

# 结构流动：基于调整—分解的分析策略<sup>\*</sup>

胡安宁

**提要：**本文系统梳理了结构流动研究在理论支撑、混淆因素控制、概念化、统计模型复杂度以及统计推断等方面的局限，并提出了调整—分解的分析策略。通过综合采用德明—斯蒂芬算法和流动表的三元分解法，这一分析策略直接针对流动表的边际分布构建“反事实”流动表，并以非建模对比的方式呈现结构流动、交换流动以及二者的交互作用。除此之外，潜在的混淆因素可以通过调整的方法予以控制，相关的统计推断可以通过参数分布建模或者自助法来实现。本文通过两个经验示例对这一分析策略进行了呈现。

**关键词：**结构流动 社会流动 调整 分解

## 一、引言

社会流动过程中个体境遇的改变既是因为行动者个人的努力和选择，又与宏观社会环境的变化息息相关。例如，伴随着产业升级和大规模城市化的进程，大量的“农家子弟”成为城市白领，实现代际间的职业阶层跃迁（Bearman & Deane, 1992）。同理，高等教育扩招也会给大量家庭带来“第一位大学生”，从而实现代际之间向上的教育流动（Breen, 2010）。将社会流动过程中的结构性因素抽离出来单独进行考察的研究需求催生了社会流动研究中的“结构流动—交换流动（structural mobility-exchange mobility）”分析框架（Breen, 1985；Cobalti, 1988）。基于这一分析框架，一个社会的总体流动情况可以进一步细分为结构流

---

\* 本文得到国家社会科学基金万人计划哲学社会科学领军人才项目（22VRC140）的支持。感谢匿名评审所提出的专业修改意见。限于篇幅，文章有所删减。文责自负。

动和交换流动两种类型,前者可以定义为由不同阶层群体的宏观“分布”差异带来的流动,后者则是指在总体流动中“扣除了”结构流动之后所剩余的不同社会阶层位置之间的交换(Sobel et al.,1985)。

“结构流动—交换流动”的二分框架是“二战”以后直至20世纪70年代社会流动问题研究的主流分析框架。但是,自20世纪70年代末以来,社会流动研究的理论兴趣逐渐发生转移,从探究结构流动转向考察以发生比(odds ratio)为测量指标的相对流动(relative mobility)及其变化模式(Breen,1985)。此时,体现结构流动的边际分布只是在考察绝对流动(absolute mobility)时使用。结构流动这一概念逐渐淡出了学者们的视野。

但是,结构流动反映了形塑个体生活境遇变化和流动过程的一个不可忽视的社会性力量,对其进行考察仍然具有重要的理论和现实意义。就学术价值而言,与一般的量化社会学研究类似,当下社会流动研究往往呈现陈云松(2022)所指出的“微观旨趣”的路径依赖,对于结构性的、宏观性的影响因素关注不足。一个具体的表现是,虽然几乎没有学者会否认结构变迁对于社会流动模式的影响,但是鲜有研究在经验层面上具体呈现结构流动的影响力。随着高质量个体数据的不断积累,学者们实际上获得了丰富的信息资源,可以使用多种策略来探究结构流动。在这个意义上,在社会流动研究中使用新的经验分析策略、恢复对结构流动—交换流动的考察是重拾社会学传统的“宏观观照”,对当下的社会流动研究有所补益。就现实价值而言,在转型期的中国社会中,个体先赋地位和自致地位之间的变化体现了诸如“改革红利”“人口红利”等社会宏观结构性变迁(干春晖、郑若谷,2009;魏杰、汪浩,2016;蔡昉,2020)。因此,结构流动研究可以联结宏观社会层面的转型和微观个体资源获取模式的变迁,反映宏观“全局性”政策调整可能达成的社会效果,为理解转型时期中国社会结构变迁的影响提供一个独特的评估路径。因此,如果说相对流动呈现的是针对特定行业或者领域的流动藩篱(高勇,2009),结构流动揭示的则是更为宏大的结构性藩篱(或者机会)。可见,结构流动仍是社会学研究中应当着力考察的一个重要议题。

综上所述,社会流动研究在借鉴国际优秀成果的同时,不应急切地抛弃结构流动的相关议题。然而,既有结构流动的经验分析策略在理论支撑、混淆因素控制、概念化、模型设定以及统计推断等多个方面都存在局限性。因此,为了更好地研究结构流动,本文在系统梳理既有结构流动分析方法的基础上,提出一个调整—分解的分析策略,并通过具体经验示例来展现其价值。

## 二、结构流动:一个应该被“放弃”的概念?

如前文所述,有关结构流动以及结构流动和交换流动的区分框架经历了一个逐渐式微的过程。在1983年的一篇论文中,社会学者索贝尔(Michael E. Sobel)甚至建议“研究者需要放弃‘结构 vs. 流动’的分析框架”(Sobel, 1983: 722)。结构流动研究式微的原因可以归结为以下三方面:首先,从概念设定上讲,索贝尔以及之前的很多学者所谈及的结构流动,指的是单一流动表的行边际分布和列边际分布之间的差异。研究者之所以关注单一流动表,在一定程度上是因为早期社会流动研究所用分析资料多是官方统计资料,故而研究人员所能确定的信息往往只有在特定的时点上根据两代人的职业特征或者教育水平所构建的列联表。与之伴随的是,很多与社会流动相关的测量指标也是针对单一流动表设计的(Blau & Duncan, 1967)。其次,从测量工具上讲,基于单一流动表来估计结构流动的方法有其局限性。索贝尔就对霍普(Keith Hope)的中途模型(Hope, 1982)(具体技术细节,参见下一节)提出了尖锐的批评(Sobel, 1983)。尽管针对索贝尔的批评也有一系列的反驳声音,但是经历了这些批评之后,加之基于发生比的对数线性模型逐渐兴起(Breen, 1985),学者们逐渐把注意力从“结构流动—交换流动”的分析框架转向适配对数线性模型的“相对流动—绝对流动”的分析框架。最后,从社会现实上讲,结构流动所代表的宏观社会变迁在发生后的一段时间内基本会维持在一个比较稳定的状态。与之相比,由于涉及社会阶层之间壁垒的微观变化,在给定结构特征后,基于发生比的相对流动会持续变化。因此,对于社会科学研究者而言,对结构流动的探究在一段时间之内有可能“毕其功于一役”。尤其是在既有研究关注较多的欧美国家中,随着社会职业与教育结构日趋稳定,结构流动逐渐成为一个“常量”。

由于上述多重原因,结构流动概念及其研究方法的影响力逐渐弱化。但是,这并不意味着结构流动本身是一个应该被放弃的学术概念(Slomczynski & Krauze, 1984)。我们针对上述三重式微原因,同样可以发现一系列“复兴”结构流动概念的有利条件。首先,从概念界定上来讲,由于大规模调查资料的不断累积,我们现在完全可以将传统单一流动表的行—列差异分析拓展到多个流动表的对比分析,同时借助个体数据来完成统计推断的工作。在这方面,一些早期的研究可资借鉴(Hazelrigg, 1974; Hazelrigg & Garnier, 1976)。其次,与

上面提到的多流动表拓展相呼应,研究人员也有可能超越既有方法策略,寻求和开发更为直观便捷的分析工具。这正是本文的研究主题,下面会展开讨论。最后,就社会现实而言,处于转型期的中国社会充分体现了社会结构转型如何形塑社会流动状态。正如白尔曼(Peter S. Bearman)和迪恩(Glenn Deane)所言,“将流动机制中的结构性和交换性成分区分开来,是理解流动在其历史背景下的社会和政治后果的基本前提”(Bearman & Deane, 1992:31)。这一判断与关于中国社会转型的社会学研究既有结论一致(例如 Whyte, 2010; Wu, 2011; Xie, 2016)。因此,无论是更好地理解当下中国社会的整体流动状态,还是进一步通过改革来提升社会整体的开放和包容程度,都离不开对结构流动的分析与考察。

### 三、结构流动的研究方法梳理

在现有文献中,针对结构流动的经验分析策略大致可以分为三类。为了便于讨论,我们将其分别命名为指数法、模型参数法和计算法。在这一部分的讨论中,我们将梳理这三类分析策略中具有代表性的方法。

#### (一)指数法

指数法是指通过建构一个特定的指数来衡量结构流动。由于对于单一流动表而言,结构流动即行边际分布和列边际分布的差异,因此一个非常直观的测量指数便是差异指数(index of dissimilarity)(Cobalti, 1988)。以教育流动为例,假设某一流动表可以表示为一个 $k \times k$ 的矩阵( $k$ 为教育类别)。样本总量为 $n_{..}$ ,针对某一特定教育水平 $i$ 的行加总表示为 $n_{i.}$ ,列加总为 $n_{.i}$ 。差异指数就是每一个教育水平 $i$ 的行—列边际差异 $n_{i.} - n_{.i}$ 的绝对值加总后除以两倍的总样本量。用 $\Delta$ 来表示差异指数,我们可得 $\Delta = \sum_{i=1}^k |n_{i.} - n_{.i}| / 2n_{..}$ 。差异指数的计算过程虽然简单直观,但是从总流动人口中减去因为边际分布差异带来的流动人口数的操作却掩盖了因为结构流动所带来的交换流动。换言之,受结构流动影响的那部分交换流动在差异指数的计算和使用过程中被归为结构流动,因此无形中会低估交换流动而高估结构流动(Cobalti, 1988)。

除了差异指数,另外一个常用的指数是流动比率(mobility ratio)。这一统计量常被用于衡量剥离了结构流动效应之后的交换流动水平,因此也间接与结构

流动的测量相关。具体而言,对于流动表第  $i$  行第  $j$  列的个体而言,其流动比率是假设行与列之间彼此独立(或者称为完全流动[perfect mobility])的情况下个体期望的频次数  $f_{ij}$ 。显然,既然行与列之间彼此独立,那么个体出现在第  $i$  行第  $j$  列的概率等于其出现在第  $i$  行的概率乘以其出现在第  $j$  行的概率。前者可以表示为  $n_{i\cdot}/n_{\cdot\cdot}$ ,后者可以表示为  $n_{\cdot j}/n_{\cdot\cdot}$ 。因此个体出现在第  $i-j$  单元格的概率就是二者的乘积,其期望频次就是  $n_{\cdot\cdot} \times n_{i\cdot}/n_{\cdot\cdot} \times n_{\cdot j}/n_{\cdot\cdot} = n_{i\cdot} \times n_{\cdot j}/n_{\cdot\cdot}$ 。流动比率就可以表示为  $f_{ij} = n_{i\cdot} \times n_{\cdot j}/n_{\cdot\cdot}$ 。如果说流动比率代表剔除了结构流动效应后的交换流动效应,而观测到的流动表背后既有结构流动也有交换流动的效应,那么结构流动就可以用二者之间的差异来表示。如果我们用  $x_{ij}$  来表示流动表的第  $i$  行第  $j$  列的某一观测值,那么对于第  $i-j$  单元格而言,其结构流动就可以表示为  $|x_{ij} - f_{ij}|$ 。

流动比率自提出之后就受到多方面的批评。例如,布劳(Peter M. Blau)和邓肯(Otis Dudley Duncan)指出,流动比率并没有把流动表的边际差异完全剔除(Blau & Duncan, 1967)。这是因为其计算过程本身涉及边际分布的频数。也正如豪特(Michael Hout)所言,流动比率作为一个统计指标,既反映了边际分布的规模,也反映了行与列之间的关联性(Hout, 1983)。因此,流动比率并不是单纯反映交换流动,而是混合了结构流动和交换流动。

有必要指出的是,除了差异指数和流动比率之外,还有一些在其基础上开发的修正指标,例如由两种交换流动的比率构成的 Yasuda 指数(Yasuda, 1964),具体表示为  $Y = \sum_{i,i \neq j} [n_{ij} - (\sum_i |n_{i\cdot} - n_{\cdot i}|/2)] / \sum_{i,i \neq j} [f_{ij} - (\sum_i |f_{i\cdot} - f_{\cdot i}|/2)]$ 。其中, Yasuda 指数的分子是观测到的流动表中去除结构流动后的交换流动,而分母是在完全流动(假设行与列彼此独立)的情况下去除结构流动后的交换流动。但由于其构成要素本身结合了差异指数和流动比率,后两者的局限性自然会体现在 Yasuda 指数之中。

## (二) 模型参数法

模型参数法是指通过拟合特定的统计模型,以模型的参数作为结构流动或者交换流动的测量。比较有代表性的是霍普的中途模型(halfway model)(Hope, 1982)以及索贝尔等人提出的拟对称模型(quasi-symmetry model)(Sobel et al., 1985)。

霍普所提出的中途模型立足于对流动表第  $i-j$  单元格预期频次的特殊参数化。具体而言,中途模型以完全流动模型(perfect mobility model)为基础进行参数

变换。如上文所言,完全流动模型是指第  $i-j$  单元格的预期频次仅取决于行效应  $a_i$  和列效应  $b_j$ ,但是不取决于行与列之间的交互效应。之所以称其为完全流动,是因为在此模型设定下,先赋地位(行变量)不影响自致地位(列变量),故而个体流动到哪一个自致地位类别都有可能。在这一模型设定中,预期频次  $f_{ij}$  可以表示为  $f_{ij} = a_i b_j$ 。基于此,霍普对完全流动模型进行了再参数化,即  $f_{ij} = a_i b_j = a_i a_j b_j / a_j$ 。在这个参数设定下,所谓的中途模型指前半部分(H 表示 halfway)  $f_{ij}^H = a_i a_j$ 。中途模型提出了一个独特的构建预期频次  $f_{ij}^H$  的方法。如上文所示,传统的流动比率通过假设行与列相互独立来建构预期频次,即  $f_{ij} = n_{i.} \times n_{.j} / n_{..} \times n_{.j} / n_{..}$ 。但是中途模型的计算为  $f_{ij}^H = n_{..} \times (n_{i.} + n_{.j}) / 2n_{..} \times (n_{i.} + n_{.j}) / 2n_{..}$ 。两相对比,基于中途模型构建出的流动模式,其行与列的边际分布一致,且生成的流动表内部频次分布对称。同时,中途模型不涉及行与列之间的交互,也没有包含行与列之间的关联。基于这些原因,霍普认为,以中途模型构建的流动表代表了去除结构因素的交换流动。与此同时,完全流动模型在去除中途模型之后剩余的部分被称为差异模型(difference model),表示为  $f_{ij}^D = b_j / a_i$ 。显然,差异模型代表了列效应与行效应之间的比值差异,因此被用以指代结构流动。

霍普的中途模型将完全流动模型分解为结构流动和交换流动两个部分,具有创新性。但是,该模型自提出之后也被多方质疑。其中比较有代表性的便是上文提到的索贝尔的研究(Sobel, 1983)。他指出,中途模型能够用来指代交换流动的前提是完全流动模型可以很好地拟合数据,但是这一点并不必然能够被保证,甚至在大多数社会学研究中根本无法保证。豪特也从模型拟合的角度对中途模型进行了批评。他认为,如果行与列之间确实存在某种关联,完全流动模型的拟合度就会很差,此时采用中途模型就会低估结构流动(Hout, 1989: 89)。除此之外,针对中途模型的另一批评是,在以完全流动模型减去中途模型所剩余的差异模型中,其构建的流动表并不对应于观测到的行与列的边际分布差异。因此,差异模型的效度也受到质疑(Marks et al., 1996)。

当然,对于索贝尔和豪特基于模型拟合度对中途模型提出的批评也存在不同的意见。例如,琼斯(Frank L. Jones)认为,中途模型将完全流动模型作为基准并不是因为完全流动模型是一个统计模型,能够以数据拟合优劣予以评判,而是因为完全流动模型作为一种理论模型代表了“机会公平”这一理论上的“基准状态”。因此,在经验研究中,完全流动模型是一个很好的理论参照模型(Jones, 1985)。除此之外,斯洛姆钦斯基(Kazimierz M. Słomczynski)和克劳兹(Tadeusz Krauze)认为索贝尔的批评有失偏颇。一则,索贝尔主要批评的对象是霍普的中

途模型,但是不能因为一个方法存在局限就否定其他所有关于结构流动的经验分析方法。再则,即使当时开发出的所有方法都有问题,也不意味着“结构流动—交换流动”的这个分析框架就应该被放弃(Slomczynski & Krauze, 1984)。斯洛姆钦斯基和克劳兹的批评对于本文颇有启示意义。

虽然索贝尔在 1983 年的论文中主张放弃“结构流动—交换流动”的这个分析框架,但是索贝尔本人并没有放弃探究结构流动。模型参数法中的另外一个模型正是索贝尔及其同事提出的拟对称模型。具体而言,拟对称模型预设了一种边际分布不对称但内部频次分布对称的流动表结构。索贝尔等人提出,为了描述这种流动模式,流动表中第  $i$  行第  $j$  列的预期频次可以写成  $f_{ij} = \alpha_j \beta_j \gamma_{ij} \delta_{ij}$  (Sobel et al., 1985)。在这个设定中,  $\alpha_j$  是指造成列边际分布与行边际分布差异的非对称结构效应,其无差别地作用于行变量的不同取值水平之上,因此其区分仅在于列变量的不同取值水平  $j$ 。在拟对称模型中,对于  $\alpha_j$  的限制条件是  $\prod_{j=1}^k \alpha_j = 1$ 。此外,  $\beta_i$  与  $\beta_j$  被称为对称结构效应,其本质上就是霍普中途模型的设定(Ultee & Luijckx, 1986)。显然,如果  $i = j$ ,那么我们有  $\beta_i = \beta_j$ 。因此,  $\beta_i \beta_j$  体现出的是对称的结构效应。除了这些参数之外,  $\gamma_{ij}$  为非对称的行—列关联效应,  $\delta_{ij}$  为对称的行—列关联效应。  $\gamma_{ij}$  和  $\delta_{ij}$  的区别在于,后者在非对角线元素上是对称的。如果拟对称模型可以很好地拟合数据,那么我们便有理由认为该流动表除了边际分布不对称之外,其内部应当是对称的,故而  $\gamma_{ij} = 1$ 。此时,这一流动表背后的结构流动效应可以表示为  $\alpha_j$ , 而交换流动则表示为  $\beta_i \beta_j \delta_{ij}$ 。

拟对称模型自提出之后,在很多社会流动议题中得到了应用(例如 Bearman & Deane, 1992; Hu & Leamaster, 2015)。但是这个模型比一般意义上的关联模型要复杂很多,且其模型设定也往往缺乏具有说服力的经验或者理论依据。例如,通常而言,交换流动被定义为剔除结构流动后的剩余流动部分,但是拟对称模型所估计出的交换流动要求  $i - j$  单元格和  $j - i$  单元格之间存在等量流动。在操作层面,交换流动也不仅仅通过给定边际分布下行与列之间的相关关系来体现,而是包含了对称化的边际分布  $\beta_i \beta_j$ 。换言之,交换流动是对称关联效应  $\delta_{ij}$  和对称边际效应  $\beta_i \beta_j$  的交互项。这些设定背后的实质性理由并不是很清晰。与上文讨论的通过边际分布差异来理解结构流动的分析视角不同,拟对称模型主张结构流动由边际分布差异  $\alpha_j$  和行与列之间非对称相关关系  $\gamma_{ij}$  组成。但是,我们并不是很清楚非对称相关关系  $\gamma_{ij}$  为何会影响结构差异。此外,正如索贝尔在对霍普中途模型的批评中所指出的那样,在拟对

称模型的设定下,为了将  $\alpha_j$  理解为结构分布的测量,拟对称模型必须很好地拟合数据以保证  $\gamma_{ij} = 1$ 。但是,这同样是一个经验问题,并不能预先确定。最后,对于结构流动的参数  $\alpha_j$ ,设置限制条件  $\prod_{j=1}^k \alpha_j = 1$  的现实依据也不清晰。这使得拟对称模型分析结果的可解释性较差。学者们不得不采用一些替代手段对结果进行诠释(例如,计算不同列变量取值水平下  $\alpha_j$  的比值,参见 Bearman & Deane, 1992; Hu & Leamaster, 2015)。

### (三) 算法

如果说 20 世纪 50—70 年代见证了结构流动研究中指数法逐渐向模型参数法的过渡,那么这部分要介绍的算法在这一时期可谓是独树一帜。虽然后续采用这种方法的研究很少,但该方法体现了不同于上述两种分析路径的独特思路。具体而言,所谓的算法是指基于一定的前提假设,通过“计算”手段来建构结构流动和交换流动的矩阵。这方面的代表性研究来自克劳泽和斯洛姆琴斯基(Krauze & Slomczynski, 1986)。假设一个特定的流动表矩阵  $N$ ,其第  $i$  行第  $j$  列的构成元素表示为  $n_{ij}$ 。那么,算法设定了  $N$  由三部分构成:对角线元素所构成的不流动矩阵  $I$ 、结构流动矩阵  $S$  和交换流动矩阵  $C$ 。显然,如果我们有一个指示变量  $\delta_{ij}$ ,其对于对角线单元而言取值为 1,非对角线单元取值为 0,则可以得出  $I = \delta_{ij} \times n_{ij}$ 。

为了求得结构流动矩阵  $S$ ,克劳泽与斯洛姆琴斯基设定了两个基本限制条件:(1)如果结构流动矩阵存在,那么结构流动矩阵的边际分布变化与实际观测到的流动表的边际分布变化相等,即  $s_{i.} - s_{.j} = n_{i.} - n_{.j}$ ;(2)由于实际观测到的流动表是结构流动和交换流动共同作用的结果,因此对于特定的  $i-j$  单元而言,结构流动矩阵的元素  $s_{ij}$  取值应当大于或等于 0,但小于实际观测到的元素  $n_{ij}$  的取值,即  $0 \leq s_{ij} \leq n_{ij}$ 。基于这两个限制条件,结构流动可以表示为结构流动矩阵的元素之和,用  $\nabla$  表示,即  $\nabla = \sum_{i=1}^k \sum_{j=1}^k s_{ij}$ 。求解  $s_{ij}$  可以通过线性规划算法完成,即在给定  $s_{i.} - s_{.j} = n_{i.} - n_{.j}$  和  $0 \leq s_{ij} \leq n_{ij}$  的条件下,最小化  $\nabla$  (Slomczynski & Krauze, 1984)。这种基于线性规划算法的分析,也是我们将其称为算法的原因。对于交换流动矩阵  $C$  的构成元素  $c_{ij}$ ,可以同样采用线性规划的方法完成计算。具体而言,交换流动矩阵  $C$  的对角线元素为 0,非对角线元素应当小于其所对应位置的观测矩阵元素  $n_{ij}$ 。这一要求可以表示为  $0 \leq c_{ij} \leq n_{ij} (1 - \delta_{ij})$  (Slomczynski & Krauze, 1984)。此外,克劳泽与斯洛姆琴斯基设定,交换流动矩阵  $C$  的流出和流入应当保持平衡,即  $\sum_{j=1}^k c_{ij} = \sum_{j=1}^k c_{ji}$ 。在以上限制条件下,通过

最大化交换流动矩阵元素之和  $\sum_{i=1}^k \sum_{j=1}^k c_{ij}$ , 我们可以获得  $c_{ij}$  的计算结果 (Slomczynski & Krauze, 1984)。

针对算法, 索贝尔等人主要提出了三点批评: 第一, 由于算法直接通过线性规划的手段生成相应的结构流动矩阵和交换流动矩阵, 因此其生成过程主要依赖于计算手段, 相应的计算过程所设定的条件也是为了满足实现最优化计算的目标, 其理论意义不是很明确; 第二, 将结构流动和交换流动各自界定为一个矩阵的做法和社会流动相关理论不是很匹配, 且以矩阵来衡量结构与交换流动在操作上缺乏便利性; 第三, 由于计算过程直接以既有数据作为计算的基础, 直接生成的结构流动与交换流动矩阵没有考虑到抽样过程所带来的统计误差 (Sobel et al., 1986)。

综上所述, 在分析结构流动问题的现有文献中, 社会学者或通过建构一种全局性的指标, 或通过拟合结构模型, 或通过算法来达成对结构流动以及交换流动的经验测量。但是正如表 1 所示, 无论哪一种方法都存在一系列的局限性。因此, 为了更好地分析结构流动, 我们需要有针对性地对已有方法的局限提出应对策略。

表 1 三类结构流动分析方法及其局限性

类别	描述	方法举例	局限性
指数法	通过建构特定的指数来指代结构流动或者控制结构流动后的交换流动	差异指数	夸大结构流动而低估交换流动
		流动比率	单一指数同时反映了结构流动和交换流动
模型参数法	以特定模型参数或者参数的组合来指代结构流动	中途模型	<ol style="list-style-type: none"> <li>1. 不能很好地衡量控制结构流动后的剩余流动</li> <li>2. 中途模型和完全流动模型之间的差异并不对应边际分布的差异</li> <li>3. 在行与列存在关联时, 中途模型会低估结构流动</li> </ol>
		拟对称模型	<ol style="list-style-type: none"> <li>1. 对结构流动和交换流动的定义限制较大</li> <li>2. 难以解释非对称流动何以影响结构</li> <li>3. 模型的限定条件难以解释</li> </ol>
算法	在特定限制条件下, 通过线性规划构建结构流动和交换流动矩阵	通过线性规划来计算结构流动矩阵与交换流动矩阵	<ol style="list-style-type: none"> <li>1. 缺乏理论支撑</li> <li>2. 操作化比较随意</li> <li>3. 未考虑抽样误差</li> </ol>

## 四、调整—分解的分析策略

### (一) 结构流动分析需要解决的方法问题

综合上文的方法梳理,在研究结构流动时,我们需要面对和解决以下五个方面的问题。

第一个问题是理论支撑问题。所谓理论支撑是指用以测量结构流动的指标是否和通行的理论主张相一致。现有研究的一个基本共识是流动研究中的结构流动需围绕流动表的边际分布展开。因此,在建构结构流动的测量指标时,一个基本要求是该指标能够体现流动表的边际分布变化。但无论是中途模型还是拟对称模型,其复杂的参数设定都不能直观地满足这一要求,因而在方法上存在共同的局限。

第二个问题是混淆因素问题。传统社会流动研究更多是一种描述性探索,虽然隐含了先赋性因素对自致性因素的因果性推论,但对于这种因果性并没有严格的经验识别。但是,随着过去几十年社会科学因果推断方法的推广,流动分析中先赋性因素对自致性因素的因果效应越发得到学界的关注(例如 Hu & Wu 2021)。为此,研究者在考察流动表行变量对列变量的影响的同时,越发需要控制潜在的混淆因素。正如戈德索普(John H. Goldthorpe)等人所言:“我们并不能够假定那些影响职业结构变迁的因素……与影响交换流动模式的因素互不相关”(Goldthorpe et al.,1978:64)。因此,在分析结构流动问题时,一些混淆因素需要予以控制,但前文所提及的方法均没有考虑这一问题。

第三个问题是概念化问题。无论是指数法还是中途模型,一个共有的方法局限在于无法在完全剔除交换流动的前提下考察结构流动。这一方法局限的根源之一或许在于结构流动的概念化局限。在传统的结构流动分析中,结构流动、交换流动和不流动三者形塑了特定的流动表。按照这种概念化路径,交换流动和结构流动之间泾渭分明,各自代表了一类流动的主效应。此时,结构流动和交换流动之间的互相影响被忽略了。实际上,那些可以在交换流动中占据优势地位的群体本身也有可能结构变迁过程中占据优势地位。这些群体的存在会让整体的社会流动过程背后包含结构流动和交换流动的交互作用(Breen,1985)。显然,对结构流动的概念化并没有考虑两种流动类型的交互作用。

第四个问题是统计模型的限制。虽然索贝尔等人偏向于通过结构模型来探究结构流动,但正如上文所述,过于复杂的结构模型本身反而会给实际研究带来

多重限制。例如,当模型设定复杂时,其解释度就会比较差,而以结构模型来分析流动问题本身也可能会遇到模型的拟合度差的问题。此外,和其他基于结构模型的量化研究类似,结构模型作为一种统计模型,本身会遇到模型的不确定性问题,即同一个流动表数据有可能存在多个拟合度好的结构模型(胡安宁,2017)。由于这些模型的参数设定不同,它们所呈现的实质性信息可能彼此矛盾,但这种矛盾性并不能通过比较模型的拟合度高低来解决。

最后一个问题是统计推论上的局限。正如索贝尔等人对算法的批评所言,在当下的流动表分析中,一个流动表中的相关信息往往来自抽样数据,因此自然会受到抽样误差的影响。换言之,我们不能将其中的数字看成是固定的值,而应看作一个包含抽样不确定性的值。但是算法和指数法均没有很好地考虑这种随机抽样带来的误差,因此在分析的完备性上存在缺陷。

综上所述,我们在研究结构流动时,需要对理论支撑、混淆因素控制、概念界定、模型设定和统计推论这五个方面的局限性进行处理。如表2所示,由于理论支撑问题涉及的是结构流动测量和边际分布的对应关系,因此一个解决思路是在构建结构流动测量的时候“直接”针对边际分布进行处理。混淆因素问题则要求在流动表的比较过程中,除了控制行变量和列变量的边际分布特征之外,还需要对其他潜在的混淆因素进行统计控制。围绕概念界定,我们需要在考察结构流动和交换流动的主效应时引入二者的交互效应,而这一过程需要尽量避免结构模型带来的诸多复杂限制。概念界定和模型设定上的限制可以通过分解的方法解决。最后,统计推断问题可以通过设定流动表中相关频次的统计分布或者自助法(bootstrap)来解决(DiCiccio & Efron,1996)。下面我们就详细介绍调整和分解的策略。

表2 结构流动分析的方法问题及其解决思路

问题	解决思路	具体解决办法
理论支撑:结构流动的测量与边际分布的对应关系	直接针对边际分布进行处理	调整
混淆因素:区分结构变化和结构背后的其他因素变化	控制边际分布变量之外的其他混淆变量	
概念界定:结构流动与交换流动的刚性区分	考虑结构流动与交换流动的交互作用	分解
模型限制:结构模型设定复杂、解释度差、数据拟合度差、模型不确定性	非参数分析	
统计推论:忽略抽样误差	考虑抽样误差	参数分布设定或者自助法

## (二) 调整的策略

严格来讲,调整并非指某种单一的统计方法,而是包含满足某一特定要求的多种方法的统称(Christiansen et al.,2021; Keele et al.,2023; Lu et al.,2023)。具体而言,调整的目的在于希望通过统计分析手段,在给定的限制条件下,将特定变量分布下的某一统计值转化为另外一个变量分布下的统计值。这里我们可以举一个简单的例子。假设在甲群体中,收入变量  $X$  的累积概率分布为  $P(x)$ ,在特定  $X$  取值为  $x$  的情况下的幸福感的取值为  $\tau$ (例如,给定收入水平  $x$  下的平均幸福感得分)。那么,在甲数据中,幸福感的期望值就是  $\varphi = \int \tau dP(x)$ 。但如果我们想知道在给定收入—幸福感联结模式的情况下,在乙群体中人们的幸福感得分,那么,我们可以收集收入变量  $X$  在乙群体中的累积概率分布  $P'(x)$ ,并求得  $\varphi' = \int \tau dP'(x)$ 。这便是基于乙群体的边际分布情况来对甲群体得出的结论进行调整。显然,从  $\varphi$  到  $\varphi'$  的调整过程取决于  $P(x)$  到  $P'(x)$  的差异大小。而为了表达从  $P(x)$  到  $P'(x)$  的差异,我们可以通过权重矩阵  $M$ ,利用最优化算法,求得  $M$  的估计值  $\hat{M}$ ,以保证  $\hat{M}P(x)$  和  $P'(x)$  之间的距离度量最小化。这里的权重矩阵  $M$  反映了如何进行调整可以让  $P(x)$  和  $P'(x)$  之间的“距离”最短。当然,衡量距离的度量有很多,比如欧氏距离(Euclidean Distance)、马哈拉诺比斯(Mahalanobis)距离或者 Kullback-Leibler 散度。在量化研究中,调整的策略经常被用来探究某一研究结论在应用于不同群体时的外部效度(Pearl & Bareinboim,2014)。对于社会流动研究而言,一个最常见的调整算法是德明—斯蒂芬算法(D-S 算法)(Deming & Stephan,1940)。这一算法通过迭代的方式,在给定一个流动边际分布的前提下计算流动表内各单元的具体人数。限于篇幅,具体的计算过程可以联系作者获取。

在社会流动研究中,通过 D-S 算法探究结构流动并非没有先例。实际上,早在 20 世纪 70 年代,黑泽里格(Lawrence Hazelrigg)便已经采用了 D-S 算法试图把结构流动和交换流动区分开(Hazelrigg,1974; Hazelrigg & Garnier,1976)。但是黑泽里格的分析思路如“昙花一现”,并没有得到学界太多的呼应。此外,如果回到上文论及的方法局限,黑泽里格的分析也没有考虑到结构流动和交换流动的交互效应以及抽样误差。从这个意义上讲,本文的分析是对黑泽里格研究思路的一种继承和发展。

具体而言,以 D-S 算法为基础的调整策略,可以帮助我们更好地探究结构流动。假设我们希望对比两个流动表  $X$  和  $Y$ (例如,不同出生世代个体的先赋地位和自致地位所构成的流动表)。显然, $X$  和  $Y$  有不同的边际分布特征和发生比

取值。此时,我们可以采用D-S算法基于X流动表的边际分布来将Y流动表调整为Y'。Y'所表现的是,当Y流动表的边际分布不是其观测值,而是X流动表的边际分布取值时的流动状态。因此,我们可以称之为反事实流动表。同理,我们也可以基于Y流动表的边际分布将X调整为X',以此建构另外一个反事实流动表。对比这些矩阵,X与Y'之间以及Y与X'之间的边际分布各自保持一致。由于D-S算法保留了发生比的取值,我们对比X和X'就会发现,二者的发生比是一样的,区别只在于边际分布。换言之,两个流动表中的行变量和列变量之间的关联度不变,而边际分布在变化,故而两相对比展现出的是结构流动的效力。另一方面,对比X和Y',二者的边际分布完全一样,因此流动模式的差异可以归因于交换流动。

调整的分析策略有助于解决上文提出的理论支撑问题和混淆因素问题。对于前者而言,D-S算法在探究结构流动的时候直接针对流动表的边际分布进行调整,因此符合理论上的共识,即边际分布体现结构流动。对于后者而言,虽然上面的经验示例仅仅对行变量和列变量进行调整,但是D-S算法调整的对象并不局限于这两个变量。除此之外,如果我们可以获得个体层面其他变量的观测数据,D-S算法完全可以对这些潜在混淆变量“同时”进行调整,以保证在对比流动表的时候被调整的这些其他变量的边际分布保持恒定。更进一步的是,我们也可以调整其他变量与行变量或者列变量之间的联合分布,这一特点对于控制潜在混淆因素而言尤为重要。

但是,仅使用调整的策略无法解决结构流动和交换流动的交互作用问题。这是因为,通过对比观测流动表和反事实流动表,我们所能获得的要么是结构流动的效应,要么是交换流动的效应,二者如何互相影响并没有直接体现,而这则需要使用下述分解的策略来解决。

### (三)分解的策略

对于量化社会学研究者而言,分解的分析策略并不陌生。例如,研究种族歧视问题时很常见的布林德—瓦哈卡分解(the Blinder-Oaxaca decomposition)将白人和黑人的组间收入差异分解为属性(例如教育水平)差异和属性回报率(例如教育回报率)差异。类似的分解策略也可以用来进行社会流动研究。假设流动模式是结构流动S和交换流动C的函数f,那么仿照布林德—瓦哈卡分解,我们可以对 $t_0$ 和 $t_1$ 两个时点(或者出生世代)的流动表之间的差异作类似的分解: $f(S_{t_1}, C_{t_1}) - f(S_{t_0}, C_{t_0}) = f(S_{t_1}, C_{t_0}) - f(S_{t_0}, C_{t_0}) + f(S_{t_1}, C_{t_1}) - f(S_{t_1}, C_{t_0})$ 。显

然,结构流动部分可以表示为 $f(S_{t_1}, C_{t_0}) - f(S_{t_0}, C_{t_0})$ ,而交换流动部分则表示为 $f(S_{t_1}, C_{t_1}) - f(S_{t_1}, C_{t_0})$ 。

但是,虽然回应了既有方法的局限,以上的分解策略仍然没有考虑结构流动和交换流动之间的交互关系。因此,为了将交互作用引入进来,我们基于别文(Martin Biewen)提出的分解策略(Biewen, 2014),对两个流动表之间的差异作三元分解,得到 $f(S_{t_1}, C_{t_1}) - f(S_{t_0}, C_{t_0}) = [f(S_{t_1}, C_{t_0}) - f(S_{t_0}, C_{t_0})] + [f(S_{t_0}, C_{t_1}) - f(S_{t_0}, C_{t_0})] + [f(S_{t_1}, C_{t_1}) - f(S_{t_1}, C_{t_0}) - f(S_{t_0}, C_{t_1}) + f(S_{t_0}, C_{t_0})]$ 。

在这一个拓展之后的分解策略下,结构流动和交换流动仍然可以分别表示为 $f(S_{t_1}, C_{t_0}) - f(S_{t_0}, C_{t_0})$ 和 $f(S_{t_0}, C_{t_1}) - f(S_{t_0}, C_{t_0})$ 。但除此之外,我们可以通过 $f(S_{t_1}, C_{t_1}) - f(S_{t_1}, C_{t_0}) - f(S_{t_0}, C_{t_1}) + f(S_{t_0}, C_{t_0})$ 来探究二者如何相互作用。这里我们可以对这一表达式作如下简单的数学变换: $f(S_{t_1}, C_{t_1}) - f(S_{t_1}, C_{t_0}) - f(S_{t_0}, C_{t_1}) + f(S_{t_0}, C_{t_0}) = [f(S_{t_1}, C_{t_1}) - f(S_{t_1}, C_{t_0})] - [f(S_{t_0}, C_{t_1}) - f(S_{t_0}, C_{t_0})]$ 。前一个方括号表示在把结构流动水平固定在 $t_1$ 时点的时候交换流动的作用,而后一个括号表示在把结构流动水平固定在 $t_0$ 时点的时候交换流动的作用。二者的差值如果不为零,则意味着交换流动的作用大小受到结构流动的影响。当然,我们把各项的顺序调换一下,也可以得到如下的表达式: $f(S_{t_1}, C_{t_1}) - f(S_{t_1}, C_{t_0}) - f(S_{t_0}, C_{t_1}) + f(S_{t_0}, C_{t_0}) = [f(S_{t_1}, C_{t_1}) - f(S_{t_0}, C_{t_1})] - [f(S_{t_1}, C_{t_0}) - f(S_{t_0}, C_{t_0})]$ 。比照上面的解释,这个表达式可以理解为结构流动的取值受到交换流动的影响。

在具体操作过程中,函数 $f$ 可以有很多选择。例如,我们可以用 $f$ 来指代总流动率,即具有流动经历的个体的占比大小。此时,还是以从 $t_0$ 到 $t_1$ 的流动表变化为例(多时点的拓展参见下文的讨论部分),前文的三元分解可以告诉研究者:(1)结构流动的变化如何影响总流动率的变化;(2)交换流动的变化如何影响总流动率的变化;(3)结构流动和交换流动的交互作用变化如何影响总流动率的变化。

除了总流动率之外,我们也可以使用向上流动率(发生了向上流动的个体占比)来指代 $f$ 。此时, $t_0$ 时点的向上流动率到 $t_1$ 时点的向上流动率的变化可以分解为结构流动变化、交换流动变化以及二者交互作用所带来的变化。这里需要说明的是,上述各个部分的效应方向可能不一致。例如,结构流动的变化有可能提升向上流动率,但同时交换流动的变化可能会阻碍整体的向上流动率。因此,这里的分解结果有可能会出现效应方向上的差异。

通过引入 $f$ 来进行三元分解的优势是,我们不再在流动表矩阵的层面上考

察结构流动变化所引发的流动表变化,而是在某一个总括性的函数上看结构流动、交换流动以及二者之间交互作用的大小。因此,我们所要研究的问题就不再是结构流动的具体数值是多少,而是看结构流动的影响力(占比)有多大。这无疑是三元分解方法的一个优势。回到上文提到的结构流动研究中的方法困境,以分解的方式呈现的结构流动和交换流动的交互作用克服了传统概念界定对于结构流动效应和交换流动效应的刚性二分。此外,无论是调整还是分解策略,二者都不涉及具体的结构模型,因此模型设定上的诸多限制便得以解决。

#### (四)关于统计推断的说明

特定的流动表展现的是一个集合性的频次数据,因此难以确定由抽样所带来的估计量的统计不确定性。在这种情况下,一个应对办法是人为地赋予流动表中的频次分布一个特定的参数统计分布,然后基于这样的参数设定来进行统计推断。但是,这种方法的参数设定随意性比较大。此外,为了简化计算过程,需要引入分布的统计趋近(statistical approximation),故而我们并不建议使用参数建模的方式来进行统计推断。由于篇幅所限,具体操作程序可以联系作者获取。如果流动表是基于个体层面的测量建构起来的,我们可以采用自助法等再抽样方法来探究调整—分解的分析过程所隐含的统计不确定性(Mooney, 1996)。需要特别强调的是,这里的自助法所适用的分析单位并不是流动表的单元格频次分布,而是特定调查样本中的个人。我们并不建议直接针对流动表的各个单元频次分布使用自助法。原因在于,流动表只能够把个体分为有限的几个单元,因此以自助法抽取的不同样本之间的变异度会非常低,不利于进行统计推断。

## 五、经验示例

为了展现本文提出的调整—分解的分析策略,我们这里提供两个经验示例。第一个示例基于戈德索普对英国社会职业流动的经典分析而展开。戈德索普在研究中多次提到当时英国社会的宏观变迁对于社会流动状况的影响(例如去工业化、石油危机等宏观事件,集中参见 Goldthorpe, 1987: 253 - 277)。但是,他的分析仍然是从绝对流动—相对流动的区分角度展开的。但正如上文所言,结构性因素带来的流动模式变化应当是结构流动分析的题中之义。采用戈德索普本人所使用的数据,但使用本文提出的新的分析结构流动的策略,正可以展示出本

文所主张的重拾结构流动分析的独特价值。

本文的第二个示例旨在分析中国综合社会调查 (Chinese General Social Survey, 简称 CGSS) 多期数据所反映的教育流动模式。选择这一分析场景的原因是自 1998 年开始的教育扩招正是从结构层面影响代际之间的教育流动。这一点虽然广泛为学界所知,但大量的社会分层和教育社会学研究关注的却是个体层面上微观入学机会的变化。与个体的交换流动相较而言,扩招所带来的结构变迁究竟有多大的影响力仍缺乏进一步的探究。此外,在第一个例子中,戈德索普采用的是宏观层面的行政性数据,因此在第一个例子中我们没有个体层面的资料,无法使用自助法来评估抽样误差,只能基于特定的参数分布设定来“近似”地计算  $P$  值(之所以说是“近似”,是因为需要使用正态分布来估计二项分布)。此外,由于我们缺乏个体层面其他变量的测量,我们也无法通过调整的方法来控制潜在混淆因素。相比于第一个例子中存在的这些局限性,在第二个例子中我们可以获取个体层面的数据和其他变量的信息,因此能够满足基于自助法来评估抽样不确定性以及控制混淆因素的需求。

无论在哪一个具体实例中,我们都可以展示如何利用本文提出的方法来实现对总体流动状况的分解,即:总流动 = 结构流动 + 交换流动 + 两种流动的交互作用。通过这种分解分析,我们可以作如下具有理论和现实启发意义的判断:(1) 结构流动在总流动中的占比;(2) 结构流动占比相比于交换流动在总流动中的占比,孰强孰弱;(3) 结构流动的作用过程是否和交换流动产生交互。除了这种总体流动的分解之外,通过第二个经验示例,我们还展现了如何在控制其他因素的情况下实现对结构流动的分析。这对于克服来自其他变量的潜在选择偏误具有实践层面的意义。

### (一) 戈德索普的职业流动分析

戈德索普的职业流动分析数据来自牛津大学纳菲尔德学院社会流动研究小组于 1972 年在英格兰和威尔士针对 20 ~ 64 岁男性所作的调查。受访者的职业等级评定根据其初职信息进行判定,受访者父亲的职业等级评定根据受访者 14 岁时父亲的职业信息进行判定。职业评级依据的是戈德索普的阶层分类框架(侯利明、秦广强,2019)。这一分类框架将职业细化为 10 个层次,分别标注为 I、II、III、IVa、IVb、IVc、V、VI、VIIa 和 VIIb。为了分析的便利,我们采用了戈德索普原文中的三分类框架,即服务阶层 (I + II, service class)、中间阶层 (III - V, intermediate class) 和工人阶层 (VI + VII, working class) (具体参见 Goldthorpe,

1987:70)。我们选取两个世代,分别是出生于1908—1917年的世代和出生于1938—1947年的世代(相关数据参见Goldthorpe,1987:71)。利用D-S算法,我们针对两个世代的职业流动表通过调整边际分布信息构建了其各自的反事实矩阵(限于篇幅,具体数据从略)。基于观测矩阵和反事实矩阵,我们采用上述的调整—分解的分析策略分析了从1908—1917世代到1938—1947世代的职业流动过程中,结构流动、交换流动以及二者的交互作用各自的影响。分析结果参见表3。

表3 戈德索普英国职业流动数据的分析结果

	点估计值	百分比	<i>P</i> 值(两端检验)
总流动	0.039	100.00%	0.02
结构流动	0.028	73.40%	0.08
交换流动	0.009	22.24%	0.64
结构流动与交换流动的交互作用	0.002	4.36%	1.00

总流动情况是所有发生职业等级流动的个体占比。可以发现,两个出生世代之间的总流动变化中有73.40%是通过结构流动来达成。与之相比,交换流动在总流动中的占比为22.24%,结构流动与交换流动的交互作用占比为4.36%。这一经验发现表明,基于英格兰和威尔士的调查资料所呈现的世代之间的职业流动很大程度上是由于整个社会的职业结构变化,这与戈德索普等人的研究结论一致。在此背景下,交换流动的解释力度不足四分之一。基于参数分布设定的方法,我们也计算了总流动、结构流动、交换流动及结构与交换流动的交互作用各自的*P*值。据此可以判断,总流动在0.05的显著性水平上显著,而结构流动的*P*值表明其只在边际意义上统计显著。交换流动和两种流动的交互作用统计上都不显著。当然,这些计算出的*P*值仅具有参考意义,不可过度解读。

## (二)当代中国的教育流动分析

本文的第二个示例是关于中国社会的代际教育流动。我们采用来自中国综合社会调查2010、2015和2021年的数据。具体而言,我们将研究对象的年龄限定在30~35岁之间,其教育水平为调查时所获得的最高教育程度,其父亲的教育程度也是所获得的最高教育程度。我们之所以选取这三年的调查资料,是因为这三次调查在询问被研究对象和其父亲的教育水平时采用了一致的问题表述

和选项设置,这有助于我们排除问题设计差异对结论的影响,进行跨时期比较。由于CGSS问卷中的教育水平分类十分详细,以此构建流动表容易出现大量零值,因此我们将被访者及其父亲的受教育程度统一编码为五类:1 = 未受过正式教育;2 = 小学教育;3 = 初中教育;4 = 高中教育;5 = 大学教育及以上。由于CGSS提供了个体层面的多个变量,我们可以基于它们来控制影响父亲教育水平和子女教育水平的潜在混淆因素。此外,我们也可以基于个体层面的观测数据采用自助法来评估调整—分解分析策略的抽样不确定性。出于示例展示的需要,我们这里选取的混淆因素是父亲的出生年份。为了便于后续的调整分析,我们将父亲的出生年份进行分组,分为1949年以前、1949—1959年、1959—1969年三组。需要特别说明的是,在CGSS2021年的调查中,少数被调查对象的父亲的出生年份在1969年以后,但考虑到2010年和2015年CGSS的调查中均不存在此类个体,故而我们仅保留三组年龄组(这一操作仅作示例使用)。

基于调整—分解的策略,我们分析了教育总体代际流动情况,相关分析结果参见图1。图1a表明,从2010年到2015年,总体教育流动率的变化比例为7.1%。当我们控制交换流动后,结构流动的变化可以带来总体教育流动率3.9%的变化。而在控制结构流动之后,交换流动的变化也可以带来等量的3.9%的总体教育流动率的提升。我们利用自助法发现,无论是总教育流动率变化、结构流动带来的变化还是交换流动带来的变化,都在统计意义上显著(其95%置信区间不包含零值)。结构流动和交换流动二者的交互作用为负,但在统计意义上不显著。在图1b中,我们控制了父亲出生年份组的边际分布、其与父亲教育水平的联合分布以及其与子女教育水平的联合分布。与没有进行这一控制操作的结果相比,结构流动的作用、交换流动的作用以及二者的交互作用的估计值都发生了扩大。这表明,伴随着中国社会大规模的教育扩招,年轻人可以普遍从中“获益”,从而获得相较于父亲一代更高的教育水平(Hu et al.,2020)。而从父代到子代,不同教育水平之间的流动壁垒也呈现随时间推移而逐渐弱化的特点。但与戈德索普的研究结果类似的是,结构流动和交换流动的交互作用在控制了父亲的出生年份之后显著为负。这也就意味着,虽然伴随着时间推移,教育代际流动的藩篱本身越发容易突破,但是在将教育扩招这种结构性变化考虑进来之后,教育扩招反而有可能会强化教育类型之间的壁垒。这或许可以用韦伯的闭合理论(closure theory)进行解释,即当高教育成就越容易获得时,人们反而越发强调教育的分层(区隔)作用(Hu,2015)。图1c和图1d展示了从2015年到2021年的教育代际流动的变化情况。如果我们将抽样误差考虑进来,可以发现教育的

代际总体流动率并没有在 2015 到 2021 年之间发生显著的变化。与此同时,无论是结构流动、交换流动还是二者的交互作用,都不再对教育代际流动变化存在显著影响。一个可能的解释是,在经历了数年快速的教育扩招之后,教育机会对教育流动的效应逐渐减弱。换言之,教育的代际流动出现了一种稳定化的态势。

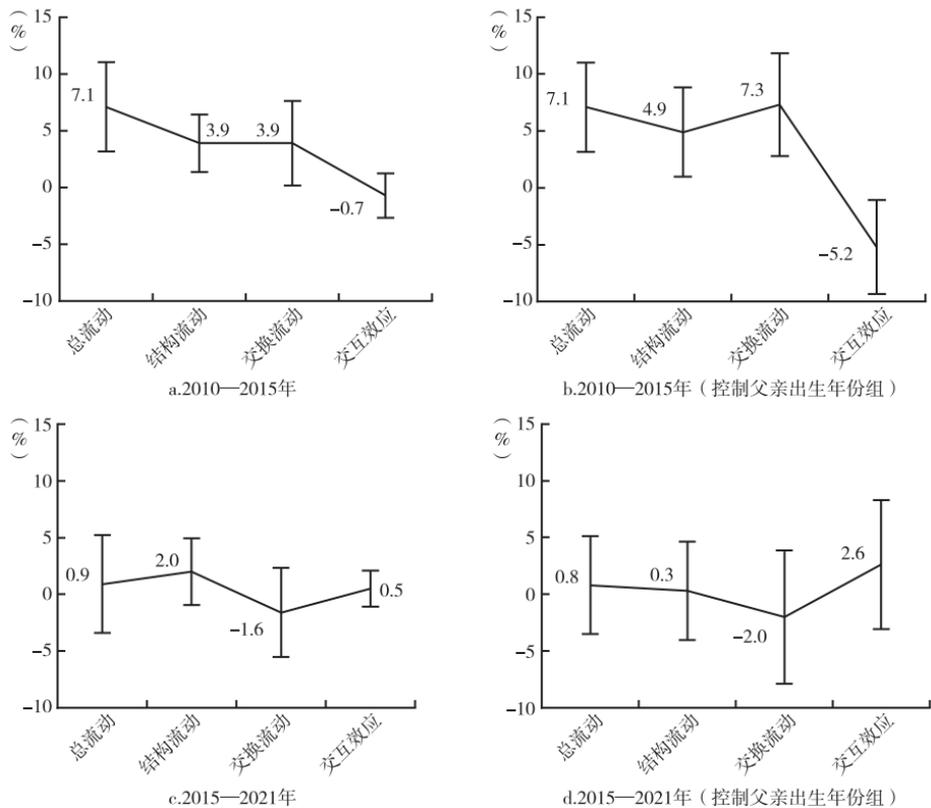


图 1 中国教育流动的分析结果

## 六、结 语

结构流动是形塑转型期社会流动基本样态的重要机制。本文系统梳理和讨论了现有文献中分析结构流动的相关方法,并指出其在理论支撑、混淆因素控制、概念化、统计模型复杂度以及统计推断等方面的局限。在此基础上,本研究提出了调整—分解的分析策略。通过综合采用德明—斯蒂芬算法和流动表的三元分解法,这一分析策略直接针对流动表的边际分布构建反事实流动表,并以非

建模对比的方式呈现结构流动、交换流动以及二者的交互作用对于流动样态变化的影响。除此之外,潜在的混淆因素也可以通过调整的方式予以控制,相关的抽样不确定性也可以通过参数分布设定或者自助法的方式进行分析。基于戈德索普经典分析的数据以及中国综合社会调查中的代际教育流动相关数据,本文展示了调整—分解分析策略的两个经验示例。

本文所提出的调整—分解的分析策略可以从多个方面予以拓展。首先,除了D-S算法之外,研究者也可以采用其他的算法达成调整流动表边际分布的目的。例如,那左迪(Anna Naszodi)和合作者提出的利用新的统计量进行交互分类表分解的方法(简称N-M算法)也可以用于分析结构流动(Naszodi,2023;Naszodi & Mendonca,2023,2024)。由于篇幅所限,具体示例可以联系作者获取。其次,在本研究中,父亲的出生年份组是作为混淆因素进行了控制,其本身并没有作为一个影响流动样态的变量进行考量。而调整—分解的分析策略完全可以更进一步,考察三个或者更多因素对于流动样态的影响。只是这方面的分解结果会过于复杂,不利于研究者对经验结果进行解读。这方面具体的分解过程可以联系作者获取。考虑到这种复杂性,我们并不建议针对多个影响因素进行调整—分解分析。再次,虽然本文以一种非模型对比的方式来使用调整—分解的分析策略,但相关的分析结果完全可以和统计模型结合起来。例如,我们如果有多期的流动表数据,那么相关的调整—分解结果就可以告诉我们,以第一期数据的流动表为参照,后面各期的流动表样态变化如何受到结构流动和交换流动的影响。这一分析结果完全可以纳入传统的多层次模型中,与其他宏观层面的变量或者个体层次的变量产生互动。最后,本文提出的调整—分解的分析策略虽然以分类变量为分析对象,但是类似的思路也可以应用到对连续型的分层变量(例如收入)的分析之中。在此情况下,我们可以选取适合连续型变量的统计量(如Copula函数等)来进行分解分析。

#### 参考文献:

- 蔡昉,2020,《如何开启第二次人口红利?》,《国际经济评论》第2期。
- 陈云松,2022,《当代社会学定量研究的宏观转向》,《中国社会科学》第3期。
- 干春晖、郑若谷,2009,《改革开放以来产业结构演进与生产率增长研究——对中国1978—2007年“结构红利假说”的检验》,《中国工业经济》第2期。
- 高勇,2009,《社会樊篱的流动——对结构变迁背景下代际流动的考察》,《社会学研究》第6期。
- 侯利明、秦广强,2019,《中国EGP阶层分类的操作化过程——以中国综合社会调查(CGSS)数据为例》,《社会学评论》第7期。
- 胡安宁,2017,《统计模型的“不确定性”问题与倾向值方法》,《社会》第1期。
- 魏杰、汪浩,2016,《结构红利和改革红利:当前中国经济增长潜力探究》,《社会科学》第1期。

- Bearman, Peter S. & Glenn Deane 1992, "The Structure of Opportunity: Middle-Class Mobility in England, 1548 - 1689." *American Journal of Sociology* 98(1).
- Biewen, Martin 2014, "A General Decomposition Formula with Interaction Effects." *Applied Economics Letters* 21(9).
- Blau, Peter M. & Otis Duncan 1967, *The American Occupational Structure*. New York: The Free Press.
- Breen, Richard 1985, "A Framework for Comparative Analyses of Social Mobility." *Sociology* 19(1).
- 2010, "Educational Expansion and Social Mobility in the 20<sup>th</sup> Century." *Social Forces* 89(2).
- Christiansen, Rune, Niklas Pfister, Martin Emil Jakobsen, Nicola Gnecco & Jonas Peters 2021, "A Causal Framework for Distribution Generalization." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44(10).
- Cobalti, Antonio 1988, "Alternative Conceptual Frameworks for the Analysis of Mobility Tables and the Log-Linear Models." *Quality and Quantity* 22(1).
- Deming, W. Edwards & Frederick F. Stephan 1940, "On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals Are Known." *The Annals of Mathematical Statistics* 11(4).
- DiCiccio, Thomas J. & Bradley Efron 1996, "Bootstrap Confidence Intervals." *Statistical Science* 11(3).
- Goldthorpe, John H. 1987, *Social Mobility and Class Structure in Modern Britain* (Second Edition). Oxford: Clarendon Press.
- Goldthorpe, John H., Clive Payne & Catriona Llewellyn 1978, "Trends in Class Mobility." *Sociology* 12(3).
- Hazelrigg, Lawrence E. 1974, "Partitioning Structural Effects and Endogenous Mobility Processes in the Measurement of Vertical Occupational Status Change." *Acta Sociologica* 17(2).
- Hazelrigg, Lawrence E. & Maurice A. Garnier 1976, "Occupational Mobility in Industrial Societies: A Comparative Analysis of Differential Access to Occupational Ranks in Seventeen Countries." *American Sociological Review* 41(3).
- Hope, Keith 1982, "Vertical and Nonvertical Class Mobility in Three Countries." *American Sociological Review* 47(1).
- Hout, Michael 1983, *Mobility Tables*. Newbury Park: Sage.
- 1989, *Following in Father's Footsteps: Social Mobility in Ireland*. Cambridge: Harvard University Press.
- Hu, Anning 2015, "Evaluating Educational Inequality Within Educational Expansion: A Formal Comparison Between Odds Ratio and the Educational Gini Coefficient." *Journal of Mathematical Sociology* 39(4).
- Hu, Anning & Reid J. Leamaster 2015, "Intergenerational Religious Mobility in Contemporary China." *Journal for the Scientific Study of Religion* 54(1).
- Hu, Anning & Xiaogang Wu 2021, "Cultural Capital and Elite University Attendance in China." *British Journal of Sociology of Education* 42(8).
- Hu, Anning, Grace Kao & Xiaogang Wu 2020, "Can Greater Reliance on Test Scores Ameliorate the Association Between Family Background and Access to Post-Collegiate Education? Survey Evidence from the Beijing College Students Panel Survey." *Social Science Research* 88 - 89.
- Jones, Frank L. 1985, "Structural and Circulation Mobility Are Alive and Well: A Critical Comment on Sobel's Premature Obituary Notice." *Sociology* 19(1).
- Keele, Luke J., Eli Ben-Michael, Avi Feller, Rachel Kelz & Luke Miratrix 2023, "Hospital Quality Risk

- Standardization via Approximate Balancing Weights.” *The Annals of Applied Statistics* 17(2).
- Krauze, Tadeusz K. & Kazimierz M. Slomeczynski 1986, “Matrix Representation of Structural and Circulation Mobility.” *Sociological Methods & Research* 14(3).
- Lu, Benjamin, Eli Ben-Michael, Avi Feller & Luke Miratrix 2023, “Is It Who You Are or Where You Are? Accounting for Compositional Differences in Cross-Site Treatment Effect Variation.” *Journal of Educational and Behavioral Statistics* 48(4).
- Marks, G. N., Y. E. Pittelkow & F. L. Jones 1996, “Structural Changes and Association in Mobility Tables: An Integration and Extension of Existing Approaches.” *Quality and Quantity* 30(1).
- Mooney, Christopher Z. 1996, “Bootstrap Statistical Inference: Examples and Evaluations for Political Science.” *American Journal of Political Science* 40(2).
- Naszodi, Anna 2023, “The Iterative Proportional Fitting Algorithm and the NM-method: Solutions for Two Different Sets of Problems.” (<https://arxiv.org/abs/2303.05515>)
- Naszodi, Anna & Francisco Mendonca 2023, “A New Method for Identifying the Role of Marital Preferences at Shaping Marriage Patterns.” *Journal of Demographic Economics* 89(1).
- 2024, “Changing Educational Homogamy: Shifting Preferences or Evolving Educational Distribution?” *Journal of Demographic Economics* 90(2).
- Pearl, Judea & Elias Bareinboim 2014, “External Validity: From Do-Calculus to Transportability Across Populations.” *Statistical Science* 29(4).
- Slomeczynski, Kazimierz M. & Tadeusz Krauze 1984, “Should the Framework of Structural vs. Circulation Mobility Be Abandoned? If So, Not Because of Faulty Logic.” *American Sociological Review* 49(6).
- Sobel, Michael E. 1983, “Structural Mobility, Circulation Mobility and the Analysis of Occupational Mobility: A Conceptual Mismatch.” *American Sociological Review* 48(5).
- Sobel, Michael E., Michael Hout & Otis Dudley Duncan 1985, “Exchange, Structure, and Symmetry in Occupational Mobility.” *American Journal of Sociology* 91(2).
- 1986, “Saving the Bath Water: An Invited Comment on Krauze and Slomeczynski’s ‘Matrix Representation of Structural and Circulation Mobility’.” *Sociological Methods & Research* 14(3).
- Ultee, Wout & Ruud Luijkx 1986, “Intergenerational Standard-of-Living Mobility in Nine EEC Countries: Country Characteristics, Competitive Balance and Social Fluidity.” *European Sociological Review* 2(3).
- Whyte, Martin 2010, *Myth of the Social Volcano: Perceptions of Inequality and Distributive Injustice in Contemporary China*. Stanford; Stanford University Press.
- Wu, Xiaogang 2011, “The Household Registration System and Rural-urban Educational Inequality in Contemporary China.” *Chinese Sociological Review* 44(2).
- Xie, Yu 2016, “Understanding Inequality in China.” *Chinese Journal of Sociology* 2(3).
- Yasuda, Saburo 1964, “A Methodological Inquiry into Social Mobility.” *American Sociological Review* 29(1).

作者单位:复旦大学社会学系  
责任编辑:胡舍之