

人口流动对生育水平的影响： 基于选择性的分析^{*}

周 皓

【内容摘要】文章借助反事实因果推论框架,利用倾向得分匹配方法以克服传统人口统计学的弱点,使用 2000 年第五次人口普查数据,以曾生子女数和普查前一年的生育作为生育水平指标,尝试讨论流动人口的选择性对估计其生育水平的影响作用;并在控制选择性后,比较流动人口与非流动人口的生育水平差异,以期为深入讨论人口流迁与生育的因果关系提供基础。Heckman 二阶段模型表明,观测到的流动与非流动人口生育水平差异由于选择性而存在偏差。三种倾向得分匹配方法得到一致结果:剔除人口流动的选择性后,流动人口的生育水平显著低于非流动人口;且两者的差异大部分可以被选择性所解释。从某种意义上说明人口流动有助于降低生育水平。

【关键词】人口流动; 生育水平; 选择性; 倾向得分匹配

【作者简介】周皓,北京大学社会学系副教授、北京大学中国社会与发展研究中心研究员。北京: 100871

Selectivity Bias in the Effect of Migration on Fertility

Zhou Hao

Abstract: With the counter-factual causal inference framework, this paper employs various propensity score matching methods to explore the effect of selectivity of migration on the fertility level of floating population in China, and estimates and compares migrant and non-migrant fertility levels after controlling for the selectivity bias. Both lifetime fertility (children ever born) and period fertility (births in the year preceding the census) are examined using data from the 2000 census. Heckman 2-stage model shows that the observed difference between migrant and non-migrant fertility is biased due to selectivity of migration. The consistent results from three types of propensity score matching methods demonstrate that migrant fertility is markedly lower than non-migrant fertility, and a larger proportion of the difference is explained by the selectivity. Anyway, the study suggests that migration does reduce fertility.

Keywords: Migration, Floating Population, Fertility Level, Selectivity, Propensity Score Matching

Author: Zhou Hao is Associate Professor, Department of Sociology, and Researcher, Center for Sociological Research and Development Studies of China, Peking University. Email: zhouh@pku.edu.cn

* 本研究得到国家社科基金项目“流动人口生育率转变研究”(09BRK003)资助。

人口流迁是否会降低流动人口(移民)的生育水平? 这个看似简单的迁移与生育之间的关系问题, 实则很难回答。以往的研究总是直接比较移民与非移民之间在生育水平上的差异。国外经验研究表明, 移民的生育水平一般总是低于迁出地非移民的生育水平、却高于迁入地非移民的生育水平; 而且, 移民的生育水平与模式由于社会融合与适应而越来越接近于迁入地的非移民群体 (Kulu, 2005)。其中主要的理论包括在迁入地的社会融合理论、移民选择性理论等。在国内, 石人炳、熊波 (2011) 曾综述了国内外有关人口迁移与生育关系的部分研究。陈卫、吴丽丽 (2006)、郭志刚 (2010)、李丁和郭志刚 (2014) 分别利用我国 2000 年第五次人口普查、2005 年小普查数据以及国家卫计委“全国流动人口动态监测调查系列数据”利用统计方法和人口统计学方法(年龄别生育率与总和生育率、婚姻结构标准化等), 讨论了流动人口的生育水平问题。得到的结论都认为: 流动人口的生育率已经很低, 不仅低于农村常住人口, 而且也低于城市常住人口。本文无意推翻这些结论。

但上述研究只是刻画了移民生育水平与非移民之间的差异问题, 而未能真正地回答迁移与生育之间的因果关系问题。一方面, 由于迁移(流动)人口的选择性, 观测到的移民(流动人口)与非移民之间在生育水平上的差异, 可能是由于选择性引起的, 而并不一定表明移民的生育水平确实低于非移民群体。因此, 要回答上述因果关系问题, 就首先应该回答: 在剔除了选择性问题之后, 移民(流动人口)的生育水平是高于还是低于非移民。另一方面, 也是更为困难的, 即两者的因果关系包含了许多外在的无法被剥离的因素, 如移民(或流动人口)在流入地的生活状况与环境 (social context) 等 (Kulu and Washbrook 2014)、移民本身携带的流出地的各种社会文化背景与蕴涵 (Amara 2014) 等, 以及迁移与生育之间的时间关系等等。这些混合在一起的各种因素使我们无法真正分析迁移/流动这个单一因素对他们生育水平的影响作用。因此, 要想回答迁移对生育的影响作用, 事实上是非常困难的。

有鉴于此, 本文将暂时忽略其他因素的混合效应, 以回答剔除流动人口^①的选择性后的生育水平为分析目标, 借助反事实因果推论框架, 利用倾向得分匹配方法以克服传统人口统计学中的弱点, 并利用 2000 年第五次人口普查数据, 比较讨论流动人口与非流动人口在生育水平上的差异, 为进一步深入讨论人口流迁与生育的关系提供基础。简单地说, 本文希望能够回答, 流动人口的生育水平是否真的低于非流动人口?

1 已有研究的思考

生育率的计算方法一般都是: 先挑选出 15~49 岁的育龄妇女, 然后区分流动与非流动, 并以调查时点前一年的生育情况为基础, 将各年龄别生育率加总得到总和生育率 (TFR), 以表示流动与非流动人口的生育水平。暂且忽略时期指标与生育模式的影响作用 (如郭志刚 2000 年以来一直讨论的去进度效应问题等), 当前观测到的流动人口和非流动人口在生育水平上的差异是否是真实的差异? 这种差异是否能够真正地表示人口流迁对生育的影响作用呢?

借用反事实因果推论的框架^②, 将生育水平定义为因变量 (其操作化定义可以是曾生子女数或普

① 我国流动人口与迁移人口由于户籍不同而存在显著差异。为简单起见, 本文先将研究对象限定于流动人口, 而将迁移人口与非流动人口合并。

② 有关反事实因果推论框架请参考相关书籍或论文 (Heckman 2005; Morgan and Harding 2006; Morgan and Winship, 2007; Brand and Xie 2010; 郭申阳、马克 W 弗雷泽 2012; Xie 2013)。其中, Morgan & Winship (2007) 一书详细介绍了反事实因果推论的框架及各种解决方法; 郭申阳等 (2012) 一书详细介绍了倾向得分匹配方法。谢宇 (Xie and Wu 2005; 谢宇 2006; Brand and Xie 2010; Xie Yu 2013) 的系列文章不仅讨论了选择性问题 (实验前的异质性), 而且还深入讨论了实验效应的异质性问题。

查前一年的生育等变量^①);将人口流动看成是一种实验,则是否参加这个实验即可被定义为一个虚拟变量:0表示不参加(即非流动人口,用c表示),1表示参加(即流动人口,用t表示);由此,实验效应就是流动人口与非流动人口在生育水平上的差异。在没有任何选择性偏差的情况下,即: $\delta = \bar{Y}^t - \bar{Y}^c$ 。但由于这两个均值中至少有一个是反事实的量而无法被直接观测到,实验效应因此也无法被直接计算。更为重要的是,由于实验前异质性的存在(pretreatment selection bias ,or heterogeneity),即是否进入实验是有选择性的,那么上述公式右边的实验组与控制组的均值可能会由于选择性而存在偏差,从而使得根据观测数据计算得到的实验效应(即公式左边的平均差异值)可能是有偏的。套用到本文中,等式右边的两个观测值分别是流动人口与非流动人口的平均生育水平;由于流动人口的选择性,根据观测数据直接计算得到的两类人群的平均差异可能是有偏的,无法真实地表示流动对生育水平的影响作用(不考虑其他混合效应)。

首先,人口流动选择性的存在。自雷文斯坦(Ravenstein,1885)发表《人口迁移规律》以来,国内外的各种实证研究和迁移理论都试图回答“谁是迁移者”这一问题(Mangalam,1968;杨云彦,1992;Massey and Douglas et al.,1993;Zhao,1999;段成荣,2000;周皓,2000;李玲,2001;周皓,2003;周皓、梁在,2006;段成荣、杨舸、张斐,2008)。尽管目前的研究无法完整地回答这一问题,但人口迁移与流动存在选择性则是明确的。正是由于人口流动在性别、年龄、婚姻状况、受教育水平、甚至社会经济地位等各种特征上存在的选择性,使得观测到的流动人口和非流动人口在生育水平上的差异可能是由两方面的因素引起的:流动的选择性与生育水平的差异。而并不能只归因于两类人口在生育水平上的差异。如郭志刚(2010)曾深入讨论了婚姻结构对流动人口总和生育率的影响作用,说明了流动人口与非流动人口“如果在年龄已婚或未婚比例存在着显著差别,那么便会影响其总和生育率,并且使得不同类别之间的总和生育率缺乏可比性”(郭志刚,2010)。李丁、郭志刚(2014)亦证明了这一点。只有对婚姻结构进行标准化以后,才能比较流动与非流动人口之间在总和生育率上的差异。这说明人口流动的选择性导致目前的研究结果可能存在偏差。

其次,人口流迁选择性的多维度特征,使传统人口学方法失效。人口流动选择性存在的同时,还有着多维度的特征。流动人口不仅在年龄结构上与非流动人口存在差异,而且,在其他诸如婚姻结构、受教育水平等多个维度上存在选择性。正是由于这种多维度特征,使得传统人口学方法无法真正有效地讨论人口流动对生育水平的影响作用。尽管传统人口学方法中,总和生育率可以控制年龄结构的影响,孩次别总和生育率可以控制孩次的影响;再利用(直接或间接)标准化可以控制其他某一结构(如婚姻结构或受教育水平等)的影响,但传统人口学方法无法综合有效地剔除其他所有结构可能产生的影响作用,“调整生育指标往往只控制了一种结构偏差产生的影响,并不能对所有结构偏差影响加以整体控制”(李丁、郭志刚,2014)。传统人口学方法无法解决由于人口流动选择性的多维度特征而导致的结构偏差。从这个意义上说,在面对人口流迁与生育水平关系问题的讨论中,传统的人口学方法是失效的。

再次,流动人口的选择性同样也可能导致他们在生育时间上的选择性,进而使观测到的流动人口和非流动人口在生育水平上的差异可能并非真正是由生育而引起的差异。“流动人口的主体实际上应该是原来农村中年轻力壮、思想活跃、素质较高的一部分人组成。而流动的目的主要是追求较高经

① 这里所用的结果变量应该是个体层次的变量,而非汇总后的变量。

济收益和转变生活方式,因此这种人口流动应该促使了晚婚晚育和生育率的下降”(郭志刚,2010)。因此,他们的生育时间,特别是流动人口在调查前一年中的生育,极有可能同样具有很强的选择性。

最后,对人口流迁与生育关系的研究中,我们更应该注意到除选择性以外的两个问题。一是因果推论问题。其关键在于时间顺序。从方法论角度看,我们应该关注原因的结果(the effects of cause),而不是结果的原因(the causes of effect)。人口流迁与生育的关系研究必须注意作为原因的流迁应该在生育之前。即通过识别两者在时间上的先后顺序,来比较有过流迁经历(且流迁发生在生育之前)与没有流迁经历(或流迁发生在生育之后)的人在终身生育水平上是否存在差异,才能说明流迁对生育水平的影响作用。但是,截面型的观测数据,却无法满足这种因果推论的前提条件,这就需要使用跟踪数据来予以说明,至少也需要用回顾性调查数据。二是指标选择。“时期生育率与终身生育数量是从不同角度测量生育,时期生育率很低并不一定代表终身生育水平就一定低”(郭志刚,2010)。这一点正适用于流动人口。当前流动人口较低的时期生育指标,由于存在各种选择性与删截等原因,可能并不代表他们的终身生育水平就一定低,进而无法说明人口流迁对生育的作用。现在的关键在于,无法匹配终身生育水平与人口流迁的关系。如人口普查数据仅提供五年内流迁人口,但无法判断五年之前的情况,从而无法确定其流迁时间与生育时间的先后顺序;而五年中的流迁人口,则更多的可能是尚未结婚或未完成生育的年轻人,从而无法得到其终身生育指标(或许可以判断流迁是否延迟了其生育时间)。这也正是当前对人口流迁与生育关系讨论中的难点所在。

正是上述这些选择性的存在,利用观测数据得到的流动与非流动人口在生育指标上的差异可能并不能够真正表示两类人群在生育水平上的真实差异。本文希望能够剔除选择性偏差后再比较两类人群在生育水平上的差异,以便为今后的进一步深入分析与研究提供基础。

2 数据与变量选择

本文选用的数据是2000年第五次人口普查数据。之所以选用该人口普查数据,而非其他大型的社会调查数据(如CGSS或CFPS),主要原因是数据可得性和样本规模问题,特别是流动人口规模和生育数量。人口普查数据在样本规模上无疑具有无可比拟的优势。当然人口普查数据也同样存在各种问题,如截面数据对因果推论的限制、漏报(特别是流动人口的漏报、及普查前一年生育的漏报等)对估计结果的影响^①等。且自2000年以来我国的人口迁移与流动发生了一些可能是本质性的变化,因此,本研究只是思路与方法上的一个实验性扩展,而并非最终结论。

2.1 迁移与流动人口的定义

迁移与流动人口,可以按照出生地和现居住地之间的关系来识别与定义,也可以按照流入时间来定义,也可以按照户口登记状况及流/迁入时间来定义。本文将以流动人口为研究对象,因此,将利用户口登记状况(即五款人),而不考虑迁移人口和迁/流入时间。具体做法是:按照R6户口登记状况,将第1、4、5款人合并成本地人口,将第2、3款人合并成流动人口。当然,应该注意到文中的流动人口,既可能是近五年流入的,也可能是远期流入人口;而且时间同样也可能会影响流动人口的生育水平。

^① 感谢匿名评审专家对所用数据的质疑。笔者同意,普查漏报可能会导致一定的偏差;而且漏报可能是选择性的非随机的(如流动人口可能漏报得更多,而非流动人口则漏报得较少;当然这种猜测尚需验证)。但五普的漏报情况需要借助其后2010年六普数据来进行评价与填补;即便填补以后,也无法得到被填补人口的各种信息(特别是其父母亲与家庭的信息)。因此,本文直接使用原始数据,而未对其进行调整。

2.2 变量选择与描述

人口普查数据中的变量相对较少。本文所选择的变量与以往研究基本相同,主要包括年龄(和年龄的平方)、性别、受教育水平、民族、婚姻状况。尽管家庭户户主特征及家庭结构等因素会影响到人口迁移与流动(周皓,2004),但人口普查资料是以迁入地为调查地点,现有的家庭特征及结构等无法反映进行迁出决策的具体背景,因此,本文未包含这些迁出地的信息。当然,另一方面本文也只是进行尝试性的分析与讨论。同时,由于本文仅关注育龄妇女的生育水平,因此,分析数据仅包括了年龄为15~49岁之间的育龄妇女,而不包括男性和其他年龄女性的案例。有关变量的描述性统计结果见表1。同时,表1还给出了流动人口与非流动人口之间在各变量上的差异是否显著的检验结果。

2.3 变量描述结果的讨论

表1中的数据清楚地表明,流动人口与总人口(和非流动人口)群体是两个完全不同的群体,存在着显著的差异。

从因变量(曾生子女数和普查前一年的生育)来看,流动人口的曾生子女数显著低于总人口;但普查前一年的生育(仅为比例,相当于是粗生育率)则高于总人口。这种看似矛盾的结果,恰恰说明利用普查前一年的生育来讨论流动人口的(终身)生育水平问题可能会存在偏差。这可能是由于流动人口在生育安排上不同于非流动人口,或者是他们的生育可能并不是随机现象,即普查前一年的生育,有可能在某种意义上无法代表流动人口的生育水平。

自变量在两类人群之间的差异则体现了迁移选择性的问题。流动人口是更为年轻的(即平均年龄较低)、以农村流向城市(即城乡变量中流动人口的数值较低,表示原流出地为农村的占32.34%)、以未婚或单身(有配偶的仅占58.36%)为主。正如李丁、郭志刚(2014)所指出的,流动人口在婚姻结构上的差异,导致计算所得的生育水平可能存在偏差。同样的,流动人口与非流动人口在其他结构上存在的差异,亦体现了流动人口的选择性。

2.4 方法

倾向得分匹配方法是基于反事实因果分析框架的一种分析方法,已有不少文献介绍其基本思路与方法(王天夫,2006;Morgan and Winship,2007;Fu Alex Z.,2008;郭申阳、马克W弗雷泽,2012)。本文将使用涉及倾向得分四大类方法中的三类:贪婪匹配(greedy matching)、权重分析法和以内核为基础的匹配(kernel)^①。其中,贪婪匹配中又包括了最近邻居法(nearest neighbor)和马氏距离法两种。需要注意的是,采用这些方法并不能够直接估计得到平均实验效应(average treatment effect,ATE),而是需要利用匹配以后的样本再重新估计。如贪婪匹配首先对流动人口(实验组)与非流动人口(控制组)进行匹配,然后仅对匹配样本进行各种分析;权重分析法即以倾向得分为权重进行新的加权分析。而以内核为基础的匹配则是非参估计。相关方法的介绍请见郭申阳等(2012)。具体命令为STATA中的psmatch2。

这几种方法本质上是相同的,之所以选用多种方法是为了检验结果的可靠性。如果不同方法之间所得到的结果在方向与显著性上都相同,则结果相对可靠;但如果多种结果之间存在差异,则需要重新考虑。

^① 未采用的另一类方法是倾向得分的分层方法。

表 1 所选变量的描述性统计量*

Table 1 Descriptive Statistics and Statistical Test of the Differences between Migrants and Nonmigrants

	全部	流动人口	统计检验
受教育水平			
文盲	6.44	2.76	Pearson chi2(4) = 7.4e+03
小学	31.86	17.44	
初中	43.30	47.65	
高中	14.64	23.91	
大专及以上	3.76	8.23	
N	306996	32547	
曾生子女数	1.33781 (1.13921)	0.6937 (0.83903)	109.9857***
生育	0.03136 (0.17429)	0.04894 (0.21576)	-19.2599***
年龄	32.02680 (9.63861)	26.7046 (8.07448)	107.3149***
年龄平方	1118.61800 (621.86180)	778.333 (489.924)	106.3139***
民族	0.91466 (0.27938)	0.94012 (0.23727)	-17.3925***
城乡(0 = 农村) ^①	0.39206 (0.48821)	0.32587 (.46871)	6.3914***
户口(0 = 农业户口)	0.24258 (0.42864)	0.32338 (0.46777)	-36.0442***
婚姻原因流动	—	0.12118 (0.32634)	— ^②
婚姻状况(0 = 无配偶)	0.76264 (0.42547)	0.58359 (0.49297)	81.1527***
N	306996	32547	

* 表中各数据为均值或频率, 括号内为标准差; 表中多数变量为分类变量, 但在比较流动人口与非流动人口的差异时则使用了 t 检验, 而未用卡方检验, 原因是希望检验结果不受同一类人口中不同类别的分布的影响作用。

注①城乡变量在流动人口中, 是以流动人口的原流出地的城乡属性替换现住地的性质;

②对非流动人口而言, 无流动原因一项。

*** $p < 0.01$ 下表同。

2.5 分析思路

本文目的是剔除人口流动的选择性后, 比较讨论流动人口与非流动人口在生育水平上的差异, 为

进一步深入讨论人口流迁与生育的关系提供基础。在此之前,首先应该考虑流动人口的选择性(内生性的一种)是否真的对目前的研究结果有影响。只有当选择性偏差确实存在时,才有必要利用统计方法剔除这种选择性。检验以后,本文将利用倾向得分匹配的各种方法来估计剔除选择性以后,流动人口与非流动人口在生育水平上的差异。其中有两点需要注意:

其一,所用的指标。正如前文所述,时期指标由于可能存在生育时间的选择性而无法真正反映流动人口的生育水平,也就无法反映生育与流迁之间的关系,因此,本文将同时选用时期指标与累积指标(曾生子女数),以更全面地反映人口流动与生育的关系。

其二,比较对象。以往在讨论人口流迁时,总是以迁入地作为标准。比如在以往讨论人口流动的原因时,总是将流动人口与流入地的非流动人口作为一个总体。这种分析事实上并不能够真正地反映流动原因,因为只有将流动人口复原到其流出地,并将他们与流出地的非流动人口相结合,才有可能真正反映他们的流动原因。因此,本文在数据处理时,先将所有的省际流动人口全部放回到其流出省份,然后再进行匹配与比较。

3 选择性偏差的检验

选择性偏差是否存在,可以利用 Heckman 2-stage 模型来进行检验。STATA 中由原来的 `treatreg` 发展到了现在的 `etregress` 和 `etpoisson` 两种方法,具体可参见 STATA 手册。其中前者 `etregress` 是对连续因变量的检验,而后者 `etpoisson` 则是对非连续的因变量的检验。所得结果中的 $\rho(\rho)$ 表示回归方程的误差项与结果变量之间的相关性。如果 $\rho(\rho)$ 显著,则表明存在选择性偏差^①。事实上,这两个模型同时还可以计算控制选择性偏差以后的实验效应(如平均干预效应 ATE: average treatment effect 和实验组的平均干预效应 ATE: average treatment effect for the treated)。但要注意的是,这种方法的本质事实上仍然是用回归方程中的系数来表示两类人群的差异,而非两类人群在相同倾向得分情况下的比较结果。因此,从某种意义上说,其所得结果并不能够真正表示平均干预效应。

由表 2 可以看到,不论是以曾生子女数(模型(1)),还是普查前一年是否生育(模型(2))为因变量,模型中所有自变量均呈现出显著;且每个模型的 $\rho(\rho)$ 也都显著。这表明,由于流动人口的选择性而导致的生育水平的偏差是显著的。即:存在选择性偏差问题。当然,从模型的角度来看,这种选择性偏差既有可能是由流动的选择性而导致的,也有可能是由于模型设定(model specification)所导致的。

表中自变量“流动人口”的偏回归系数都是显著的,这说明,在剔除了部分选择性偏差后,流动人口与非流动人口之间在曾生子女数和普查前一年的生育可能性上都存在差异。从曾生子女数来看,该系数为 1.0881,这说明流动人口的曾生子女数比非流动人口的高 1.0881。从普查前一年生育情况来看,该系数为 -0.2009(取指数后为 0.8180),即在普查前一年,流动人口生育的可能性仅为非流动人口的 81.8%,下降了近 20%左右。

这是个比较有意思的结果。首先,将这一结果与描述性分析相比较,可以看到,曾生子女数由控制前流动人口低于非流动人口,转变为在控制选择性以后,流动人口高于非流动人口;而从普查前一年的情况则正好相反,即控制前较高的粗生育率转变为控制后较低的水平。这种转变同样说明流动人口的选择性对于生育率的影响作用。其次,曾生子女数并不能够完全用以说明妇女的终生生育率,即使是在控制了流动的选择性以后,因为分析过程仍然没有控制妇女的年龄与婚姻结构等能够导致妇女生育水平不同的因素。再次,流动人口在普查前一年的生育水平^②相对较低,这与以往的研究结

① 对于 $\rho(\rho)$ 是否能够真正表示存在选择性偏差,尚存在争论(郭申阳 2014)。

② 通常都简单地称之为流动人口的生育水平,而不再强调是普查前一年的生育水平。

果相同(陈卫、吴丽丽 2006; 郭志刚 2010)。可能的原因还是在于: 流动人口的生育安排可能并不是随机的, 如果流动人口在流动前已经生育过(不论多少), 那么, 他们在普查前一年中的生育显然可能会低于非流动人口。这一点尽管无法反映流动与生育的关系, 但有可能的是, 已生育的流动人口降低了普查前一年该类人口的生育水平, 但流动却并不一定能够降低他们的生育水平。

4 倾向得分匹配方法的分析结果

利用倾向得分匹配方法所得的具体结果见表 3。该表包括了两个因变量: 曾生子女数和普查前一年的生育。每个因变量的结果由两部分构成: 前一个 t 检验部分, 是基于贪婪匹配与以内核为基础的匹配这两种方法, 对挑选出的匹配案例进行 t 检验的结果; 后一个回归部分是对挑选出的匹配案例进行线性或 logistic 回归的结果^①。当然, 在回归分析中同样加入了部分自变量。表中回归部分的“系数”是“流动人口”这一虚拟变量在回归方程中的系数; odds ratio 是该虚拟变量在 logistic 回归中, 流动人口与非流动人口的发生比率。

4.1 曾生子女数的分析

表中第一行表示未匹配以前, 非流动人口的曾生子女数远高于流动人口的曾生子女数, 两者相差 0.7205 (sig. = 0.000)。这一结果和前面的结果相同。表示流动人口较低的平均曾生子女数可能是由于流动人口的选择性造成的。为此需要进行倾向得分匹配处理。

首先, 匹配以后的结果, 从整体来看, 除无设置的倾向得分匹配方法外, 其他各种匹配方法得到的结果有很好的一致性, 不论是两组人之间的差异, 还是差异的显著性水平。在差异幅度上, 流动人口的平均曾生子女数均一致低于非流动人口; 而且这种差异在 0.05 水平下显著。这种一致性表明: 其一, 倾向得分匹配方法所得的结果是可靠的; 其二, 再次验证了流动人口选择性的存在; 其三, 剔除流动人口选择性后再来进行讨论与分析的必要性。

其次, 匹配以后的结果再次证明流动人口的平均曾生子女数低于相同特征的非流动人口的平均曾生子女数。如果用反事实的因果推断来分析, 即流动人口如果不流动的话, 那么, 她们的平均曾生子女数应该比现在流动人口的水平要高。以贪婪匹配 - 最近邻居 1 的结果为例, 现在流动人口的平均曾生子女数为 0.6937; 但如果这批人她们不流动的话, 她们的平均曾生子女数应该是 0.8888。由于流动导致的平均曾生子女数减少了 0.1951, 下降了 22%。因此, 流动人口的生育水平低于非流动人口; 或者是, 人口流动有助于降低生育水平。尽管人口流动与生育水平之间的因果关系尚需深入讨论。

再次, 从差异幅度来看, 匹配后流动人口与非流动人口在曾生子女数上的差异比原来未匹配以前的差异有显著的减少。两类人口原有的差异为 -0.72; 即: 流动人口的平均曾生子女数比非流动人口的少 0.72。正如前文所述, 这种差异既有可能是由于流动人口本身较低的生育水平引起的, 也有可能是由于流动人口的选择性引起的。在剔除了流动人口的选择性以后, 即匹配以后, 以贪婪匹配 - 最近邻居法 1 的结果为例, 流动人口的平均曾生子女数仅比非流动人口的少 0.1951 个。差异从原来的 0.72 降低到了 0.19, 下降了近 73%; 且两种差异间的降低是显著的。其他几种方法所得的结果类似。在上述 0.72 的差异中大部分(近 73%)是由选择性而引起的; 实际上生育水平并没有这么大的差异。

^① 本文之所以选用线性回归, 而未用 poisson 回归来做的原因在于: 文章的目标在于控制其他个体特征以比较流动与非流动两组人口在平均曾生子女数上的差异, 即强调平均数的差异, 而不再强调它们的分布, 尽管线性回归可能会存在违反假设的问题。

表 2 Heckman 选择模型的结果

Table 2 The Estimation of Heckman Selection Models

	曾生子女数	普查前一年生育
年龄	0.0835 *** (0.0014)	0.6568 *** (0.0236)
年龄平方	-0.0003 *** (0.0000)	-0.0131 *** (0.0004)
户口	-0.4650 *** (0.0042)	-0.3324 *** (0.0330)
小学	-0.2635 *** (0.0057)	-0.1747 *** (0.0474)
初中	-0.4895 *** (0.0060)	-0.3256 *** (0.0473)
高中	-0.6056 *** (0.0073)	-0.4178 *** (0.0584)
大学及以上	-0.7954 *** (0.0098)	-0.5486 *** (0.0733)
民族	-0.2554 *** (0.0054)	-0.2398 *** (0.0313)
婚姻状况	0.5942 *** (0.0059)	6.5610 *** (0.2062)
流动	1.0881 *** (0.0040)	-0.2009 *** (0.0521)
已有曾生子女数		-2.3410 *** (0.0268)
常数项	-0.8508 *** (0.0211)	-15.0042 *** (0.3676)
流动的选择性		
年龄	-0.0171 *** (0.0026)	0.0335 *** (0.0031)
年龄平方	-0.0004 *** (0.0000)	-0.0010 *** (0.0000)
户口	0.3144 *** (0.0075)	0.0823 *** (0.0085)
受教育水平	0.1700 *** (0.0038)	0.1825 *** (0.0044)
民族	0.3507 *** (0.0107)	0.1916 *** (0.0125)
婚姻状况	-0.0409 *** (0.0097)	-0.1039 *** (0.0114)
常数	-0.9203 *** (0.0382)	-1.7133 *** (0.0446)
athrho		
_cons	-1.3211 *** (0.0042)	1.6600 *** (0.2767)
lnsigma		
_cons	-0.1991 *** (0.0014)	-6.3776 (14.9137)
N	306996	306996
chi2	422795.0190	18118.4451

表 3 流动人口与非流动人口的曾生子女数和普查前一年生育情况在各种匹配方法下的结果
Table 3 The Estimation of Various Propensity Score Matching Methods

方法	流动	非流动	差异	标准误	t 值	案例数	
曾生子女数							
匹配前	0.6937	1.4142	-0.7205 ***	0.00655	-109.99	流动	非流动
t 检验							
无设置 ^①	0.6937	0.6170	0.0767	0.0711	1.08	-	-
贪婪匹配: 最近邻居 1 ^②	0.6937	0.8888	-0.1951 ***	0.0071	-27.34	32549	32549
贪婪匹配: 最近邻居 2 ^③	0.6937	0.7995	-0.1056 ***	0.0071	-14.93	32540	32549
贪婪匹配: 马氏距离法 ^④	0.6937	0.7949	-0.1012 ***	0.0088	-11.56	32549	1338
以内核为基础的匹配 kernel ^⑦	0.6937	0.8229	-0.1292 ***	0.00627	-20.61	32547	274449
回归 ^⑥			系数	标准误			
贪婪匹配: 最近邻居 1 ^②			-0.1022 ***	0.0048			
贪婪匹配: 最近邻居 2 ^③			-0.1236 ***	0.0044			
贪婪匹配: 马氏距离法 ^④			-0.0984 ***	0.0043			
权重分析法 1 ^⑤ -ATE			-0.2001 ***	0.0072			
权重分析法 2 ^⑤ -ATT			-0.1489 ***	0.0031			
普查前一年的生育							
匹配前	0.0489	0.0293	-0.0197 ***	0.0010	-19.258	流动	非流动
t 检验							
无设置 ^①	0.0489	0.0445	0.0045	0.0190	0.24	32 549	274 449
贪婪匹配: 最近邻居 1 ^②	0.0489	0.0650	0.0161 ***	0.0018	8.8642	32 549	32 549
贪婪匹配: 最近邻居 2 ^③	0.0489	0.0657	0.0167 ***	0.0018	9.1965	32 549	32 548
贪婪匹配: 马氏距离法 ^④	0.0489	0.0539	0.0049	0.019	0.25	32 549	1 338
以内核为基础的匹配 kernel	0.0489	0.0546	0.0056 ***	0.0013	4.42	32 549	274 449
logistic 回归 ^⑥			Odds ratio	标准误			
贪婪匹配: 最近邻居 1 ^②			0.7432 ***	0.0312	-7.08		
贪婪匹配: 最近邻居 2 ^③			0.7621 ***	0.0314	-6.59		
贪婪匹配: 马氏距离法 ^④			0.7317 ***	0.0250	-9.13		
权重分析法 1 ^⑤ -ATE			0.7736 ***	0.0326	-6.09		
权重分析法 2 ^⑤ -ATT			0.7317 ***	0.0250	-9.13		

***: $p < 0.05$

注: ①“无设置”系指在 psmatch2 命令中不加任何其他设置而计算得到的结果^a;

- a 本命令程序为: psmatch2 floatp age agesqi hukoui. educi. nati. marriage i. rural2 ρout(ceb)。匿名评审专家指出: 本方法中的标准误可能未考虑匹配样本中存在的随机原则而引致的抽样误差。作者同意这种观点, 但本文使用这种简单的设置主要只是提供一个基础, 后续的分析采用不同方法, 并逐步加入各种设置, 以显示剔除选择性偏差以后的结果。如果使用 bootstrap 方法重新估计 psmatch2 中估计的标准误, 可能得到与其他后续分析相同的结果。

- ②以 0.25 个对数发生比($\log((1-p)/p)$) 的标准差为半径进行 1:1 的贪婪匹配;
- ③以 0.1 为半径进行 1:1 的贪婪匹配^b;
- ④马氏距离法中,由于匹配案例以加权形式出现,因此,对其利用 poisson 回归,此处的差异系数指方程中变量“流动人口”的系数,方程中的自变量与上述模型相同;
- ⑤权重分析法是以计算得到的倾向得分值为权重,对整个数据进行重新建立回归方程,由于权重不同,可计算得到平均实验效应 ATE 和实验组的平均实验效应 ATT 两种效应;
- ⑥回归分析的结果由于篇幅问题而未给出;
- ⑦此处以内核为基础的匹配 kernel 中所有的设置均为默认值,不加其他参数。

从方法角度来看,这种差异的下降是由于匹配过程仅利用了非流动人口中那些与流动人口的各种特征基本相同(即倾向得分基本相同)的部分案例,而未包括那些特征不相同的非流动人口案例。表 3 最后两列给出了流动与非流动人口在匹配以后可使用的案例数。本数据中,非流动人口的案例数为 274449 人;流动人口的案例数为 32549 人。贪婪匹配—最近邻居 1 所使用的非流动人口案例也为 32549 人;但在贪婪匹配—最近邻居 2 中,流动人口中有 9 人无法找到匹配;在马氏距离法中,非流动人口则仅挑选了 1338 人。

从本质上讲,差异减少的主要原因在于剔除了各种结构在流动人口与非流动人口之间的差异。或者说,上述结果说明,原来观察所得的流动人口与非流动人口的差异大部分是由流动人口的选择性导致的,而并不是真正的由生育水平的差异而导致的。正如李丁、郭志刚(2014)的文中对流动人口婚姻结构的调整。而本文的匹配则不仅控制、调整了婚姻结构,而且还控制调整了年龄、城乡、受教育水平等各种特征。

4.2 普查前一年生育情况的分析结果

普查前一年的生育情况事实上是妇女的一般生育率,因为分子是普查前一年是否生育,分母则是全部育龄妇女(以妇女人数代替妇女的平均生存人年数,即将妇女人数视为平均人年数)。而一般生育率受到了年龄结构等的影响,因此它并不能够真正地表示妇女的生育水平。这也是构建总和生育率指标的原因所在。但即使是总和生育率,它仅仅只是控制了年龄结构(或用孩次别总和生育率控制孩次的作用),而未能控制其他结构(如教育、婚姻结构等)。本文所使用的倾向得分匹配方法,类似于人口学中的标准化方法,即将其他包括年龄、婚姻状况、曾生子女数等结构进行标准化,然后再来比较一般生育率,以期讨论不同人口之间的生育水平。

表 3 的后半部分给出了普查前一年生育情况的分析结果,与曾生子女数的结果正好相反。未匹配以前,流动人口在普查前一年的生育水平(4.89%)显著高于非流动人口(2.93%),差异达近 2% 左右。但匹配以后,除无设置的倾向得分值匹配方法外,其他各种方法都得到基本一致的结果,不论是方向还是显著性水平。这说明流动人口的选择性导致目前观测到的流动人口的生育水平可能是有偏的,也即说明选择性对因变量的影响作用的存在。

其次,一致性的分析结论是:流动人口的生育率显著低于非流动人口。正如前文所述,经过匹配以后,流动人口与非流动人口在生育水平上的差异,由原来的负值转变为正值,即由未匹配前流动人口的生育水平高于非流动人口,转变为低于非流动人口。仍然以贪婪匹配—最近邻居法的第一种结果为例,普查前一年流动人口的一般生育率为 4.89%;如果非流动人口的各种特征与流动人口相同,则

b 本文中仅用了 1:1 匹配的方法。匿名评审专家指出,可以利用不同的匹配案例数(如 1:5 和 1:10)进行匹配,再检验所得结果。事实上笔者在模型建立过程时已试验性地比较了 1:1 和 1:5 两种匹配方法,所得结果无显著差异(即显著性与系数的符号),仅在具体估计值上存在差异。为简便起见,仅给出 1:1 匹配的结果。

他们普查前一年的一般生育率则由原来的 2.93% 提高到 6.50% ,高于流动人口的生育水平。这说明流动人口在普查前一年较高的一般生育水平可能是由各种结构特征引起的 ,而并不一定是其真正的生育水平。

再次 ,流动人口与非流动人口在普查前一年的一般生育率上的差异是显著的。即上述例子中非流动人口的 6.50% 与流动人口的 4.89% 之间的差异是显著的。

最后 ,如果再利用匹配样本 ,控制各种人口学特征后 ,流动人口在普查前一年的生育概率同样低于非流动人口(见表 3 的 logistic 回归部分)。仍然以贪婪匹配—最近邻居法的第一种结果为例 ,流动人口在普查前一年生育的发生比仅为非流动人口的 74.32% ,即其在普查前一年的生育可能性比非流动人口降低了 26% 左右。这也从另一方面再次说明流动人口的生育率低于非流动人口。

5 结论与讨论

本文以反事实因果框架为基础 ,利用第五次人口普查数据和倾向得分匹配方法 ,重新分析讨论了流动人口与非流动人口的生育水平差异问题。主要结论如下:(1) 由于人口流动的选择性 ,直接对流动人口生育水平的分析及其比较可能存在偏差;(2) 不论是以曾生子女数 ,还是普查前一年的生育情况为生育水平的测量指标 ,结果都表明 ,剔除流动人口的选择性以后 ,流动人口的生育水平显著低于非流动人口;(3) 在未控制选择性时流动人口与非流动人口在曾生子女数上的差异 ,大部分可归结为流动人口的选择性导致的 ,而并不能真正反映流动人口与非流动人口在生育水平上的差异;(4) 普查前一年生育情况 ,由原始数据中 ,流动人口高于非流动人口 ,在经过匹配以后转变为 ,流动人口低于非流动人口。这同样说明较高的流动人口生育水平可能是由其选择性引起的。

本文不论是方法还是结论 ,仍有值得商榷的地方 ,需要进一步深入讨论。

第一 ,生育指标问题。本文选用了曾生子女数和普查前一年生育情况这两个指标作为生育水平的指标。选用这两个指标的好处在于其特征:前者可以被视为计数变量 ,可以使用 poisson 回归等方法来做;后者则是虚拟变量 ,既可以被视为一般生育率 ,也可以用 logistic 回归等方法进行分析。但是 ,如果从人口流动与生育之间的关系来看 ,曾生子女数只表示到普查时点为止的生育情况 ,可能存在许多删截问题 ,即大部分的流动人口未完整地度过生育期 ,从而无法真正地表示人口流动对生育的影响作用。普查前一年生育情况尽管看似更符合人口学的生育指标 ,但是 ,其一从本质上讲 ,本文中所使用的是一个比例 ,而非真正的率;其二 ,从方法上讲 ,文中所用的均值(比例)既存在数值过小(如 logistic 回归要求取值为 1 的一类在总体中所占的比例至少要高于 5%)的问题 ,而且也有可能存在异方差(即不同类别的因变量方差不相等) ,从而导致结果存在偏差。因变量存在的这些问题需要我们审慎地对待上述结果。

第二 ,方法问题。倾向得分匹配方法是基于反事实因果推论的框架发展出来的 ,其本质是控制实验前的异质性后再讨论实验效应问题。其中稳定的干预值假设(Stable Unit Treatment Value Assumption, SUTVA) 是该方法的基础。本文中 ,尽管可以假定人口流动在短期内对流动人口的影响作用是相同的 ,但由于流动人口的各种背景及在流入地的生活与工作环境等都有可能不同;更为重要的是 ,流动人口在流入地的居住时间长短不同 ,导致人口流动对流动人口的个体作用可能是不同的 ,从而有可能违反假定条件而得到可能有偏的结果。其次 ,文中可能存在模型设定误差问题(Model Specification Error)。模型设定误差可能是定量研究中普遍存在的问题。但具体到本研究中 ,则可能涉及两个方面:一是流动人口选择性模型的识别与设定问题;二是实验效应干预模型的设定。

流动人口的选择性 ,不仅仅体现在本文所使用的这些人口学特征上 ,而且诸如家庭特征、迁出地社区、迁移网络、迁移流向等诸多因素同样会有影响作用。但文中未考虑这些变量 ,可能使倾向得分

模型(或选择性模型)存在设定误差,进而影响到具体结果。同时,利用倾向得分方法讨论实验干预效应时,理论上要求,干预效应模型中的自变量与倾向得分模型中的自变量至少有一个是不相同的。由于人口普查数据中变量本身就较少,因此在两个模型中直接加入了相同的自变量。尽管影响作用有限,但从理论上可能是存在问题的。

倾向得分匹配方法并不是万能的,其方法本身也存在缺陷。一方面,匹配过程中可能剔除了部分实验组或控制组的案例(特别是倾向值的两端),从而可能使得结果有偏;另一方面,即使忽略上述由于流入时间而引起的违反稳定干预值假设的情况,是否还存在讨论人口流动对生育的异质性影响作用的问题?即上述分析讨论的是实验(人口流动)前的异质性问题,而未能回答实验效应的异质性,即人口流动对哪些人会起到降低生育水平的作用,对哪些人的作用可能是相反的呢?这或许是今后的研究方向之一。

第三,本质上,本文并未能真正证明流动与生育之间的关系问题。从方法论来看,我们仍然是在寻找结果的原因(the causes of effects),而并不是真正地探讨原因的结果(the effects of causes)。一方面,我们所用的截面数据,本身就存在着原因与结果在同一时点的问题(当然,这是所有截面数据或观测数据的问题);另一方面,如果想要真正讨论人口流动对生育的影响作用,就需要以跟踪数据为基础(哪怕是回顾性调查数据),结合迁移与生育在时间上的关系、考察目前是迁移或流动的人、曾经有过流迁经历的人以及那些没有流迁经历的人这三类人在生育水平上的差异。本文既未能结合迁移与生育的时间关系,也受数据限制而未讨论那些曾经有过流迁经历的人与其他两类人群之间的差异,甚至未讨论迁移人口与流动人口之间的差异。而这些问题的研究将有助于更好地理解人口流迁与生育的关系。

参考文献/References:

- 1 Amara Mohamed. 2014. Multilevel Modelling of Individual Fertility Decisions in Tunisia: Household and Regional Contextual Effects. *Social Indicators Research* 1: 1 – 23.
- 2 Brand Jennie E. and Yu Xie. 2010. Who Benefits Most From College? Evidence for Negative Selection in Heterogeneous Economic Returns to Higher Education. *American Sociological Review* 2: 273 – 302.
- 3 Fu Alex Z. 倾向得分法综述. *中国药物经济学* 2008; 2: 27 – 34
Alex Z. Fu. 2008. Review on Propensity Score Matching Methods. *China Journal of Pharmaceutical Economics* 2: 27 – 34.
- 4 Heckman James J. 2005. The Scientific Model of Causality/Discussion: Response to Sobel. *Sociological Methodology* 1: 1 – 97.
- 5 Kulu Hill. 2005. Migration and Fertility: Competing Hypotheses Re-Examined. *European Journal of Population* 1: 51 – 87.
- 6 Kulu Hill and Elizabeth Washbrook. 2014. Residential Context, Migration and Fertility in a Modern Urban Society. *Advanced in Life Course Research* September: 168 – 182.
- 7 Mangalam J. J. 1968. *Human Migration: A Guide to Migration Literature in English, 1955—1962*. Lexington: University of Kentucky Press.
- 8 Massey Douglas S., Joaquin Arango, Graeme Hugo, Ali Kouaouci, Adela Pellegrino and J. Edward Taylor. 1993. Theories of International Migration: A Review and Appraisal. *Population and Development Review* 3: 431 – 466.
- 9 Morgan Stephen L. and Christopher Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press.
- 10 Morgan Stephen L. and David J. Harding. 2006. Matching Estimators of Causal Effects: Prospects and Pitfalls in Theory

- and Practice. *Sociological Methods & Research* 1: 3 – 60.
- 11 Ravenstein E. G. 1885. The Laws of Migration. *Journal of the Royal Statistical Society* 2: 167 – 227.
 - 12 Xie ,Y. 2013. Population Heterogeneity and Causal Inference. *Proceedings of the National Academy of Sciences* 16: 6262 – 6268.
 - 13 Xie Yu and Xiaogang Wu. 2005. Market Premium ,Social Process ,and Statisticism. *American Sociology Review* 1: 865 – 870.
 - 14 Zhao Yaohui. 1999. Labor Migration and Earnings Differences: The Case of Rural China. *Economic Development and Cultural Change* 4: 767 – 782.
 - 15 陈卫 吴丽丽 . 中国人口迁移与生育率关系研究 . 人口研究 2006; 1: 13 – 20
Chen Wei and Wu Lili. 2006. Research on the Relationships between Migration and Fertility in China. *Population Research* 1: 13 – 20.
 - 16 段成荣 . 影响我国省际人口迁移的个人特征分析 . 人口研究 2000; 4: 14 – 22
Duan Chengrong. 2000. Individual Level Determinants of Interprovincial Migration in China: On the Effects of Time Sequence in Migration Studies. *Population Research* 4: 14 – 22.
 - 17 段成荣 杨舸 张斐 卢雪和 . 改革开放以来我国流动人口变动的九大趋势 . 人口研究 . 2008 6: 30 – 43
Duan Chengrong ,Yang Ge Zhang Fei and Lu Xuehe. 2008. Nine Trends of Changes of China's Floating Population Since the Adoption of the Reform and Opening-up Policy. *Population Research* 6: 30 – 43.
 - 18 郭申阳 . 因果分析高级讲座: 倾向值分析及相关模型 . 北京: 北京大学社会学系暑假班 . 2014
Guo Shenyang. 2014. Advanced Topics in Casual Inference: Propensity Score and Related Models. Summer School of Sociology Department ,Peking University.
 - 19 郭申阳 马克 W 弗雷泽 . 倾向值分析: 统计方法与应用 . 重庆大学出版社 2012
Guo Shenyang and Frasher ,Mark W. 2012. Propensity Score Analysis: Statistical Methods and Applications. Chongqing: Chongqing University Press.
 - 20 郭志刚 . 时期生育水平指标的回顾与分析 . 人口与经济 2000; 1: 3 – 10
Guo Zhigang. 2000. Review and Analysis on the Indicators of Period Fertility. *Population & Economics*1: 3 – 10.
 - 21 郭志刚 . 流动人口对当前生育水平的影响 . 人口研究 2010; 1: 19 – 29
Guo Zhigang. 2010. Impacts of Floating Population on Current Fertility in China. *Population Research* 1: 19 – 29.
 - 22 李丁 郭志刚 . 中国流动人口的生育水平——基于全国流动人口动态监测调查数据的分析 . 中国人口科学 . 2014 3: 17 – 29
Li Ding and Guo Zhigang. 2014. Fertility Level of Migrant Population: Based on National Migrant Population Dynamic Monitoring Survey Data. *Chinese Journal of Population Science* 3: 17 – 29.
 - 23 李玲 . 改革开放以来中国国内人口迁移及其研究 . 地理研究 2001; 4: 453 – 462
Li Ling. 2001. Internal Population Migration in China since the Economic Reform: A Review. *Geographical Research* 4: 453 – 462.
 - 24 石人炳 熊波 . 迁移流动人口生育特点及相关理论——中外研究述评 . 人口与发展 2011; 3: 73 – 79
Shi Renbing and Xiong Bo. 2011. Review on the Impacts of Migration on Fertility. *Population & Development* 3: 73 – 79.
 - 25 王天夫 . 社会研究中的因果分析 . 社会学研究 2006; 4: 132 – 156
Wang Tianfu. 2006. Causal Analysis in Social Research. *Sociological Studies* 4: 132 – 156.
 - 26 谢宇 . 2006: 社会学方法与定量研究 北京: 社会科学文献出版社
Xie Yu. 2006. *Sociological Methodology and Quantitative Research*. Beijing: Social Science Academic Press.
 - 27 杨云彦 . 80 年代中国人口迁移的转变 . 人口与经济 1992; 5: 12 – 16
Yang Yunyan. 1992. The Transition of Population Migration in China in 1980s. *Population and Economics* 5: 12 – 16.
 - 28 周皓 . 中国人口迁移研究: 中国人民大学博士论文 2000

- Zhou Hao. 2000. Study on Migration in China ,PhD Thesis ,Renmin University of China.
- 29 周皓. 中国人口迁移与家庭户研究. 北京大学社会学系博士后出站报告 2003
Zhou Hao. 2003. Study on Family and Migration in China. Thesis for Post-doctoral Research ,Peking University.
- 30 周皓. 中国人口迁移的家庭化趋势及影响因素分析. 人口研究 2004; 6: 60 - 69
Zhou Hao. 2004. Family Migration in China: Trends and Determinants. Population Research 6: 60 - 69.
- 31 周皓 梁在. 中国的返迁人口: 基于五普数据的分析. 人口研究 2006 3: 61 - 69
Zhou Hao and Liang Zai. 2006. Return Migration in China: An Analysis of the Fifth Census Data. Population Research 3:
61 - 69.

(责任编辑: 宋 严 收稿时间: 2014 - 11)