁	31—40岁	41—50岁	51—60岁
总系数	0.310	0.182	0.147	0.310	0.311	0.282	0.315	0.327	0.301	0.288	0.273
户籍	54.076			53.850	53.252	31.281	54.926	55.339	53.184	59.448	70.448
性别	1.219	2.596	0.275			4.692	1.012	0.763	1.027	1.502	4.171
民族	0.222	2.239	1.686	0.042	0.500			0.124	0.093	3.793	1.289
年龄	12.516	30.894	37.642	12.047	12.820	24.695	12.689				
地域	22.450	59.570	30.833	26.286	19.380	27.504	21.726	26.229	31.919	30.396	19.903
PE	6.449	3.375	20.530	5.413	9.186	9.195	6.426	10.717	4.518	3.852	2.946
PO	3.043	1.326	9.035	2.362	4.834	2.634	3.221	6.828	9.036	1.000	1.244
		东	中	西	东北	高中以下	高中及以上	缺失	低技能	高技能	缺失
总系数		0.271	0.298	0.281	0.245	0.333	0.265	0.277	0.308	0.375	0.304
户籍		34.436	72.395	78.406	18.618	53.616	41.680	55.220	54.957	30.227	52.591
性别		4.461	1.058	0.758	4.670	2.192	0.719	4.875	1.462	10.440	2.894
民族		-0.654	0.226	0.624	0.351	0.568	1.459	0.105	0.422	4.676	0.096
年龄		33.737	14.095	11.521	49.683	9.852	25.123	17.079	11.153	34.391	15.471
地域						26.571	27.379	22.231	22.844	14.319	23.655
PE		19.701	6.946	5.464	15.273				9.127	5.947	5.283
PO		7.369	5.090	3.029	11.404	7.129	3.640	0.484			

注: PE 代表父母受教育程度,PO 代表父母职业,Shapley 分解展示了每个机会因素对总系数的贡献度占比(%)。

在上述对于整体机会不平等测算的基础上,本文进一步使用 Shapley 分解估算了每一个机会变量的贡献度。表 2 第(1)列的结果显示,在全样本中,户口对于低收入状态机会差异的贡献度最大,为 54.076%,这也与我国先前的扶贫工作重点为什么在农村相关,同时也在一定程度上印证了表 1 的结果。地域因素对低收入状态机会不平等系数的贡献度为 22.450%,其重要性排在第二,说明区域性的扶持政策可能对走出低收入陷阱具有很大的作用。另外,年龄组和家庭背景(包括父母教育和职业的加总)的贡献在都在 10%左右,相对也较高,结合表 1 的结果,在政策制定中可能要重点关注家庭背景较差的年轻群体。另外,性别和民族要素对机会差异的贡献度很小,分别为 1.219%和 0.222%,与表 1 结果一致。

除此之外,本文也对每一个子样本的机会要素进行了分解,有几点值得注意的信息:(1)在农

村和城市样本中,由于剔除了户口的影响,地域的影响被强化了,而在其他样本中户口都具有非常重要的作用;(2)家庭背景的重要性相对年轻群体而言远远大于年老群体;(3)对于家庭背景较差的个体,户口的影响要远远大于来自更好家庭背景的个体。

本文做了如下稳健性检验:第一,虽然本文按照相关文献的年龄标准选取样本,只保留了20—60岁这个处于劳动力市场活跃阶段的年龄群体,但是可能有少部分学生或者离退休、伤残等人员是没有工资收入的,这在一定程度上可能会影响低收入指标,因此进一步剔除了这类人群,只保留了处于工作状态的样本,表3检验1的结果显示整体机会不平等系数为0.347,略高于上述参数估计的结果,但整体差异不大。第二,由于本文在父母职业和父母受教育程度这两个指标上具有较多的缺失值,因此参考了Jusot et al. (2013)的做法,将缺失值分别作为受教育程度和职业的一个类别进入回归中,但是这样的做法可能会影响整体机会不平等系数的测算,因此检验2、检验3和检验4中,本文分别剔除了父母受教育程度缺失的样本,父母职业缺失的样本以及两者都缺失的样本,并重新对机会差异系数进行测算,结果发现系数整体差异不大,分别为0.344、0.325和0.323。同时上述四个检验对于不同机会变量分解的结果也不存在特别大的差异,户口和地域依旧是最重要的两个要素,年龄和家庭背景也相对重要,而性别和民族则相对贡献度很小。第三,在低保或者社会救助项目的识别中,确实可能出现“精英俘获”现象,比如有些研究证实当家庭中具有较强的政治联结的时候,可能更容易得到社会救助,尽管这些群体可能并不是真正的瞄准对象。由此,本文在稳健性检验中进一步加入了父母的政治面貌(其中至少有一方是党员)作为机会变量进行测度(检验5),结果发现整体机会不平等系数基本保持不变(0.311)。第四,本文参照史新杰等(2018)和雷欣等(2018)的思路,在收入决定方程中,对随机干扰项进行分解,剥离出体现在不同环境类别之间的异方差。在考虑这种异质性后,整体机会不平等系数依旧为0.310,各环境变量的贡献度变化不大。第五,参照Brunori et al. (2018)以及李金叶和郝雄磊(2019)的操作过程,本文使用条件推断森林进行进一步分析,结果发现条件推断森林结果略高于参数估计和条件推断树,为0.354(相异指数)和0.314(泰尔指数)。同时在条件推断树中进行了变量重要性排序,虽然该重要性指标与参数估计的Shapley分解不直接可比,但也反映了不同机会变量的贡献程度。结果显示户口同样是最为重要的因素,与表2结论一致。<sup>①</sup>

本部分为共同富裕为什么要关注机会因素提供了依据。通过延伸上述结果可以发现,要想实现共同富裕,依靠转移支付在短期内提升低收入群体的收入可能只是扬汤止沸,因为个体间的收入差异有31%来源于与个体努力无关的机会性壁垒。这就意味着如果不破除诸如户口、地域、家庭背景和年龄等机会障碍,个体很难通过内源式发展从真正意义上掌握获得高收入的能力。

表3 稳健性检验

	检验1	检验2	检验3	检验4	检验5
机会不平等系数	0.347	0.344	0.325	0.323	0.311
户籍	46.610	49.738	51.081	51.652	53.003
性别	0.751	1.454	1.517	1.494	1.090
民族	0.0930	0.397	0.560	0.578	0.202
年龄	15.072	10.019	11.145	11.244	12.114
地域	26.557	23.185	18.817	18.869	22.316
父母受教育程度	7.296	7.890	10.300	9.503	5.870
父母职业	3.548	7.291	6.514	6.594	2.783
父母政治面貌					2.607

<sup>①</sup> 第四、五点的图表,由于篇幅所限没有在正文列出,感兴趣的读者可以向作者索取。

## (二) 机会不平等渠道分析

上文的分析证实了机会因素确实对低收入状态产生重要影响,然而这种影响是如何发生作用的还未可知。由文献回顾可知,个人教育和职业选择可能在其中发挥着重要作用,同时两个渠道相互关联,从个体生命周期的角度来看,教育资源很大程度上会决定个人在劳动力市场上的表现。由此,本部分基于(6)一(9)式的方法,将教育变量视为前置中介变量,在渠道分析中先引入教育变量再引入职业变量,从而可以在控制教育影响的前提下进一步测算职业途径的作用。具体而言,个体教育变量为个人的文化程度,包括未上过学、小学、初中、高中、职高/技校、中专、大专、大学本科、研究生等9类;个体职业选择采取与父母职业相同的分类方式,划分为低技能和高技能两类。<sup>①</sup>其结果如表4所示。

首先,教育作为先于职业发生的前置变量,是低收入机会差异中极为关键的一个传递渠道。按照方法部分(6)式先将教育变量放入回归,按照(8)式将分离出的残差继续对职业变量回归,进而可以基于(9)式将机会变量对收入的作用分解为教育渠道、(控制教育影响后的)职业渠道以及其他渠道。按上文所述,如果基于相异指数,整体的机会不平等系数为0.31,而表8显示42.237%的影响是通过教育传递的。该结论说明机会因素的差异影响了个体是否能够接受高中教育,进而导致了个体是否陷入低收入陷阱。相较于以往的研究而言,这个数字是非常可观的,比如Palomino et al.(2019)针对欧洲多个国家收入机会不平等的一项研究发现,教育渠道对于整体的机会不平等系数的贡献最高的国家也只是达到了30%左右;李莹和吕光明(2019)针对国内的一项研究发现,教育渠道在中国收入机会不平等的贡献比例在30%—34%之间。从分样本的情况来看,在西部地区的年轻农村女性群体,教育渠道的重要性更大,更应该成为机会扶持关注的焦点。值得注意的是,家庭背景较差的个体其教育传递渠道的比例虽然也很高,但是却低于家庭背景较好的个体,其原因可能在于家庭背景较差的个体其整体的教育水平还相对较低,内部差异性不大。

表4 机会差异的传递渠道:教育和职业

	整体机会差异系数	教育渠道(%)	职业渠道(%)
全样本	0.310	42.237	16.594
户籍			
农村	0.182	36.236	18.142
城市	0.147	33.564	30.805
性别			
女性	0.310	51.333	16.084
男性	0.311	41.918	24.889
民族			
少数民族	0.282	39.226	15.356
汉族	0.315	42.237	16.163
年龄组			
20—30岁	0.327	52.596	9.031
31—40岁	0.301	37.235	13.752
41—50岁	0.288	28.796	4.203
51—60岁	0.273	33.624	13.102

<sup>①</sup> 为了结论的稳健性,本文同时利用受教育年限和原始的职业代码(8类)进行了估计,所得结果没有太大变化。

续表 4

	整体机会差异系数	教育渠道(%)	职业渠道(%)
地域			
东	0.271	40.900	30.281
中	0.298	41.200	12.760
西	0.281	45.579	19.680
东北	0.245	31.947	24.155
父母受教育程度			
高中以下	0.333	41.377	18.179
高中及以上	0.265	49.059	5.759
缺失	0.277	39.027	10.666
父母职业			
低技能职业	0.308	45.487	15.635
高技能职业	0.375	57.189	21.988
缺失	0.304	40.896	15.538

其次,在控制教育影响的基础上,职业渠道的贡献度相对更小一些,<sup>①</sup>在整体样本中为16.594%。这一比例相较于李莹和吕光明(2019)发现的4%—6%以及Palomino et al.(2019)计算得出的1%—8%而言依然较大。另外在分样本中,职业渠道的贡献度呈现出与教育渠道不同的特征,具体表现为城市高于农村,男性高于女性。原因可能在于,出生在城市的群体其职业分化情况相较于农村而言会更大,而男性群体的就业多样性也高于女性。

本文的渠道分析可为从因果视角来解读机会因素和收入状态之间的关系提供一些思考。如前文所述,机会因素一定程度上(31%)导致低收入状态的发生,很大程度上是由于机会差异性影响了个体教育获得性(能否上高中)以及职业选择,进而影响了收入状态。具体而言,13%(31%×42.237%)的影响通过“机会—教育—低收入状态”的渠道产生,5%(31%×16.594%)的影响通过“机会—职业—低收入状态”的渠道产生。该结论为如何提升低收入人群的收入进而推进共同富裕提供依据,即要从源头上关注低收入人群的人力资本积累和劳动力技能提升。如果能阻断机会差异性的负面影响,在一定程度上也能提升机会弱势群体的人力资本,拓宽其就业渠道,进而能够从源头上提升低收入群体的收入水平,相较于短期转移支付而言是更加可持续的发展路径。

### (三) 进一步讨论<sup>②</sup>

本部分基于教育渠道在机会差异传递中的重要作用,进一步分析机会因素能在多大程度上决定个体受教育程度的高低,<sup>③</sup>通过参数估计和机器学习两种方法计算了教育机会不平等系数,其结果显示两种方法获得的系数非常接近,分别为0.297和0.305。这说明30%左右个体受教育程度的不平等来源于机会因素的差异,一定程度上为上文的渠道分析提供了证据,即个人教育获得很大程度上也受制于机会要素,由此才可能对个人的低收入状态产生影响。在此基础上有必要明晰的另一个问题是弱势群体如果顺利脱离低收入陷阱,其收入的增长在多大程度上依旧受到机会因素的桎梏。由此本文进一步测算了收入分配中的机会不平等系数,结果显示收入分配的机会不平等绝对系数为0.278,而相对系数为0.441,这说明如果不从根源上解决机会的差异性问题,即便弱势群体脱离了低收入状态,个体为了进一步增加收入能够努力的空间依旧被机会因素挤压,因为

① 这是因为职业渠道的影响也可能来源于教育的差异,这也是本文模型先加入教育变量,后加入职业变量的原因。

② 由于篇幅所限本部分具体表格没有在正文列出,感兴趣的读者可以向作者索取。

③ 为了更好地与低收入状态对比,将个人教育变量处理成二元变量,1代表高中及以上,0代表高中以下。

44. 10% 的收入差异不是由个体努力可以改变的。而从分样本的机会不平等系数以及各个机会因素的分解结果来看, 性别对于收入分配中的机会不平等具有较大的贡献度( 32. 08%) , 这背后的一个重要原因可能是, 虽然女性在收入增长方面相较男性依旧存在一定的差距, 但是在低保等政策的目标人群识别过程中女性群体( 尤其是农村地区) 由于话语权相对薄弱, 往往更容易出现漏保现象。比如, 朱梦冰和李实( 2017) 基于 CHIP 2013 年数据的研究发现, 农村低保的保准率水平较低。

为了证实这种差异性, 本文进一步基于收入变量和 2017 年全国的贫困标准构建了低收入指标, 并使用参数估计和机器学习方法分别估计了机会不平等系数。结果显示, 其整体机会不平等系数为 0. 321, 与前面使用的低收入指标计算结果几乎没有差异, 佐证了该机会不平等系数的稳健性; 机器学习结果显示为 0. 356, 略高于参数估计结果。但是与收入机会不平等系数类似, 在分解结果中发现, 性别对于基于收入指标的机会不平等系数贡献高达 46. 212% , 由此进一步证实了在低收入识别过程中的性别偏差问题, 这对于相应扶持政策的制定来说, 也具有较大的参考价值。

共同富裕的重要内涵是“共享与发展”, 现有文献大多以人均国民收入来衡量“发展”, 以“收入基尼系数”来衡量“共享”( 万海远和陈基平, 2021) 。本文的一个启示是机会不平等系数也可以作为衡量“共享”的一个重要维度。除了关注收入的总体不平等( 基尼系数) 外, 也应该要关注收入的机会不平等。通过机会公平促进结果公平是实现低收入群体收入跃升进而促进共同富裕的有效路径。当然, 机会平等是共同富裕的必要条件, 但不是充分条件, 有了均等的机会后还需要设计合理的制度激励公民付出努力去追求“共同富裕”。

## 六、结论与政策建议

随着脱贫攻坚圆满收官, 中国经济社会发展进入新阶段, 如何从根本上破解共同富裕的难题, 寻找实现共同富裕的最优路径成为社会关注的焦点。本文认为机会不平等是实现全体人民共同富裕的主要障碍之一, 应该从低收入人群如何破除机会障碍实现收入跃升的视角理解这一命题。因此, 本文利用 CHIP 2018 年的最新数据, 通过引入机会不平等的分析框架并加以拓展, 从学理上对这一问题进行了系统的分析。本文的研究发现, 个体是否陷入低收入陷阱很大程度上取决于个体出生即很难改变的机会因素。基于传统参数估计方法和机器学习中的条件推断树方法, 本文测算发现与个体低收入状态相关的机会不平等系数为 0. 310( 参数估计结果) / 0. 336( 机器学习结果) , 这意味着超过 30% 的低收入状态与相对弱势的机会因素有关。本文发现机器学习方法在一定程度上能够弥补参数估计方法的缺陷, 尤其是在样本数量较小的情况下, 机器学习的结果会更精确; 而参数估计方法虽然在小样本情况下具有一定劣势, 但是在在大样本估计中其估计结果相对可靠, 并且能够更为便捷地对整体机会不平等系数进行分解, 以分析个别机会要素的作用。本文为后续研究在二元结果变量的情境下如何选择合适的方法测度机会不平等系数提供了新的思路。同时, 本文的渠道分析进一步解释了其中的传导机制, 机会要素导致个体教育获得的差异是主要的渠道, 对于总体机会差异系数的贡献比率高达 42. 237% , 在控制教育渠道之后, 就业渠道的贡献比率为 16. 594% 。当然, 后续对于机会渠道的探索也应该关注健康这一重要的人力资本路径。

共同富裕的重点在于提高低收入群体的收入, 而要“提低”不能仅仅依靠短期转移支付, 更应该从源头上破除机会性壁垒, 提高低收入群体的人力资本( 如受教育程度) 和职业技能, 从而赋予其长期发展的能力和动力。具体而言, 本文分析揭示了四个重点方向:

第一, 推进户籍制度改革, 致力于缩小城乡机会性差异。本文研究发现, 户口是造成个体陷入低收入状态的最重要机会因素。目前大多数的低收入人口还是在农村, 部分农村居民虽然已经脱贫, 但是面临较大的生计脆弱性和返贫风险, 关注这部分人的收入提升问题应该是共同富裕推进的工作重点。中国的户籍制度虽然已经在不同城市有不同程度的改革, 但从目前的效果来看, 拥有农村户籍的人口依旧在

劳动力市场上面临劣势,医疗和教育等公共服务还和城市居民有着较大差距。第二,优化区域发展政策,加大对中西部地区低收入群体的扶持力度。地域因素是影响低收入群体收入的第二大机会因素,出生在中西部地区的人口相较于东部而言,面临更少的发展机会和平台。更多的资金应该用于投资中西部地区的人力资本建设,同时合理引导人口在不同区域之间的流动,为出生在相对欠发达地区的个体提供更多的选择空间。第三,重视代际传递问题,畅通社会上升渠道。大量代际传递文献指出父母的收入和人力资本在很大程度上会传递给子代,而本文的研究也发现家庭背景是影响个体低收入状态的重要机会因素。从共同富裕的视角来看,当一个社会拥有畅通的向上流动渠道时,社会的财富才更有可能被共享。要推进共同富裕,应该更关注那些家庭背景较弱的个体,通过公平的基础教育体系提升其人力资本,借助职业技能培训帮助其拓宽职业选择路径。第四,关注农村老年群体收入,打通共同富裕薄弱环节。农村老年群体是共同富裕推进容易忽略的目标人群,其养老金和医疗保险等公共服务相较于城市老年群体而言差距较大,更容易陷入低收入状态。对于这部分群体而言,应该采取兜底性的社会保障政策,基于最低富裕标准,在确保财政可负担的前提下,通过转移支付等方法进一步提升其生活水平。

本文加深了对共同富裕实现路径的认识,但是本文的结论对于全面理解共同富裕依旧有两点局限:首先,“如何实现共同富裕”是一个系统性命题,本文的研究结论无法为其提供全面的答案。机会平等是促进共同富裕的一个重要因素,但两者并不完全等同。共同富裕是一个与财富积累有关的动态过程,而非一种静态结果(罗明忠,2022)。本文的机会不平等测算是相对静态的视角,但本文为后续研究如何利用机会不平等框架做动态分析提供了重要参考。其次,共同富裕是一个多维度概念,不仅包括本文讨论的收入和教育等领域,也囊括了医疗、社保、精神等方面的内涵,后续研究可以在更为全面的共同富裕指标体系下进行机会不平等及相关问题的研究。

#### 参考文献

- 陈东、黄旭锋 2015 《机会不平等在多大程度上影响了收入不平等?——基于代际转移的视角》,《经济评论》第1期。
- 陈宗胜、沈扬扬、周云波 2013 《中国农村贫困状况的绝对与相对变动——兼论相对贫困线的设定》,《管理世界》第1期。
- 程名望、Jin Yanhong、盖庆恩、史清华 2014 《农村减贫:应该更关注教育还是健康?——基于收入增长和差距缩小双重视角的实证》,《经济研究》第11期。
- 陈丽君、郁建兴、徐钦娜 2021 《共同富裕指数模型的构建》,《治理研究》第4期。
- 范从来 2017 《益贫式增长与中国共同富裕道路的探索》,《经济研究》第12期。
- 龚锋、李智、雷欣 2017 《努力对机会不平等的影响:测度与比较》,《经济研究》第3期。
- 胡安宁 2014 《教育能否让我们更健康——基于2010年中国综合社会调查的城乡比较分析》,《中国社会科学》第5期。
- 黄乃静、于明哲 2018 《机器学习对经济学研究的影响研究进展》,《经济学动态》第7期。
- 江求川、任洁、张克中 2014 《中国城市居民机会不平等研究》,《世界经济》第4期。
- 厉以宁 2002 《以共同富裕为目标,扩大中等收入者比重,提高低收入者收入水平》,《经济研究》第12期。
- 李实、宋锦、刘小川 2014 《中国城镇职工性别工资差距的演变》,《管理世界》第3期。
- 李实 2021 《共同富裕的目标和实现路径选择》,《经济研究》第11期。
- 李实、朱梦冰 2022 《推进收入分配制度改革促进共同富裕实现》,《管理世界》第1期。
- 李莹、吕光明 2019 《中国机会不平等的生成源泉与作用渠道研究》,《中国工业经济》第9期。
- 刘波、胡宗义、龚志民 2020 《中国居民健康差距中的机会不平等》,《经济评论》第2期。
- 刘波、王修华、彭建刚 2015 《我国居民收入差距中的机会不平等——基于CGSS数据的实证研究》,《上海经济研究》第8期。
- 刘培林、钱滔、黄先海、董雪兵 2021 《共同富裕的内涵、实现路径与测度方法》,《管理世界》第8期。
- 刘国光 2011 《关于国富、民富和共同富裕问题的一些思考》,《经济研究》第10期。
- 李春玲 2014 《“80后”的教育经历与机会不平等——兼评〈无声的革命〉》,《中国社会科学》第4期。
- 李金叶、郝雄磊 2019 《机会不平等的测度:回归树模型的应用与比较》,《统计与信息论坛》第10期。
- 罗楚亮、滕阳川、李利英 2019 《行业结构、性别歧视与性别工资差距》,《管理世界》第8期。
- 罗明忠 2022 《共同富裕:理论脉络、主要难题及现实路径》,《求索》第1期。
- 雷欣、贾亚丽、龚锋 2018 《机会不平等的衡量:参数测度法的应用与改进》,《统计研究》第4期。

- 马超、曲兆鹏、宋泽 2018 《城乡医保统筹背景下流动人口医疗保健的机会不平等——事前补偿原则与事后补偿原则的悖论》,《中国工业经济》第2期。
- 沈扬扬、詹鹏、李实 2018 《扶贫政策演进下的中国农村多维贫困》,《经济学动态》第7期。
- 史新杰、卫龙宝、方师乐、高叙文 2018 《中国收入分配中的机会不平等》,《管理世界》第3期。
- 宋扬 2017 《中国的机会不均等程度与作用机制——基于 CGSS 数据的实证分析》,《财贸经济》第1期。
- 汪三贵、胡骏 2020 《从生存到发展: 新中国七十年反贫困的实践》,《农业经济问题》第2期。
- 王芳、王宣艺、陈硕 2020 《经济学研究中的机器学习: 回顾与展望》,《数量经济技术经济研究》第4期。
- 万广华、吴万宗 2019 《走向共同富裕的制度力量——评陈宗胜等著〈中国居民收入分配通论〉》,《经济研究》第11期。
- 万海远、陈基平 2021 《共同富裕的理论内涵与量化方法》,《财贸经济》第12期。
- 吴晓刚、张卓妮 2014 《户口、职业隔离与中国城镇的收入不平等》,《中国社会科学》第6期。
- 徐舒 2010 《技术进步、教育收益与收入不平等》,《经济研究》第9期。
- 章元、万广华、史清华 2013 《暂时性贫困与慢性贫困的度量、分解和决定因素分析》,《经济研究》第4期。
- 周京奎、王文波、龚明远、黄征学 2020 《农地流转、职业分层与减贫效应》,《经济研究》第6期。
- 朱梦冰、李实 2017 《精准扶贫重在精准识别贫困人口——农村低保政策的瞄准效果分析》,《中国社会科学》第9期。
- 郑冰岛、吴晓刚 2013 《户口、“农转非”与中国城市居民中的收入不平等》,《社会学研究》第1期。
- 赵西亮 2017 《教育、户籍转换与城乡教育收益率差异》,《经济研究》第12期。
- 周国富、陈菡彬 2021 《产业结构升级对城乡收入差距的门槛效应分析》,《统计研究》第2期。
- Bourguignon, F., F. H. Ferreira, and M. Menéndez, 2007, “Inequality of Opportunity in Brazil”, *Review of Income and Wealth*, 53(4), 585—618.
- Brunori, P., P. Hufe, and D. G. Mahler, 2018, *The Roots of Inequality: Estimating Inequality of Opportunity from Regression Trees*, World Bank, Washington, DC.
- Brunori, P., V. Peragine, and L. Serlenga, 2019, “Upward and Downward Bias When Measuring Inequality of Opportunity”, *Social Choice and Welfare*, 52(4), 635—661.
- Ferreira, F. H., and J. Gignoux, 2011, “The Measurement of Inequality of Opportunity: Theory and an Application to Latin America”, *Review of Income and Wealth*, 57(4), 622—657.
- Golley, J., Y. Zhou, and M. Wang, 2019, “Inequality of Opportunity in China’s Labor Earnings: The Gender Dimension”, *China & World Economy*, 27(1), 28—50.
- Golley, J., and S. T. Kong, 2018, “Inequality of Opportunity in China’s Educational Outcomes”, *China Economic Review*, 51, 116—128.
- Jusot, F., S. Tubeuf, and A. Trannoy, 2013, “Circumstances and Efforts: How Important Is Their Correlation for the Measurement of Inequality of Opportunity in Health?”, *Health Economics*, 22(12), 1470—1495.
- Jia, R., X. Lan, and G. P. I. Miquel, 2021, “Doing Business in China: Parental Background and Government Intervention Determine Who Owns Business”, *Journal of Development Economics*, 151, 1—12.
- Juárez, F. W. C., and I. Soloaga, 2014, “Iop: Estimating Ex-ante Inequality of Opportunity”, *Stata Journal*, 14(4), 830—846.
- Li, H., L. Meng, X. Shi, and B. Wu, 2012, “Does Having a Cadre Parent Pay? Evidence from the First Job Offers of Chinese College Graduates”, *Journal of Development Economics*, 99(2), 513—520.
- Paes De Barros, R., M. De Carvalho, and S. Franco, 2007, “Preliminary Notes on the Measurement of Socially-Determined Inequality of Opportunity when the outcome is Discrete”, Unpublished Manuscript. Rio de Janeiro: IPEA.
- Palomino, J. C., G. A. Marrero, and J. G. Rodríguez, 2019, “Channels of Inequality of Opportunity: The Role of Education and Occupation in Europe”, *Social Indicators Research*, 143(3), 1045—1074.
- Roemer, J. E., 1998, *Equality of Opportunity*, Cambridge: Harvard University Press.
- Shi, X., 2019, “Inequality of Opportunity in Energy Consumption in China”, *Energy Policy*, 124, 371—382.
- Shorrocks, A. F., 2013, “Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value”, *Journal of Economic Inequality*, 11(1), 99—126.
- Singh, A., 2012, “Inequality of Opportunity in Earnings and Consumption Expenditure: The Case of Indian Men”, *Review of Income and Wealth*, 58(1), 79—106.
- Yang, X., B. Gustafsson, and T. Sicular, 2021, “Inequality of Opportunity in Household Income, China 2002—2018”, *China Economic Review*, 69, 1—18.
- Zhang, Y., and T. Eriksson, 2010, “Inequality of Opportunity and Income Inequality in Nine Chinese Provinces, 1989—2006”, *China Economic Review*, 21(4), 607—616.

# Understanding Common Prosperity from the Perspective of Equality of Opportunity: An Empirical Study Based on the Low-income Groups

SHI Xinjie<sup>a, b</sup>, LI Shi<sup>b, c</sup>, CHEN Tianzhi<sup>d</sup> and FANG Shile<sup>e</sup>

( a: China Academy for Rural Development, Zhejiang University; b: School of Public Affairs, Zhejiang University;

c: Institute for Common Prosperity and Development, Zhejiang University; d: Australian National University;

e: School of Economics, Zhejiang Gongshang University)

**Summary:** With the major victory of poverty alleviation, how to fundamentally solve the challenges of common prosperity and find the optimal path to achieve common prosperity has become the focus of Chinese society. Relevant studies show that common prosperity should focus on two dimensions including overall development and shared development. In addition to economic development, attention should also be paid to the sharing of development opportunities and various public services. This paper argues that inequality of opportunity is one of the main obstacles to achieving common prosperity for all people, and this proposition should be understood from the perspective of how low-income people can break the barriers of opportunity to achieve an income jump.

This paper uses the data of the Chinese Household Income Project (CHIP) for 2018 to expand the research framework of inequality of opportunity to the discussion of binary variables. With low-income groups as the starting point, this paper systematically examines the current situation of inequality of opportunity in China, that is, to what extent circumstances (such as gender, household registration, place of birth, family background, age, etc.) that are beyond the control of individuals lead to their falling into a low-income state and thus affect the realization of common prosperity. There are three main findings in this paper. First, whether an individual falls into the low-income trap largely depends on circumstances that are difficult to change at birth. Based on the traditional parameter estimation method and the conditional inference tree method (machine learning), this paper estimates that the coefficient of inequality of opportunity related to individual low-income status is as high as 0.310 (parameter estimation result) and 0.336 (machine learning result), which means that more than 30% of low-income status is related to relatively weak opportunity factors. Second, the difference in individual education acquisition is the main channel through which circumstances affect the low-income status, accounting for 42.237% of the overall inequality of opportunity. After controlling for the education channel, the contribution rate of the employment channel is 16.594%. Third, this paper further finds that, in addition to the low-income status, individuals' educational level and income level are also constrained by circumstances.

This paper contributes to existing research in three ways. First, this paper starts from the perspective of income jump of the low-income population, discusses inequality of opportunity in the context of common prosperity, and extends the theoretical connotation and practical significance. Second, at present, the measurement of inequality of opportunity is mainly based on continuous variables (such as income) using the parameter estimation method. In this paper, we use both parameter estimation and machine learning methods, which provides the possibility of causal analysis for the study of inequality of opportunity. We find that the machine learning method can make up some defects of the parameter estimation method to a certain extent. Particularly, when the sample size is small, the results of machine learning will be more accurate. Although the parameter estimation method has certain disadvantages in the case of small sample size, its estimation results are relatively reliable in the case of large sample size, and it can more easily decompose the coefficient of overall inequality of opportunity to analyze the role of individual opportunity factors. This paper provides a new idea for subsequent research on how to choose an appropriate method to measure the coefficient of inequality of opportunity in the context of binary outcome variables. Third, this paper discusses how circumstances play a role in influencing low-income status through education and employment.

This paper emphasizes the role of equality of opportunity in promoting common prosperity, and the relevant conclusions provide important reference for policies supporting low-income people in the context of common prosperity.

**Keywords:** Equality of Opportunity; CHIP 2018; Common Prosperity; Machine Learning

**JEL Classification:** D63, H53, J10

(责任编辑: 陈小亮) (校对: 王红梅)