

从阶层分类到阶层量化： 社会阶层量化方法新范式的一种探索

孙静含 李升

摘要：在以往的社会分层与流动研究中，研究者们常使用国际社会经济地位指数 ISEI 指标或 EGP 阶层分类框架等主要方法对社会阶层进行测量研究。然而，一方面 ISEI 指标主要考虑经济和教育两个因素且前提是依据主观测量的职业声望，因而容易受到质疑；另一方面 EGP 阶层分类框架则主要依据职业状况进行分类，虽然考虑了雇佣关系及技术能力等客观因素，却也面临阶层分类边界是否清晰适用于国内研究的局限。基于 EGP 阶层分类法，将收入和教育因素纳入 EGP 阶层分类中，提出一种新的计算方法，即通过计算获得“阶层地位分”，从而实现从阶层分类转向阶层量化的社会阶层连续性测量的新方式。通过使用 CGSS 数据进行验证分析，发现使用迭代法能够计算出收入和教育两个因素在阶层分类中的比重，以此计算出的“阶层地位分”也更易反映社会现实。该方法能够解决困扰国内研究的阶层量化问题，为今后的社会阶层研究方法拓展出一种新的可能性。

关键词：社会阶层研究；EGP 阶层分类；阶层量化；阶层地位分

①作者简介：

李升，北京工业大学文法学部社会学系教授、北京社会管理研究基地研究人员。
孙静含，北京工业大学文法学部。

From Class Classification to Class Quantification: An exploration of a new paradigm of social class research technique

Jinghan SUN, Sheng LI

ABSTRACT

In the previous studies about social stratification and social mobility, researchers always take the principal methods, such as the ISEI (International Socio-economic Index) or the EGP classification framework, to measure social class. However, the ISEI index mainly focuses on economic and educational factors, and the EGP classification framework primarily classifies according to the occupation--both are limited by the lack of consideration of other factors. Based on the EGP classification framework, a new calculation method is proposed by incorporating income and education factors into EGP classification framework. To realize the transformation from class classification to class quantification in the study of social class, the "class status score" is calculated. Illustrated by the CGSS data, the proposed iterative method can calculate the proportion of income and education in class classification, and the "class status score" is also easier to reflect the social reality, compared to the traditional methods. The method provides a solution for the long-standing problem of class quantification and offers a new possibility for the future study of social class.

KEY WORDS

Social class research; EGP classification framework; Class quantification; Class status score

一、问题提起：社会阶层何以测量

如何对社会阶层进行测量，以更为直观有效地开展社会结构分析，一直以来都是阶层研究领域中最重要的问题之一。无论是马克思主义的经济要素决定，抑或韦伯主义的财富权力声望维度区分，学者们都认可需要将社会阶层进行测量及分类研究，以更清晰理解社会结构特征及其构建起的社会秩序，以此也发展出了新马克思主义式（Wright, 1997）及新韦伯主义式（秦广强，2011；Goldthorpe, 1987）的阶层分析框架。由于社会阶层的类型化或水平化是社会分层与流动研究的基础，因而社会阶层何以测量自然就成为推动社会阶层研究的重要动力机制。

对社会阶层进行测量主要可分为分类性指标测量与连续性指标测量。已有研究普遍是以阶层分类为基础，基于定类或定序的尺度进行测量分析，并被认为是适合于经验社会中人们对社会分化的群体分类认知且易于研究分析，其中包含了依据客观标准测量及主观标准测量。分类性指标测量具有重要意义，阶层分类体系对于解释宏观社会结构、群体社会经济差异、个体态度行为差异等都具有一定的解释力（李春玲，2001），分类标准会根据研究者的研究目的而多样化（李强，2006），适合于不同的研究情境。然而，经验社会中的群体及个体分化是否存在明确的分类边界是需要讨论的，理论研究中对社会阶层结构的量化也是需要持续探索的重要领域。以社会经济地位指数为主的职业声望法即是通过连续性指标测量来进行阶层量化分析，尽管此类方法受到了很多批评（Grusky, 2001），但其量化方法不仅使阶层分析更易于实证，也被认为实际上连续性测量才更符合经验现实中人们社会经济地位的难以分界

性特征。基于国内研究中关于阶层分类体系应用丰富而阶层量化分析不足的现实, 本文将在讨论经典阶层分类方法与现有研究应用情况的基础上, 重点探索对社会阶层进行量化分析的新范式。

(一) ISEI 指标与 EGP 阶层分类

在社会阶层研究早期, 学者们主要尝试以职业为标准对社会阶层进行分类, 用以创建一个较为合理的、与经验事实相符合的阶层结构模型。美国社会学家布劳和邓肯最早对美国的职业结构进行了较为详尽的考察, 将与职业关联的教育和收入纳入回归方程并以此计算出各个职业的社会经济地位 (Blau & Duncan, 1967), 同时提出了 SEI 指标, 即社会经济地位指数 (Socio-economic Index)。在 SEI 指标的基础上, 甘泽布姆、格拉夫和特雷曼等学者以国际标准职业分类 (ISCO) 为基础, 使用 16 个国家的 31 套数据, 为 271 个不同类别的职业进行了分数的分配, 得到了一个具有一定量化性质的职业分数表, 即国际社会经济地位指数 ISEI (International Socio-economic Index) 指标 (Ganzeboom et al., 1992), 这也成为后来测量社会阶层地位高低的重要依据。在围绕以“职业”为核心的阶层测量中, 除了如“社会经济地位指数”进行阶层量化测量外, 还有如“职业声望量表”的基于主观数据的量化测量 (Bukodi et al., 2011)。然而, 由于前述测量时的依据涉及职业声望等主观性评价, 因而也常被人质疑, 学者们也更为追求科学性高的客观评价框架。

国际社会经济地位指数 (ISEI) 的测算方法虽为社会阶层研究提供了很好的阶层分类参照, 但也存在一定的局限。一方面, ISEI 指标与 SEI 指标一样, 指标的连续性质同时造成了阶层界限的不明确, 另一方面,

测算时都只将与职业关联的教育和收入的两类因素纳入考察范围，忽略了许多其他职业特性，比如上升渠道、雇佣关系和生产资料所有等。因此，ISEI 指标虽然对职业进行了一定程度的量化并且获得了较为准确的结果，但却只是有限地被应用于社会学的研究。相比于 ISEI 指标，EGP 阶层分类框架被国际上更多学者所认可并采用。该分类方法主要由 Erikson、Goldthorpe、Portocarero 的三位学者在 1979 年提出，三人以雇佣关系、生产资料所有等因素作为标准，对职业进行了分类，并且逐渐扩展和完善了最早的分类方法，最终形成了 11 分类、10 分类、7 分类等多个不同的 EGP 阶层分类框架^①，同时该框架被应用于社会流动领域的相关研究（Erikson et al., 1982、2010；Tahlin, 2007），也对欧洲很多国家的政府及统计部门等职业归类应用产生了十分重要的影响。

国内分层与流动的早期研究中，应用 EGP 阶层分类框架的研究并不多，近年来逐渐得到重视并开始形成研究成果，已有研究也将 EGP 阶层分类框架在中国调查数据中的操作化过程进行了详细介绍（侯利明、秦广强，2019）。然而，即便 EGP 阶层分类框架已逐渐得到认可，却也并不完美。首先，该方法缺乏对经济和教育因素的直接考察（虽然职业本身已经一定程度上包含这两个因素），正如传统韦伯主义所表明的，经济因素仍是影响社会阶层的重要指标（Chan & Goldthorpe, 2007），这就导致该框架缺乏足够清晰完整的理论基础且与实际情况存在一定偏差（秦广强，2011）。其次，由于该框架提供了相对过多且不可被量化的类别，导致后续研究者在使用该框架时，为了使模型不至于过于复杂，经常会

^①EGP 阶层分类框架通常包含 7 个大类，在 7 大类下，又可以划分为多个不同小类，于是产生了 10 分类或 11 分类的模型。

出于研究目的的需要将原有的框架进行简化。最后,该框架仅仅是泛化的,缺少对阶层间距离和阶层内部可能存在的差异的考察,因而导致在使用该阶层分类框架时会忽视一些本该被发现的事实。

(二) 现有研究中的阶层分类应用

在现有阶层研究的定量分析中,对阶层变量的处理主要有两种模式:其一是直接依据已有研究的阶层分类作为变量纳入模型进行分析,其二是以阶层分类为基础对数据进行再次分类,再分别构建模型。在这两种模式下,研究者们对阶层变量会进行不同的选择及处理,其中较为常见的主要有以下方式。

第一种是直接使用主观阶层认同作为阶层变量进行分析(陈云松、范晓光, 2016; 李升、倪寒雨, 2018; 许琪, 2018; 陈云松等, 2019; Tang & Tan, 2021)。主观阶层认同的确可以在一定程度上反应个体所处的阶层且使用起来更为容易,但却只能用于探索个体的社会心理层面,而不是客观阶层地位。同时,由于主观阶层认同和客观所属阶层之间往往有较大差异(范晓光、陈云松, 2015; Benjamin et al., 2013),因此主观阶层认同不能直接作为客观所属阶层进行使用,这导致其在使用上有较大局限,无法有效的应用于回答更多的社会事实问题,于是将主客观阶层因素同时考虑或尝试消解主客观偏误成为重要的补充研究方式(李培林、张翼, 2008; 李升, 2017; 宋健、张晓倩, 2020)。

第二种是依据国际社会经济地位指数(ISEI)或EGP阶层分类法以及在此基础上进行重新分类。多数研究主要是以减少分类类别、降低模型的复杂度为目的(李姚军、王杰, 2021; 赵晓航、李建新, 2017;

Graaf et al., 1995), 如将 EGP 职业阶层简化为“非体力劳动者阶层-体力劳动者阶层”的二分类(尚进云、王振振, 2020), 再如以 SES (socio-economic status) 分数为标准建构二分类模型(Walpole, 2003)以及分析中国社会的声望分层(李春玲, 2005), 以 ISEI 为基础区分上层、中层与下层(李强, 2021), 减少分类的方法也为研究中国社会流动状况提供了可能(李路路、朱斌, 2015)。然而, 如前文所述, 减少分类的方法虽然简化了模型的复杂度, 且更易有助于研究, 却同时增加了结果中可能出现的偏误。对此, 一些学者开始尝试构建 CNLR 模型以提高结果的可信程度(Chen, 2018), 但是该方法依然无法从根本上解决问题, 同时又增加了模型自身的复杂度。

第三种是国内学者采用认为是更适合本土经验的其他分类方法。如考虑组织资源、经济资源和文化资源占有状况的十大阶层分类框架(陆学艺, 2002), 基于赖特的新马克思主义阶层分类框架(张翼, 2008), 考虑工作状况与组织权威特征的阶层分类体系(李路路等, 2012), 考虑行政协调和市场协调关系的支配权大小的七分类阶层框架(刘欣, 2018)以及为便于分析社会流动而简化为三阶层分类(李路路等, 2018)、区分“体力与非体力”的二分类(张顺、祝毅, 2021)等。实际上, 还有很多在具体研究中根据研究需要进行阶层分类的实例, 这里不再一一列举。整体来看, 国内学者对于社会阶层结构的探讨多基于职业、教育、收入等采取单一或组合的方式(刘欣、田丰, 2018), 且已有阶层分类方式或多或少都受到新马克思主义和新韦伯主义的阶层分类模式影响(李春玲, 2019)。尽管已有研究的阶层分类方式一定程度满足了研究者分析现实社会的研究需要, 但考虑到阶层变量现实内涵的复杂性

特征，相关的阶层分类研究也会面临因考虑因素不足而造成分析结果与现实状况上的偏差。因此，仍有必要对现有阶层分类框架进行改进及进行量化分析的完善，以弥补现有框架的一些局限，尤其是 EGP 阶层分类框架的不足。

二、“阶层地位分”：从阶层分类到阶层量化

（一）阶层分类的现实复杂性

根据前文所述，目前主流研究会以 EGP 阶层分类框架或其他简单明晰的阶层分类框架开展研究，且核心是依据“职业”的基本特征，对职业进行区分，并最终构建了多分类的阶层分类框架。在使用 EGP 阶层分类法或其他简单分类法进行研究时，意味着其认可一个基本假设：即同一阶层内部的所有成员都具有相似的经济水平、行为表现、意识特征等，这显然与经验事实并不完全一致。举例而言，小学教师、中学教师和高校教师，虽然在职业上都可分类为教师，但是各类教师之间也存在一些显著差异。该引例仅仅是同一职业内，若扩大到同一阶层时，其差别就可能更大。因此，已有分类方法为同一阶层内部的所有成员赋予完全相同的社会位置，显然具有科学研究的局限性。

从现实情况来看，社会成员个体所处的实际社会地位，并不是由其所属职业决定的，而是由其工作状况、雇佣关系、技术能力、上升渠道、收入水平、受教育程度等因素共同决定的。换言之，每个社会成员个体不是因为他的职业是什么而处于什么社会位置，而是取决于其职业特征和自身特征等综合状况确定处于什么社会位置。因此，职业虽然成为阶

层分类的核心指标，但不能完全涵盖或指代所处的社会位置，关键在于职业和其自身的资源，同一职业的人可能有不同的资源，不同职业的人也可能有相似的资源。只要充分掌握了职业特征和自身特征等综合因素信息，每个社会成员个体的阶层就都是可以单独量化的。

（二）阶层量化的新内涵

基于对阶层量化的考量，本文将量化的结果值命名为“阶层地位分”，其影响因素可以与阶层分类分析完全相同，只是表示的内容并不是社会阶层分类，因为社会阶层分类一定是有限多个。事实上，“阶层地位分”所表示的是个体所拥有的、可以影响其所属阶层的在职业与自身方面关联资源量的多少，是一个从社会阶层中分化出来的分析概念，而非独立概念，其存在的目的就是进行对阶层进行量化，为后续的阶层研究提供帮助。在讨论阶层的社会流动时，就可以使用“阶层地位分”来测量，以更准确地表明阶层流动的实际距离。

例如，通常的代际流动研究中，尽管依据 EGP 阶层分类能够分析从农业体力阶层流动到普通体力阶层、普通体力阶层流动到中间阶层等情况，然而这些流动距离的衡量并不完全相同。例如，虽然可以区分年收入 5 万的自雇农民流动到年收入 50 万的自雇小资产者，与年收入 50 万的自雇农民流动到年收入 5 万的自雇小资产者，但二者之间显然是不同的社会流动状况。若能使用“阶层地位分”这样的分析方法考察个体所属的阶层，就可以更好地表明年收入 5 万的自雇小资产者和年收入 50 万的自雇小资产者之间的社会位置差别。

“阶层地位分”主要包括四个层面的内涵，由此构成其方法成立的

基础命题。第一，由于社会成员现实中的职业状况关联了诸多内容，彼此会形成多样化的社会状态，因此社会成员个体所属阶层都是可以单独量化，而不是简单分类。第二，阶层量化后，下一阶层收入和受教育水平最高者，和上一阶层收入和受教育水平最低者，二者的阶层地位分差距很小，且无限接近于 0。例如，年收入 50 万的农民，其地位不会明显低于年收入 20 万的半技术和非技术的体力工人。当样本足够大时，阶层地位分应是连续的，不会出现某一中间数值缺失或断裂的情况。第三，由于阶层地位分以资源为依据，而同一阶层内部的个体所拥有的资源实际可能有很大差别，因此，同一阶层内部的阶层地位分差距可能会很大，甚至可能会大于两个阶层分类之间的差距。例如，尽管职业分类相同，但年收入 100 万的自雇小资产者与年收入 5 万的自雇小资产者之间的差距可能是巨大的。第四，阶层地位分是连续型的定距变量，而非定比变量，不存在阶层地位分为 0 的个体，只存在阶层地位分数相对为 0 的个体。虽然阶层地位分本质表示的是资源，但是必须注意的是，其所表示的资源是一种复合式资源，即由许多其他资源通过一定方式相叠加而成。由于这些被叠加的资源本身具有类型上和数值上的差异，因此所生成的变量数值之间不存在倍数关系。

三、“阶层地位分”的设计思路和计算方法

（一）设计思路

考虑到阶层量化的复杂性特征，如果从最开始的零对社会成员个体职业和自身资源进行重新综合考察，那将是极为繁琐且易错的。不过有

利的是,EGP 阶层分类框架通过不断修正已经按照个体的难以量化的信息对职业进行了大致的划分,且 EGP 阶层分类中并未考虑到的收入和教育因素都是可以直接量化的,因此,最直接的解决方式就是在 EGP 阶层分类的基础上,通过加入收入和教育因素再对各类别和类别内的个体进行重新界定,以赋予相应的“阶层地位分”。

根据这样的思路,阶层量化所需解决的问题就转化为如何将收入和教育因素纳入 EGP 阶层分类框架之中。这实际上涉及一个数学算法,该算法所解决的根本问题是,在有序分类变量的基础上,对数值连续型变量进行重新赋值,要求下一类别的最高值必须低于上一类别的最低值。该算法首先需要得出各阶层的一个基础值,然后以基础值为基准分,标准差为距离,确定各阶层间的分界值,再以上下两个分界值为界限,对各类别内部的所有个案进行重新赋值并得出各个案的阶层地位分。

(二) 计算方法

根据上述设计思路,“阶层地位分”的计算方法需要解决以下五个基本问题。

第一,如何确定经济因素和教育因素的占比。这是一个长期困扰阶层研究的问题,事实上也很难为二者的占比给出一个明确的数值。这一方面是因为缺少明确的、既符合经验事实也符合理论框架的阶层分类模型,以对该数值进行有效的调整和确定。事实上的逻辑关系是,必须有确定的阶层模型才能获得该数值,但同时没有该数值就无法确定一个足够可信的阶层模型。另一方面因为该数值总是因地而变、因时而变,甚至可能因阶层而变的。为了更好地对该数值进行确定并进行后续的调整,

在此引入占比系数 β , β 初始默认值为0.5, 该值越高, 教育因素占比越高, 反之则经济因素占比越高。

第二, 如何对经济变量和教育变量进行操作化处理。首先, 在算法中假设经济变量和教育变量都符合正态分布, 因此对经济变量首先进行了自然对数的处理, 教育变量用受教育年限表示; 其次, 由于要将经济变量和教育变量通过 β 值整合为一个变量, 称之为基本分数(bCode), 因此两个变量必须处于同一量级之中。好在经济变量进行自然对数处理后, 其与受教育年限刚好处在同一量级, 因此不需要再进行进一步处理。

第三, 如何确定各阶层的基础值。对该问题的解决有两个不同方法, 方法1是: 用各阶层基本分数的均值减去全样本基本分数的均值(调整系数 α)除以2, 再加上所处阶层, 其中 i 为阶层, 除以2的原因是避免下一阶层的基础值高于上一阶层的基础值。

$$\frac{(S_i^{bas} - \alpha)}{2} + i$$

该方法所得出的模型逻辑上会大致符合实际情况, 但是由于该方法只是对阶层之间的相对位置进行了有限调整, 因而会导致忽略一些可能的情况。例如, 最低一个阶层占据了0-5的阶层地位分, 剩余阶层只占据了5-10的阶层地位分。同时, 该方法中所使用的“除以2”是一个缺乏理论论证的经验数值, 会导致其科学性下降。

因此, 有必要在此创建方法2: 最高阶层的基础值等于其基本分数均值, 在此基础上, 每一阶层的基础值均为该阶层上一阶层的基本分数均值加上该阶层的基本分数均值, 减去该阶层之上所有阶层的基本分数均值的均值, 于是倒数第二个阶层的基础值也等于其基本分数均值。该

方法的一个缺陷是，由于是逐级进行比较且参与比较的阶层逐级增加，因此会缩小被计算的最初几个阶层的基础值分差而扩大最末几个阶层的基础值分差。为此，可以从最低阶层和最高阶层分别开始进行计算，然后使用两次计算的基础值差的均值作为新的基础值分差。在修改后的方法 2 中，各阶层的基础值分差实际就是下一阶层的基本分数均值和其上所有阶层的基本分数均值的差，与上一阶层基本分数均值和其下所有阶层的基本分数均值的差，二者的均值。其中， N 为阶层总数。该方法在计算过程中，必须从最高阶层开始逐级向下计算。

$$B = \begin{cases} \bar{X}_i & i = N \\ B_{i+1} - \frac{\left(\bar{X}_{i+1} - \frac{\sum_{j=i+1}^N \bar{X}_j}{N-i} \right) + \left(\bar{X}_i - \frac{\sum_{j=1}^i \bar{X}_j}{i} \right)}{2} & i < N \end{cases}$$

第四，如何确定各阶层间的分界值。基本思路是以基础值为基础，以标准差为标准来确定各阶层之间的距离，简而言之，就是通过上下两阶层标准差的比例，在上下两阶层基础值中间找到两个阶层之间的分界值，即用上一阶层的基础值减去下一阶层的基础值，除以两阶层的标准差之和，再乘以下一阶层的标准差，加下一阶层的基础值。该方法中存在两个特殊值，当 i 为最小值时，则其下限为两倍其基础值减去其与上一阶层的分界值，当 i 为最大值时，其上限为两倍其基础值减去其与下一阶层的分界值。由于个体所拥有的资源不可能小于 0，因此有必要固定阶层地位分的最低值为 0，即 i 为最小值时，其下限为 0。为了使该值固定，在此需要引入一个增加系数 Xt ，该数值用于对所有阶层的基础值进行调整，可以得到公式：

$$B_1 \times 2 - L_2 = 0$$

其中, B_1 为最低阶层的基础值, L_2 为最低阶层和倒数第二阶层之间的分界值, S 为标准差。公式展开可得:

$$2(B_1 + X_i) - \left[\frac{((B_2 + X_i) - (B_1 + X_i)) \times S_1}{S_1 + S_2} + (B_1 + X_i) \right] = 0$$

因此, 各阶层的实际基础值为:

$$B = \begin{cases} \bar{X}_i & i = N \\ B_{i+1} - \frac{\left(\bar{X}_i + 1 - \frac{\sum_{j=i}^i \bar{X}_j}{i} \right) + \left| \bar{X}_i - \frac{\sum_{j=i+1}^i \bar{X}_j}{N-i} \right|}{2} & i < N \end{cases} + \frac{B_2 - B_1}{S_2 + S_1} \times S_1 - B_1$$

前半部分为原有基础值, 后半部分为增加系数。各阶层间分界值为:

$$L = \begin{cases} 0 & n = 1 \\ \frac{(B_n - B_{n-1}) \times S_{n-1}}{S_n + S_{n-1}} + B_{n-1} & 1 < n < N + 1 \\ B_{n-1} \times 2 - L_{n-1} & n = N + 1 \end{cases}$$

第五, 如何为各个案最终赋值。对个案的赋值显然也是以基本分数为依据的, 由于已经通过阶层分界值固定了各阶层所包含的区间, 因此只需要确定各个案在该阶层内部所处的位置, 就可以得到其阶层地位分。即用该个案所属阶层的下界, 加上该个案的基本分数减去该阶层内个案基本分数的最低值, 除以该阶层内基本分数的最高值减该阶层内部基本分数的最低值, 再乘以该阶层上界和下界的差, 最终得到个体的阶层地位分值。

$$F = L_i + \frac{X - \min(X_i)}{\max(X_i) - \min(X_i)} \times (L_{i+1} - L_i)$$

（三）系数调整

在上述计算方法中，定义了一个用于界定经济因素和教育因素占比的系数 β ，并默认该数值为0.5。但显然该值并不具有足够的可信度，甚至不具有任何理论依据，因此有必要对该值进行调整。由于可信的 β 值的确定需要依托可信的阶层模型，而可信的阶层模型的确定又需要使用可信的 β 值，因此可以使用迭代的方法得到该值，通过反复计算逐渐与最终可信的结果靠拢。

具体方法是：（1）通过上述计算为所有个案赋予一个阶层地位分，然后以所得到的阶层地位分作为因变量，收入因素和教育因素作为自变量（在第二个基本问题中假设这两个变量符合正态分布），进行普通最小二乘法回归（OLS, Ordinary Least Squares），得到两个回归系数。（2）以教育因素回归系数的占比 [教育因素的回归系数 / （教育因素的回归系数 + 经济因素的回归系数）] 作为新的 β 值，重新进行上述计算。（3）如此反复计算，直到新得到的教育因素回归系数占比与该次迭代时使用的 β 值的差在0.0001以内，就可以得到调整后的占比系数。由于该迭代方法的基本性质，无论最开始默认的 β 值为多少，最后都会迭代出一个相同的结果，这也从能够侧面表明该迭代方法的科学性。

四、数据验证和结果分析

（一）数据验证

由于“阶层地位分”计算方法的一个前提是大样本基础，是一种

大数据思想的再现，因此本文选择了中国综合社会调查 cgss2011 到 cgss2017 之间共 5 次的调查数据进行测试。首先使用侯利明和秦广强（2019）的方法，使用 stata16 对 ISCO 编码进行了重新编码，得到所有个案所属的 EGP 阶层分类，并且依照他们的建议将原始的 10 分类模型缩减为 9 分类模型，其中自营农民和农业劳动者两个类别合并为一个类别，并将 6 个数据集合并到同一个数据集中，其分类框架如表 1 所示：

表 1 EGP 阶层框架（9 分类）

1	I	Higher service	高级专业技术人员或行政管理者，大企业管理者、大资产所有者
2	II	Lower service	较低专业技术人员或行政管理者，小型企业管理者
3	III	Routine clerical / sales	行政或商业部门的非体力雇员，其他服务业雇员
4	IV a	Small employer	自雇雇人的小资产者
5	IV b	Independent	自雇不雇人的小资产者
6	V	Manual foreman	低级技术人员，体力雇员的监督者
7	VI	Skilled manual	熟练技术型体力工人
8	VII a	Semi-unskilled manual	半技术和非技术的体力工人
9	IV c	Farmers / Farm managers	农业劳动者，自营农民

根据前述“阶层地位分”的设计思路与计算方法，数据验证的过程共分为六步，该验证计算过程使用 Python3.8 编程语言，使用 pandas, numpy 开源库进行计算，数据类型为 DataFrame。

表 2 “阶层地位分”的计算步骤

1	操作步骤 1	使用各个案的收入 (REV) 和教育水平 (EDU), 求出其基本分数 bCode
2	操作步骤 2	将数据集所有个案按照 EGP 阶层分类框架进行分类, 聚类到 9 个数据集中
3	操作步骤 3	求出各阶层的原始基础值和标准差
4	操作步骤 4	求出增加系数 X_i 并为所有基础值加上增加系数
5	操作步骤 5	计算各阶层分界值, 形成分界值列表
6	操作步骤 6	据各个案的基本分数和所处阶层, 确定各个案的阶层地位分值

经过上述步骤, 在完成单次计算后, 通过 statsmodels 开源库中的 add_constant 函数和 OLS 类进行普通最小二乘法的线性回归, 并得到新的 beta 值。在第 5 次迭代后, 教育因素回归系数占比与该次迭代的 β 值得占比之差小于 0.0001, 得到一个可信的结果, 该数值为 0.44175, 可以看出教育因素的占比相对少于经济因素。

(二) 结果分析

在前述对“阶层地位分”的计算方法内容中, 提出了两种基础值计算的方法, 除去第一种方法, 其中第二种方法可以拆分为从最高阶层开始计算、从最低阶层开始计算及结合同时进行计算的三种方法, 加之最后一种对 β 值进行了迭代求解的计算方法, 一共可分为五种方法。“阶层地位分”根据五种方法所得到的基础值如表 3 所示, 图式呈现如图 1 所示, 各阶层基础值差呈现如图 2 所示。

表 3 阶层地位分的基础值量化分值结果

阶层分类		方法 1 基础计算	方法 2 从高层 计算	方法 3 从低层 计算	方法 4 计算均值	方法 5 迭代后 均值
1	Higher service	11.7060	13.4164	14.2122	13.8143	12.8255
2	Lower service	10.5428	13.0899	11.7581	12.4240	11.5193
3	Routine clerical/sales	8.8955	11.6322	9.3265	10.4794	9.7457
4	Small employer	7.7062	10.2818	7.9998	9.1408	8.5565
5	Independent	6.1886	8.2339	6.8622	7.5480	7.0721
6	Manual foreman	5.9098	8.0378	6.7341	7.3859	6.9194
7	Skilled manual	4.3315	6.7177	4.6403	5.6790	5.3315
8	Semi-unskilled manual	3.1546	5.2324	3.2345	4.2334	3.9834
9	Farmers/Farm managers	1.1026	1.8288	1.1305	1.4797	1.3983

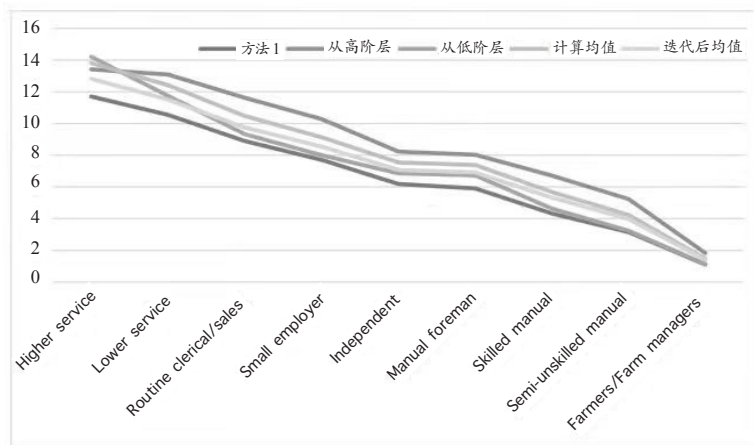


图 1 阶层地位分的基础值量化分值图示

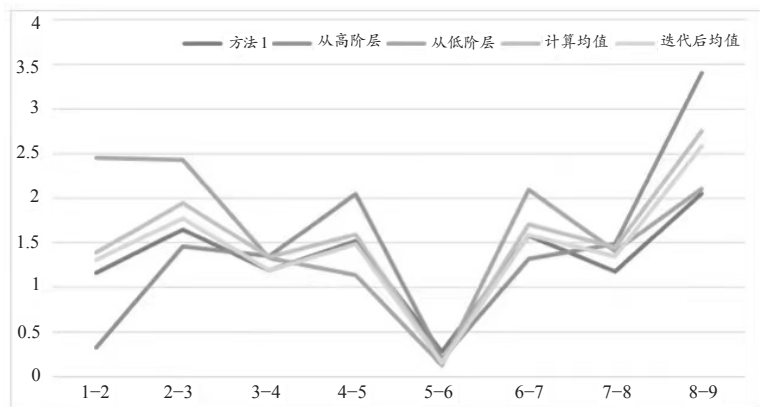


图2 阶层地位分的基础值差量化分值

从图表的显示结果可以看到，从最高阶层（方法2）和从最低阶层（方法3）开始计算时，都会明显导致最初计算的几个阶层的基础值差缩小，而最后计算的几个阶层的基础值差扩大。在方法1中，虽然各阶层基础值差的比例没有显著偏差，但是由于其只进行有限调整的性质，因而导致几乎所有的基础值差都被缩小了，最终使得结果无法完整的反映客观事实。计算均值法（方法4）可以较好地反映原数据集的实际情况，且其基础值曲线也与方法1较为相似，二者主要的区别在于：前者中两端阶层（阶层分类的1-3、6-9）的差值明显大于后者，而中间阶层（阶层分类的5-6）的差值则明显小于后者。在通过迭代法将占比系数 β 修改为0.44175后，除5-6阶层的差值外，所有阶层的差值相比于计算均值法都有一定的下降，且下降程度相似。而5-6阶层的差值则表现出了小幅度的增加，这表明相比于第6阶层的低级技术人员/体力雇员的监督者，第5阶层的自雇雇人的小资产者在教育因素上的优势要小于在经济因素

上的优势,这与经验事实相符,因而也表明迭代后均值求得的结果比计算均值求得的结果更具有说服力。

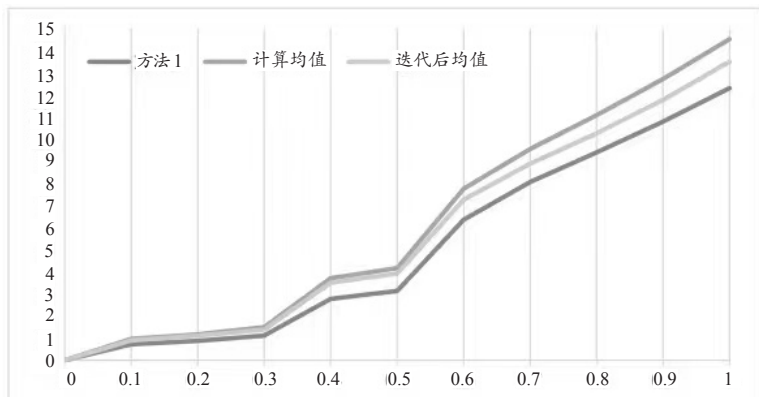


图3 阶层地位分的十分位数结果

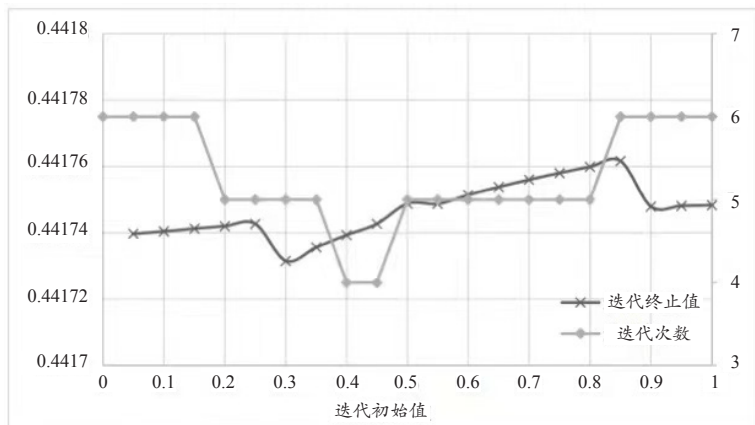


图4 阶层地位分计算的占比系数

为进一步判断迭代法的有效性,通过方法 1、计算均值法和迭代后均值法分别计算了数据集的十分位数结果,如图 3 所示。从图 3 的数据计算结果可以看出,迭代后均值(方法 5)相比计算均值(方法 4)和方法 1,在 0.9 分位数到最高值之间有一个细微的上扬,这显示出最上层和其下层存在更大的差距,这也与经验事实相符,表明迭代后均值能够更好地反映实际情况。

最后对计算阶层地位分的占比系数 β (即收入因素与教育因素的占比)进行讨论。图 4 显示了在以不同数值为初始值时,迭代后得到的占比系数 β 计算结果。可以看出,占比系数 β 最低值为 0.441731,该值是以 0.25 为初始值进行 5 次迭代后的结果,最高值为 0.441762,该值是以 0.8 为初始值进行 5 次迭代后的结果,在进行 4-6 次迭代后,最终 β 值都会取到 0.44175 附近。也就是说,在对阶层分类进行量化时,从数据中得到的收入因素与教育因素的占比大致会维持在这个水平,这也成为上述 β 值取 0.44175 的主要依据。使用 0.44175 作为占比系数也能够确定阶层分界值(表 4)以及最终的阶层地位分值(图 5),从而实现了阶层量化。

表 4 各阶层间界限值

阶层分类	计算均值	迭代后均值
最低值	0.0000	0.0000
1~2	2.9593	2.7967
2~3	4.9715	4.6727
3~4	6.5308	6.1258
4~5	7.4663	6.9944
5~6	8.3375	7.8087
6~7	9.7567	9.1059
7~8	11.5598	10.7293
8~9	13.1182	12.1699
最高值	14.5104	13.4811



图 5 阶层量化后的阶层地位分呈现图

五、结论与讨论

在社会分层与流动的研究中，关于社会阶层的分类及量化始终是一个基础前提。本研究提供一种可以将社会阶层进行量化的方法，将所得到的阶层量化结果命名为“阶层地位分”，由此推进从阶层分类到阶层量化的研究范式新发展。该方法不但可能解决阶层变量的量化问题，帮助后续研究简化模型、提升结果的可读性并可用于探查更多值得关注的结论，而且还通过占比系数的确定和调整，解决了在一定时空条件下无法确定收入因素和教育因素占比的问题。获取占比系数的迭代方法对于后续研究具有基础性的作用效果，该方法可以直接用于任何时空条件下，且操作化过程也不复杂。本研究使用中国综合社会调查 cgss2011 到

cgss2017 共 5 次的调查统计数据，建立了一个具有较大样本的数据集进行“阶层地位分”量化方法的验证，数据结果不仅能够支持提出的阶层量化方法，也为后续研究提供了诸多应用方式，主要包含以下可行应用。

第一，可以使用收入因素与教育因素的占比系数 0.44175 生成基本分数，然后直接在创建的数据集中找到与研究所需使用的数据集的个案具有相同基本分数和 EGP 分数的个案，并赋予该个案以相同的“阶层地位分”，该方法本质是一种个案对照法，如果研究者所使用的数据集较小，那么所生成的结果相对偏差也是最小的，但是应用过程会较为繁琐和困难。第二，可以使用 0.44175 作为占比系数生成基本分数，然后利用本研究中所提供的各阶层基础值和分界值，对新数据集中的个案按照前述方法进行重新赋值。第三，可以直接使用本研究所提供的计算方法对新数据集进行计算，并迭代出适用于新数据集的占比系数，需要注意的是，如果研究者所研究的是其他时空调查的数据，则必须对数据集进行重新计算。第四，可以将本研究的数据集以 EGP 分数和基本分数作为自变量，阶层地位分作为因变量进行普通最小二乘法回归，或者将教育因素和经济因素拆分为两个不同自变量进行回归，并将研究者所要研究的数据集中的数据代入回归方程得到其阶层地位分，不过该用法的一个主要缺陷在于，处于下一阶层的个案其阶层地位分可能会高于处于上一阶层的个案，从而导致数据自身的偏误。第五，可以直接将本研究所提供的各阶层的基础值作为个案的阶层地位分，该用法主要用于无法获得个案的收入或受教育水平之时，例如在 cgss 数据库中，缺少父辈的收入变量，因此在探讨阶层流动距离时，只能使用其所处阶层的基础值，此法由于并未使用阶层地位分，因此应用起来受限条件较少。

本研究所提供的方法,实际上是在有序分类变量基础上对连续型变量进行重新赋值,且赋值成一个定距连续型变量的数学方法,因此其在使用过程中有相比于一般方法更具有多样性。例如,本研究使用了9分类的EGP阶层分类框架,如果研究者不认可或其所使用的数据来源于其他地区,则可以依照研究者的偏好选择其他分类框架甚至更复杂的微观阶层分类方法。再如,本研究所使用的连续型变量包含教育因素和收入因素,研究者也可以依据所使用的分类框架自行选择其他变量作为该连续型变量。但是必须注意的是,所选择的变量都必须是符合正态分布的连续型变量,其占比系数数量应为 $N-1$ (N 为变量数)。

最后,本研究虽然提供了一个阶层量化的新方法,但是在算法层面仍具有一定改进空间。大数据模拟结果表明,如果所使用的分类框架和连续型变量之间出现了极端情况,例如二者的相关度过低或下一阶层的均值远高于上一阶层的均值,则可能出现无法计算出结果的情况。另外,本文对占比系数的确定,是建立在所有阶层均拥有相同占比系数的前提下,若每个阶层都有不同的占比系数,则该迭代法无法得出有效结论,且若使用的阶层数过多也会导致结果产生一定误差。虽然上述问题并不常见,但是仍期待未来能够持续修正文本所提出的算法,以进一步促进社会阶层研究的不断发展。

参考文献

- 陈云松、范晓光（2016）. 阶层自我定位、收入不平等和主观流动感知（2003-2013）. 中国社会科学, (12), 109-126.
- 陈云松、贺光烨、句国栋（2019）. 无关的流动感知：中国社会“阶层固化”了吗？社会学评论, (6), 49-67.
- 范晓光、陈云松（2015）. 中国城乡居民的阶层地位认同偏差. 社会学研究, (4), 143-168.
- 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, (2), 16-26.
- 李春玲（2001）. 建构社会阶层分类体系的几个问题. 中国人民大学学报, (2), 26-29.
- 李春玲（2005）. 当代中国社会的声望分层——职业声望与社会经济地位指数测量. 社会学研究, (2), 74-102.
- 李春玲（2019）. 我国阶级阶层研究 70 年：反思、突破与创新. 江苏社会科学, (6), 22-30.
- 李路路、石磊、朱斌（2018）. 固化还是流动？——当代中国阶层结构变迁四十年. 社会学研究, (6), 1-34.
- 李路路、秦广强、陈建伟（2012）. 权威阶层体系的构建——基于工作状况和组织权威的分析. 社会学研究, (6), 46-76.
- 李路路、朱斌（2015）. 当代中国的代际流动模式及其变迁. 中国社会科学, (5), 40-58.
- 李培林、张翼（2008）. 中国中产阶级的规模、认同和社会态度. 社会, (2), 1-19.
- 李强（2006）. 试析社会分层的十种标准. 学海, (4), 40-46.

- 李强 (2021). 21 世纪以来中国社会分层结构变迁的特征与趋势. 河北学刊, (5), 190-199.
- 李升 (2017). 主客观阶层位置与社会政治态度研究——兼论中国中产阶层的“稳定器”功能. 社会发展研究, (2), 73-88.
- 李升、倪寒雨 (2018). 中国城镇居民的中层意识研究——基于对工作状况、地区差异与生活方式的分析. 社会学评论, (4), 64-76.
- 李姚军、王杰 (2021). 社会流动与传统性别意识——以“干得好不如嫁得好”为例. 社会学评论, (2), 92-114.
- 刘欣 (2018). 协调机制、支配结构与收入分配：中国转型社会的阶层结构. 社会学研究, (1), 89-115.
- 刘欣、田丰 (2018). 社会结构研究 40 年：中国社会学研究者的探索. 江苏社会科学, (4), 33-46.
- 陆学艺 (2002). 当代中国社会阶层研究报告. 北京：社会科学文献出版社.
- 秦广强 (2011). 新韦伯主义阶级框架：起源与演变、操作与理论. 国外社会科学, (6), 95-103.
- 尚进云、王振振 (2020). 家庭背景、大学文凭与职业阶层地位获得的城乡差异——基于 EGP 职业阶层分类的新观察. 复旦教育论坛, (5), 79-85.
- 宋健、张晓倩 (2020). 主客观测量结果的一致性：原因分析与解决策略. 中国人口科学, (6), 112-123.
- 许琪 (2018). “混合型”主观阶层认同：关于中国民众阶层认同的新解释. 社会学研究, (6), 102-129.
- 张顺、祝毅 (2021). 代际流动轨迹与分配公平感——影响机制与实证分析. 社会学评论, (3), 199-218.
- 张翼 (2008). 当前中国中产阶层的政治态度. 中国社会科学, (2), 117-131.

赵晓航、李建新（2017）. 社会经济地位与社会政治态度——基于“中国家庭追踪调查”（CFPS2012）的实证分析. 社会学评论, (3), 34-50.

Tang, B. W., & Tan, J. J. (2021). Subjective social class and life satisfaction: Role of class consistency and identity uncertainty. *Asian Journal of Social Psychology*, (25), 60-74.

Benjamin S., David B., & Frenk S M. (2013). Class in Name Only: Subjective Class Identity, Objective Class Position, and Vote Choice in American Presidential Elections. *Social Problems*, 60(1), 81-99.

Blau P M., & Duncan O D. (1967). *The American Occupational Structure*. New York: Wiley.

Bukodi, E., Dex, S., & Goldthorpe, J. H. (2011). The Conceptualisation and Measurement of Occupational Hierarchies: A Review, A Proposal and Some Illustrative Analyses. *Quality and Quantity: International Journal of Methodology*, 45(3), 623-639.

Chan T W., & Goldthorpe J H. (2007). Class and Status: The Conceptual Distinction and Its Empirical Relevance. *American Sociological Review*, 72(4), 512-532.

Chen, M. (2018). Does Marrying Well Count More Than Career? Personal Achievement, Marriage, and Happiness of Married Women in Urban China. *Chinese Sociological Review*, 50(3), 240-274.

Erikson R., Goldthorpe J H., & Portocarero L. (1982). Social Fluidity in Industrial Nations: England, France and Sweden. *The British Journal of Sociology*, 33(1), 1-34.

Erikson R., Goldthorpe J H., & Portocarero L. (1983). Intergenerational class mobility and the convergence thesis: England, France and Sweden. *The*

- British Journal of Sociology*, 34(3), 303-343.
- Ganzeboom H B G., Paul M. De Graaf., & Treiman D. J. (1992). A Standard International Socio-economic Index of Occupational Status. *Social Science Research*, 21(1), 1-56.
- Goldthorpe J. H. (1987). *Social Mobility and Class Structure in Modern Britain*. Oxford: Clarendon Press.
- Grusky D B. (2001). *Social Stratification: Class, Race, and Gender in Sociological Perspective*. Boulder: Westview Press.
- Degraaf, N., Nieuwbeerta, P., & Heath, A. (1995). Class Mobility and Political Preferences: Individual and Contextual Effects. *American Journal of Sociology*, 100(4), 997-1027.
- Tahlin M. (2007). Class Clues. *European Sociological Review*, 23(5), 557-572.
- Walpole M. (2003). Socioeconomic Status and College: How SES Affects College Experiences and Outcomes. *Review of Higher Education*, 27(1), 45-73.
- Wright E. O. (1997). *Class Counts: Comparative Studies in Class Analysis*. New York: Cambridge University.