职业代际流动的对数线性模型*

孙 凤

ABSTRACT

Two way contingency tables are main research method for occupational generation mobility. More often than not, the researcher directly interprets the descriptive statistics and reaches substantive conclusions without resort to explicit modeling. However, the method is very imprecise when the research analyzes multi-way tables. In this paper, I introduce loglinear model for occupational generation mobility; Meanwhile I discuss urban Chinese occupational generation mobility according to a cross-classified data in 2000.

关键词：线性模型；列联表；交互影响；主要作用

一、问题的提出

父母所从事的职业是衡量家庭背景的一个重要指标，因为父母的职业等级能够反映一个家庭所掌握的经济权利、政治权利和社会声望等。鉴于此，许多社会学家通过相关数据的分析对各国父母职业与子女职业之间的关系进行实证研究，探讨家庭背景对劳动者就业的影响。如 Ducan(1967)通过对美国人口调查局的相关数据分析，发现美国存在明显的职业代际效应。Benjamin(1958)利用英格兰和威尔士的人口普查数据，发现在十三类职业分布中，父亲从事较高层次的管理工作，其儿子将近三分之二也从事管理工作；父亲是体力劳动者，其儿子相对多数也从事体力劳动。郭丛斌、丁小浩(2004)运用中国国家统计局2000年的全国城镇住户调查数据，分析中国的职业代际效应，发现职业在代际之间存在一定程度的流动性，但流动范围有一定程度的分割(见表1)。

笔者通过对以往代际职业流动的研究方法考证发现，这些研究方法主要是建立在父代职业和子代职业的二维列联表基础之上，用以反映职业的继承性和流动性。运用列联表研究问题，无论是简单的列联表还是复杂的列联表，其中所有频数之间分布的关联都可以分解为两种效应：一种反映了变量自身的频率分布影响，称之为主效应(main effect)；另一种反映变量之间关联所产生的效应，称之为交互效应(interaction effect)。对于两个变量构成的频数列联表是一张二维列联表，其主效应有两个，交互效应只有一个；当变量数增加时，交互的维数就会增加。

常规的列联表是二维列联表，通常只分析两个变量之间的联系，我们可以直接从列联表的分布中读取主效应和交互效应。然而，当进行多个变量的分析时，该方法就无法把握变量之间的关系了。

表1 **中国职业的代际继承性和流动性(2000)**

		子女职业						
		机关、企事业负责人	各类专业技术人员	办事人员	商业工作人员	服务性工作人员	生产运输工人	不便分类的其它项
父母职业	机关、企事业负责人	2.46	1.19	1.57	0.78	0.69	0.62	0.65
	各类专业技术人员	1.01	2.02	1.02	0.68	0.62	1.74	0.27
	办事人员	0.85	0.8	1.33	0.93	0.92	1.9	0.89
	商业工作人员	0	0.44	0.85	2.82	0.98	1.64	2.78
	服务性工作人员	1.05	0.87	0.51	1.53	1.71	0.99	1.25
	生产运输工人	0.5	0.52	0.5	0.91	1.35	1.57	1.28
	不便分类的其它项	0	0.15	0.9	1.04	1.12	1.38	2.94

引自：郭丛斌、丁小浩(2004)《职业代际效应的劳动力市场分割与教育的作用》，《经济科学》第3期。

在实际研究中，研究者通常采用一次只分析两个变量之间的交互表，经过多个两两交互分析，企图再拼接成多个分类变量之间复杂关系的整体。尽管这种做法也能得到一些信息，然而正如多个简单回归并不能代替多元

*本文受国家自然基金《中国消费者行为研究》(批准号70173040)和教育部人文社科基金《跨时选择与消费者行为研究》资助。

万方数据

回归一样,这种缺乏综合性的分析方式是不可能以多个个别分析叠加出整体的多元联系。特别是由于整个频数分布被分成多张二维交互表,只能大致分析每一张二维交互表的主效应和交互效应,而更多变量之间的联合交互效应(或高阶交互作用)将无法分析,然而,正是联合交互效应才真正反映代际职业流动之间的关联。

鉴于以上分析方法中存在的问题,本文认为 Loglinear 模型是一种有效处理高维列联表信息的统计方法,本文运用 loglinear 模型讨论中国职业的代际流动,关注的问题是在市场化转型中,职业代际流动具有什么特征,政策关注点在哪里。

二、Loglinear 模型

(一)数据介绍

研究代际间的职业流动需要掌握父代职业与子代职业信息,最好的数据来源应是人口普查或人口抽样调查数据。由于作者不能直接获得这样的调查数据,只能间接采用前人的分析数据。作者通过对现有文献的考察,认为郭丛斌、丁小浩(2004)所计算的中国代际职业流动表(见表 1)的数据,来源于国家统计局城市社会经济调查队 2000 年在全国范围内进行的城镇住户调查,从数据的调查和管理机构来看,该数据的来源渠道及抽样调查方式应该具有权威性。

表 1 中行变量为父亲的职业,列变量为儿子的职业。关于职业的划分和排序,由高到低分别为:机关企事业负责人、各类专业技术人员、办事人员、商业工作人员、服务性工作人员、生产运输工人、不便分类的其它劳动者。

表 1 反映代际职业的流动性指数和继承性指数。继承性指数测度的是代际职业的继承性水平;流动性指数测度的是代际职业的流动性水平。如果父母所从事的职业与子女所从事的职业类型彼此独立,则频数分布的观察值与独立假设下的期望值应该趋于相等,二者之比应该接近于 1,否则该比值远离 1。因此通常用观察值与期望值之比定义代际职业流动性指数($i \neq j$)和继承性指数($i = j$)。当比值大于 1,表明职业为 i 的子女进入 j 职业的可能性大;当比值小于 1,表明父母职业为 i 的子女进入 j 职业的可能性小。

独立假设通常采用卡方检验,公式为:

$$\chi^2 = \sum \frac{(f_{ij} - h_{ij})^2}{h_{ij}}$$

式中:f_{ij} 为实际观察频数,h_{ij} 为预期频数,i 和 j 分别代表父亲的职业和儿子的职业。卡方检验值越小,二者越独立。预期频数的计算公式为:

$$h_{ij} = \frac{f_{i+} f_{+j}}{f_{++}}$$

我们用符号 + + 表示合计,下标 $i+$ 代表行边际频数(父亲职业的边际频数),以下标 $+j$ 代表列边际频数(儿子职业的边际频数)。

表 1 呈现的是 2000 年中国代际职业的流动状况。对角线上的值为代际职业的继承性指数,非对角线上的值为代际职业的流动性指数。表 1 的计算结果表明,所有职业的代际继承性指数均显著大于 1,子从父业的继承性十分明显。

表 1 只给出了代际职业流动的基本描述,当职业的类型划分有很多,或者要进行不同国家、不同地区或不同人口特征的比较时(如考虑教育、地区、性别和年龄等因素),基于二维列联表就难以为之了,需要建立一个综合模型进行比较,loglinear 模型为我们提供了一个具有可操作性的方法。

(二)loglinear 模型

以 R 表示行变量(父亲的职业),C 表示列变量(儿子的职业)。R 的下标 i,$i = 1,2\cdots I$;C 的下标为 j,$j = 1,2\cdots j$,i 和 j 分别代表父与子的职业分类。以组频数作为因变量,行和列的分类作虚拟自变量所建立的回归模型即为对数线性回归模型,亦称泊松回归方程,表 1 可表示为:

$$\log(F_{ij}) = \beta_0 + \beta_1 R_1 + \mathrm{L}\ \beta_7 R_7 + \beta_8 C_1 + \mathrm{L}\beta_{14} C_{14} + \beta_{15} R_1 C_1 \mathrm{L} + \beta_{63} R_7 C_7 \qquad (1)$$

对数线性模型有一套专用符号系统,(1)式亦可以表示为:

$$\log(F_{ij}) = \lambda + \lambda_i^R + \lambda_j^C + \lambda_{ij}^{RC}$$

其中 F_{ij} 代表列联表中的频率,$i = 1,2,\cdots\cdots,i$,表示父亲的职业分类;$j = 1,2,\cdots\cdots j$,表示儿子的职业分类;λ 是总均值,λ^R 是行边缘(row marginal)效应,λ^C 是列边缘(column marginal)效应,λ^{RC} 是行列交互作用,交互作用反映的是行与列之间的相关。

如何反映行边缘效应和列边缘效应及其交互作用呢?首先需要对变量进行虚拟化处理。对于对数线性模型,常采用 0 总合限制,亦称为方差编码,如 $\beta1 + \beta2 = 0$,则 $\beta1 = -\beta2$。如:

$$\sum_i \lambda_i^R = \sum_j \lambda_j^C = \sum_i \lambda_{ij}^{RC} = \sum_j \lambda_{ij}^{RC} = 0 \qquad (3)$$

λ^R 有 $(i-1)$ 个参数;λ^C 有 $j-1$ 个参数;λ^{RC} 有 $(i-1)\times(j-1)$ 个参数。

Loglinear 模型与线性回归模型有很大的不同,突出表现为前者的因变量是频数而不是结果变量,结果变量和解释变量同时出现在 loglinear 模型中,这就使得研究者能够从模型参数推断出二者之间的关系;Loglinear 模型可以包含许多参数,在实际应用中,大量有意义的参数都是变量有交互作用的参数。

方程(1)和(2)是饱和模型,即包括行变量、列变量以及二者的所有可能关系的模型。除饱和模型外,还有非饱和模型,常见的有:

万方数据

1. 独立模型。

由假定交互作用为0得到:

$$\mathrm{Log}(F_{ij}) = \lambda + \lambda_i^R + \lambda_j^C \quad (4)$$

即当行变量与列变量独立时,类别频数只取决于行合计与列合计的分布。

2. 准独立模型(quasi-independence)。

$$\mathrm{Log}(F_{ij}) = \lambda + \lambda_i^R + \lambda_j^C + \delta_{ii} \quad (5)$$

其中 δ_{ii} 是对角线各类的参数。

观察表1的职业流动数据,可以发现对角线单元格的数值比较大,这些数值反映了"继承"的特点,人们希望将这种特征分离出来,进一步讨论其中所包含的信息,因此假定除了主对角线以外的行变量和列变量是独立的,我们称其为准独立模型。

(三)参数估计与拟合优度检验

Loglinear是非线性模型,参数估计的最好方法是极大似然估计法。运用极大似然估计法需要知道变量的分布特征。对于社会科学的列联表建模,主要以二项分布、泊松分布和多项式分布为主。

对数线性模型的统计检验包括模型的整体检验和单个参数估计的检验。模型的整体拟合优度检验主要有对数似然比卡方和皮尔逊卡方检验。对数似然比卡方检验的思想是以饱和模型的对数似然比为基准,计算对数似然比,衡量非饱和模型的拟合优度。

假如以Ls表示饱和模型的似然函数,Lr表示非饱和模型的似然函数,对数似然比卡方就是:

$$G^2 = -2\log\left(\frac{L_r}{L_s}\right) = 2\left(\sum_{i=1}^{I}\sum_{j=1}^{J} F_{ij}\log\left(\frac{f_{ij}}{F_{ij}}\right)\right) \quad (6)$$

当交叉表的数据不稀疏时,该统计量符合卡方分布。

单个参数估计的检验,有单项效应检验和单个参数估计的检验。单项效应检验反映的是如果从模型中撤销一个效应以后对似然比的影响,称为偏关联检验(tests of partial associations)。对于单个参数估计的检验我们用标准正态分布检验Z。

(四)实证分析

我们以表1的职业流动表为例,建立Loglinear独立模型和准独立模型,通过编制Stata程序估计模型参数,并进行拟合优度检验,估计结果见表2和表3。

从表2的估计结果看,(1)独立模型的各个参数均不显著,这表明各个职业间的流动性不强。(2)加入反映继承性的对角线影响后,该准独立模型则显示出职业间有较强的继承性特点,如职业层次位居上层的机关企事业负责人和专业技术人员的父母,其子女仍为机关企事业负责人和专业技术人员的发生比分别为1.9和1.5,较其它职业显著的高。位居中层的商业服务人员的父母,其子女仍为商业服务人员也显著的高,发生比为1.7。位居下层的不便分类的其他项,其子女仍从事同类职业的发生比为1.82,该数值仅仅低于机关企事业负责人的代际继承,排名第二。(3)从两个模型的检验指标看,准独立模型的总体拟合优度检验明显优于独立模型,这意味着考虑了职业的继承性后,使得模型的拟合优度提高了。

表2 loglinear模型参数估计表

	独立模型		准独立模型	
	参数	Z检验	参数	检验
常数	0.9237	2.52	0.5587	1.86
父职业:				
机关企事业负责人	0(a)			
各类专业技术人员	-0.224	-0.59	-0.1777	-0.61
办事人员	-0.1914	-0.50	0.0194	0.07
商业工作人员	0.0786	0.21	0.1029	0.35
服务性工作人员	-0.0071	-0.02	0.1857	0.64
生产运输工人	-0.1900	-0.50	-0.0223	-0.08
不便分类的其它项	-0.0614	-0.16	-0.0500	-0.17
子职业:				
机关企事业负责人	0(a)			
各类专业技术人员	0.0171	0.05	0.0637	0.22
办事人员	0.1157	0.30	0.3266	1.13
商业工作人员	0.4029	1.06	0.4271	1.47
服务性工作人员	0.2214	0.58	0.4143	1.43
生产运输工人	0.1386	0.36	0.3063	1.06
不便分类的其它项	0.5986	1.57	0.6100	2.10
子承父业:				
机关企事业负责人			1.9013	3.28
各类专业技术人员			1.5753	2.72
办事人员			0.4253	0.73
商业工作人员			1.7313	2.99
服务性工作人员			0.5513	0.95
生产运输工人			0.7273	1.26
不便分类的其它项			1.8213	3.15

注:(a)父职业和子职业的参照类均为机关企事业负责人;(b)置信水平为0.05。

表3 总体模型拟合优度检验

名称	对数似然比	自由度	BIC
(1)独立模型	-45.338686	36	-121.8493835
(2)准独立模型	-22.235262	29	-105.7527696

三、结论

本文采用由郭丛斌、丁小浩(2004)依据中国国家统计局2000年的全国城镇住户调查数据所建立的代际职业流动表,运用loglinear模型分析中国职业流动的代际效应。研究结果表明:

(一)从独立模型看,中国城镇职业间的流动性不强,特别是在不同职业间缺乏流动。这说明中国存在二元劳动力市场分割,即存在主要和次要劳动力市场。主要劳动力市场收入高、工作稳定、工作条件好、培训机会多,具有良好的晋升机制;而次要劳动力市场则与之相反,主要劳动力市场和次要劳动力市场之间相互分割缺乏流动。

万方数据

基本养老保险体系中个人帐户继承比例的优化分析*

张 勇

ABSTRACT

This paper designs an actuarial model of Individual Retirement Account (IRA), and analyses the relationship between the heritage and the divisor, which is used to calculate the pension of IRA. We find that the heritage will greatly decrease the amount of premiums used to pay pensions when making both ends meet. Then, we modify the present IRA model by reducing the heritage, and use them to increase the pensions. The positive outcomes show that the modified model is more steady and operational than the present one.

关键词：个人账户；继承数额；计发月数；精算现值

一、引言

根据国务院最近发布的《关于完善企业职工基本养老保险制度的决定》(以下简称《决定》)，个人账户作了重大调整：计发月数不再固定为120①，而是由职工退休时城镇人口平均预期寿命、本人退休年龄和利息等因素来确定，并且根据在岗职工平均工资年增长率的一定比例来调整养老金数额。

在此之前的许多学者认为，随着人均寿命的增加，计发月数固定为120将会导致个人账户的收不抵支，并对此进行了探讨(张畅玲，吴可昊，2003；段家喜，2005；罗良清和王茶香，2005；高建伟和丁克诠，2005)。对于养老金增长率，吕江林等(2005)认为可以采取等差数列形式，但由于通货膨胀呈几何级数增加，养老金购买力水平是逐年下降的；邓大松和刘昌平(2001)则以养老金增长率等于工资增长率为研究前提；程永宏(2005)则认为，由于退休者不直接参加生产活动，经济负担又低于在职职工，养老金增长率应低于工资增长率。

* 本文是本人博士论文相关内容进行整理后的结果，感谢导师王美今教授在研究和写作过程中提供的悉心指导。对于文章中可能出现的错误，全部由作者负责。

① 国务院在1997年发布了《关于建立统一的企业职工基本养老保险制度的决定》，规定个人账户养老金月标准额为个人账户储存数除以120。许多学者指出，此规定相当于假设退休职工的期望寿命为10年。

(二)从准独立模型看，中国城镇存在职业代际流动效应，对于不同的职业，都有明显的子承父业特征，特别是在收入和声望排名前两位的职业，如机关事业单位负责人和专业技术人员，其特征更加明显。相应地，父母在劳动力市场从事排序较低的工作，如不便分类的其他职业，其子女从事相同工作的可能性也比较大。这可以在一定程度上说明，劳动者工作所属的劳动力市场与其家庭背景有密切的关系，劳动者在进入劳动力市场之前，其家庭背景对他能进入主要或次要劳动力市场起了相对重要的作用。

(三)相关的政策建议涉及教育问题。经济学的人力资本理论认为教育具有提高劳动者个人收入，进而促进社会公平的作用。而社会学家布迪厄则认为教育具有复制社会生产关系，维系代际之间职业稳定的功能。两种理论差异背后的逻辑起点是教育的公平与否问题，这也是我们希望验证的问题，由于本项研究缺乏反映教育状况的数据，我们只能作为一个问题提出。

无论如何，使一个社会职业流动达到正常的重要举措就是保证教育的公平，从而能够为社会各阶层的劳动者子女提供较为公平的就业平台，这能够使得整个社会的职业系统有足够的平衡性和动力保持良性运转。否则，当教育不具公平性，教育就会成为社会优势阶层占有社会资源，并在代际之间传承的工具，这势必影响国家的长治久安。

万方数据

但是到目前为止,各种研究都没有考虑到个人账户的一个基本特征——剩余财产的可继承性(简称"当前模型")①。在当前模型中,基金平衡只是指缴费与养老金之间的精算平衡,而不是缴费与所有给付(养老金与继承数额之和)之间的精算平衡,得到的计发月数就小于实际计发月数,若按照此结果来确定养老金,个人账户就会出现收不抵支,在财务上是不可持续的。因此,只有弄清楚计发月数与可继承性之间的内在关系,才能合理确定计发月数和保持财务收支平衡,这正是本文所要研究的主要内容。

本文运用精算技术构建了个人账户的当前模型和实际模型,并进行比较分析。结果表明,当考虑到可继承性时,实际计发月数要远高于当前模型的计发月数;实际模型虽然能保持收支平衡,但只有小部分缴费用于支付养老金,继承人的收益程度超过了退休职工,个人账户没有发挥其养老保障功能。这表明,可继承性通过作用于计发月数而对个人账户的养老保障程度产生了很大的影响。为此,本文在不违背个人账户效率原则的情况下,对模型中存在的不足进行了修正,并得到了可操作性更强的解决办法。

《决定》没有说明计发月数和养老金增长率调整比例的具体操作办法,本文的结论在这方面提供了重要的参考价值,并且为类似问题提供了研究方法。

全文结构如下:第二节构建了我国个人账户的精算模型,分析了计发月数与继承数额之间的定量关系,并探讨了相应的修正办法;第三节对实际模型与当前模型进行比较实证分析,以及修正办法的合理程度;最后一节为结语。

二、个人账户的精算模型

1. 缴费。

假设职工在 t 岁时的工资等于 w_t,年工资增长率为 g_w,则有 $w_t = w_a(1+g_w)^{t-a}$;个人账户缴费率为 θ,职工在 t 岁时的年缴费额就等于 θw_t;起始工作年龄和退休年龄分别为 a 和 R;个人账户投资收益率等于 $i(i>0)$;${}_kp_t$ 表示 t 岁的职工在 k 年($k\geqslant 0$)之后还活着的概率,它是 k 的单调减函数。本文以年为时间单位,但只需相应的调整,就可变为以月为时间单位的结果。

在上述变量设定下,个人账户缴费的精算现值等于:

$$P = \sum_{t=a}^{R-1}\theta w_t v_{t-a}^{t-a} p_a = \theta w_a \sum_{t=a}^{R-1}\beta_{t-a}^{t-a} p_a \tag{1}$$

其中 $v=(1+i)^{-1}$,$\beta=(1+g_w)v$。P 与 θ 和 g_w 同方向变化,与 i 则呈反方向变化。

2. 养老金。

退休时个人账户储存额等于:

$$S = \sum_{t=a}^{R-1}\theta w_t v^{R-t} = \theta w_a(1+i)^{R-a}\sum_{t=a}^{R-1}\gamma^{t-a} \tag{2}$$

其中 $\gamma=(1+g_b)v$。i、θ 或 g_w 增加,S 就随之增加。

假设计发月数等于 m,最大生存寿命为 Ω,退休职工在 t 岁($R\leqslant t\leqslant\Omega$)时的养老金为 b_t。由《决定》可知,职工刚退休时领取的养老金为 $b_R = S/m$。如果养老金调整比例为 ϕ,即养老金增长率等于 $g_b=\phi g_w$,则有 $b_t = b_R(1+g_b)^{t-R}$。

以退休前一年的工资为基准,替代率水平等于 $\rho = b_R/w_{R-1}$②。计发月数越小(或收益率越高、退休年龄越大),替代率水平就越高。

经过化简,可得养老金的精算现值:

① 根据个人账户的相关规定,如果职工死亡时个人账户还剩有余额,则全部由继承人享有。以下把没有考虑可继承性的个人账户模型称为"当前模型",而把考虑到可继承性的个人账户模型称为"实际模型"。

② 王晓军(2002)分析了以最后工资来确定替代率的合理性。

(四)本项研究还需要提及的是关于 Loglinear 模型的应用。Loglinear 模型不仅能够分析代际之间的职业流动,更重要的是该方法作为进行高维列联表数据的处理方法,在进行属性数据分析上具有广阔的用途。我们只建立了两个变量的 loglinear 模型,实际上,在控制变量较多且可获得数据的情况下,我们还可以进行更高维度问题的研究。

参考文献

[1]Daniel A. Powers, and Yu Xie(2001), Statistical Methods for Categorical Data Analysis. Academic Press. Long, J. S., 1997, "regression models for categorical and limited dependent variables". Sage publication.

[2]Goodman, L. A. (1984), The Analysis of Cross-Classified Data Having Ordered Categories, Cambridge: Harvard University Press.

[3]Peter M, Blau; Otis Dudley Ducan. The American Occupational Structure, New York, 1967.

[4]李沛良.社会研究的统计应用,社会科学文献出版社.2001.

[5]郭丛斌、丁小浩.职业代际效应的劳动力市场分割与教育的作用.经济科学,2004.3.

作者简介

孙凤,1999年毕业于中国人民大学统计系获经济学博士学位,之后分别在南开大学经济研究所和美国密歇根大学消费者调查中心从事研究工作,现为清华大学人文学院副教授。主研究领域:社会科学定量分析方法、消费者行为、收入分配不平等。

万方数据