

# 人口流动与婚姻的历史关系回顾

## ——兼谈生命史数据处理

朱 农

生命史 (Life History), 是某个人一生中各种特征事件发生的过程。生命史调查, 即通过对被调查人的采访, 顺次记录下被调查者自进入社会生活到接受采访为止的期间内, 各种事件、行为 (如婚姻、职业、迁移、生育等) 发生的时间 (或年龄)。它从纵向时间序列上完整地提供被调查者各方面的信息, 使我们仅用一次问卷调查, 便能收集到丰富的数据。因而生命史调查在人口社会调查中占有十分重要的地位, 成为人口学研究的一个有力工具, 被国内外许多人口学者所应用。由于所获得的信息量大且具有随机性, 因此很多学者都致力于不断完善生命史调查的方法和数据处理技术。

生命史调查数据为若干组时间序列, 它的处理方法异于一般的社会调查数据。国外学者对此作过很多研究, 设计了多种数学模型, 并开发了一系列计算机软件与之配套。如SAS中的 Lifetest、Lifereg, BMDP中的PL1、PL2等。法国国家人口研究所的达尼埃尔·古诺教授 (Daniel Gourgeau) 是著名的迁移问题研究专家。他在其新著《Analyse Démographique des Biographies》中, 提出了一套关于生命史数据分析的新理论。其基本方法是运用数学模型, 考察不同情况下生命史中诸事件发生的概率, 揭示各种事件之间的相互关系。如结婚与迁移, 生育与迁移等, 以及其它社会、经济因素对这些事件的影响。笔者拟通过应用其中非参数模型和半参数模型, 结合我们的生命史调查数据, 对结婚与流动两种行为的相互作用, 及其它社会、经济特征的影响, 对如何运用这种数据分析方法作一扼要的介绍。

### 一、分析模型

此项研究分为两个步骤: 第一步是使用非参数模型分析结婚与流动的相互关系。所谓“非参数”, 意为不引入其它变量, 仅考虑这两种行为的相互作用。下面介绍这种模型的基本思想。

假定 $t$ 时刻, 未婚且未流动的样本数为 $N_0(t)$ , 已婚但未流动的人数为 $N_1(t)$ , 未婚但已流动的人数为 $N_2(t)$ , 以0、1、2分别代表这三种状态。定义 $T_1$ 和 $T_2$ 两个随机变量,  $T_1$ 为结婚年龄,  $T_2$ 为流动年龄, 从 $T_1$ 与 $T_2$ 出发, 可定义4个条件概率变化率, 以描述样本随时间的推移在各状态间的流动。其中:

$$h_{01}(t) = \lim_{\Delta t \rightarrow 0} \frac{P(T_1 < t + \Delta t \mid T_1 \geq t, T_2 \geq t)}{\Delta t}$$

$h_{01}(t)$ 为 $t$ 时刻,  $N_0(t)$ 中的结婚率, 即从状态0向状态1过渡的变化率。同理定义 $h_{02}(t)$ 为 $N_0(t)$ 中的流动率。此外:  $h_{21}(t/u)$ 为状态2下 $N_2(t)$ 个已流动但未婚的样本中的结婚率,

$$h_{21}(t/u) = \lim_{\Delta t \rightarrow 0} \frac{P(T_1 < t + \Delta t \mid T_2 = u, T_1 \geq t)}{\Delta t}$$

其中 $u$ 为流动年龄。同理定义 $h_{12}(t)$ 为 $N_1(t)$ 中的流动率。图1形象地说明了分析思想。

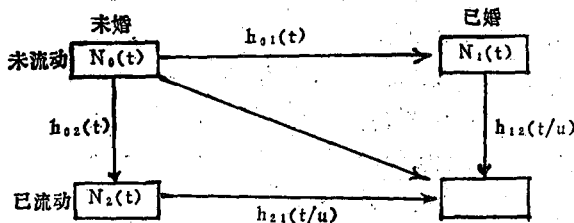


图1 婚姻与流动关系示意图

通过比较结婚前后的流动率, 即 $h_{02}(t)$ 与 $h_{12}(t)$ , 即可推断出结婚对流动的影响。例如,  $h_{02}(t) = h_{12}(t)$ , 则表明结婚与否对流动行为不起作用, 二者无关。同理, 比较 $h_{01}(t)$ 与 $h_{21}(t)$ , 可看出流动对结婚的影响。

在分析过程中, 样本整体从初始状态出发, 随时间 (或年龄) 的推移, 沿两条路径 (即先结婚再流动或先流动后结婚) 流向最终状态, 直到调查时

刻为止。其中有些人可能停留在某个状态下，如已流动但一直未结婚，则停在状态2；既未结婚又未流动，则停在状态0。此处时间t为年龄，计算出每个年龄下的4个变化率，然后进行比较。在计算中，流动和结婚同时发生的情况被均匀分布到4个变化方向上。

第二步使用半参数模型，考察其它因素（如文化程度、出生地等）对流动及结婚的影响。所谓“半参数”，意为假设这些被引入的变量对迁移和结婚的作用不随年龄变化。

变量分为两种。其一为独立变量，该变量在考察其对某一行为的影响时，与另一行为的发生无关。例如，在研究文化程度对流动的作用时，文化程度与另一行为结婚无关，即结婚与否对文化程度的高低不发生影响。其二为对应变量，即对应于另一行为，它只有在另一行为发生后才有意义。如在考察流动前的子女数对流动的影响时，子女数这个变量只有在结婚后才开始发生作用，在婚前没有意义。

模型的公式表述为： $h_{ij}(t, z) = h_{ij}(t) \exp [z\beta_1 + H(t-u)(\beta_0 + z'\beta_2)]$  ( $i, j=0,1,2$ )  
 $h_{ij}(t, z)$  为t时刻从状态i到状态j过渡的变化率， $h_{ij}(t)$  为一未知函数，其值为非参数分析中得到的结果。

Z为另一行为发生前引入变量的取值，因此时对应变量不起作用，所以Z仅包括独立变量；Z'为另一行为发生后引入变量的值，包括独立变量及对应变变量。因独立变量的取值与另一行为无关，所以其值不变，仍为Z中的值。

$\beta_1$ 、 $\beta_0$ 和 $\beta_2$ 为所要估计的回归系数。 $\beta_1$ 和 $\beta_2$ 分别对应于Z和Z'， $\beta_0$ 为另一行为本身的回归系数。模型假设各引入变量彼此互不关联。使用最大似然法则，取得系数的估计量，并对系数进行显著性检验，从而推断出这些因素对事件发生所起的作用。

变量的引入有两种方式：一是分别引入，即在模型中每次只引入一个变量，单独考察这个变量的作用；二是同时引入多个变量，综合考察诸变量各自的作用。本文将同时采用这两种方式，以更好地观察各种情况。

## 二、数据资料

1988年，武汉大学人口研究所与美国布朗大学人口研究所及培训中心合作，在湖北省进行了一次大规模的社会调查，主题为“迁移、生育与经济发展”，调查样本包括全省52个社区的4 390户家庭，这些家庭的人口总数为18 725人，其中7 702个15岁以上的人作为被调查者回答了问卷。

在本次调查中，我们使用了社区、家庭户和个人三级问卷，个人问卷中包括一部分生命史的问题，这些问题涉及婚姻史、职业史、流动史和生育史，调查顺次记录了被调查者15岁以来各种行为发生的年龄。表1为样本的城乡分布情况。

在问卷的生命史部分，“流动”一词的定义为常住地的改变，因为以下将分析结婚与流动之间的相互作用，所以将两种行为分别取为首次结婚和首次非婚姻性流动。表2为这次流动的原因。

由于样本总体的年龄跨度较大（从15岁至80岁以上），为了能从历史的纵向上来分析问题，将样本划分

	家庭户数	总人口数	个人卷数 <sup>①</sup>
城	1 035	4 768	1 673
镇	940	3 861	1 613
乡	2 415	10 096	4 408
合计	4 390	18 725	7 694

①生命史部分有效问卷数为7694份。

原 因	百分比
寻找更高的收入	34.3
变换职业	18.3
家庭原因	9.5
教育原因	4.3
其它原因	11.2
未 回 答	1.9

表3 样本的性别及代际分布

	一	二	三	合 计
男	1509 19.6%	1278 16.6%	1399 18.2%	4186 54.4%
女	996 13.0%	1099 14.3%	1413 18.4%	3508 45.6%

为三代人：生于1950年以前为第一代；生于1950～1959年的为第二代；生于1960年以后的为第三代。这三代人在进入青年期开始就业时，经历了各不相同的历史年代。第一代经历了共和国的创始阶段；第二代经历了十年文革；第三代经历了改革开放的巨大转变。每一代人的历程都带有他们所处的时代的痕迹。表3为按性别及代际划分后的样本分布。

由表3得到6个群体，分别对这6个组应用模型，然后对得到的结果进行比较分析。

### 三、结果讨论

(一) 结婚与流动的相互作用。首先看结婚对于流动的影响。图2为第三代人在不同婚姻状况下的按年龄累计流动率。其他两代人的情况也基本相同。

从图2可看出两个特点：(1) 男性较女性有更大的流动倾向；(2) 结婚在较大程度上降低了流动率。从表2可以知道，职业、经济因素在流动原因中占有重要位置，而在这方面男性通常比女性具有更大的主动性。另一方面，大部分人的常住地即是他们的户口所在地，在中国由于长期以来实行严格的户籍管理制度，户口的变迁一般不太容易，尤其是对于家庭来说。因此结婚以后，建立了家庭，常住地一般就比较稳定了。

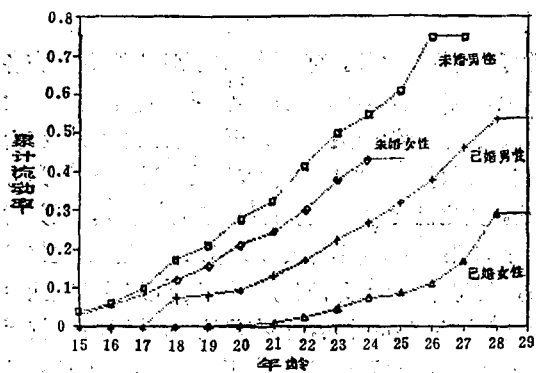


图2 按年龄累计流动率(生于1960年以后)

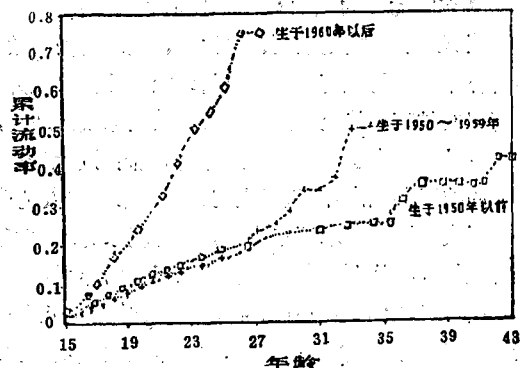


图3 按年龄累计流动率(未婚男性)

图3为不同代际未婚男性的按年龄累计流动率。从中可看出不同代际之间的差异。在中国，人口迁移受政策的影响很大，从图3可以看到，60年代后出生的这一代人的流动倾向远高于前两代人，这是由于他们刚步入就业年龄即遇上70年代末开始的经济体制改革。市场经济的繁荣与发展，极大地促进了城乡之间的人口流动，这个结果亦可以从第二代人的曲线中看到，第二代人的流动率在25岁以前与第一代基本一致，而在25岁左右以后明显上升。第一代人的流动倾向基本上变化不大。

反过来看流动对于结婚行为的作用。图4为第二代人流动前后的按年龄累计结婚率。另外两代人的情况也基本相同。

从图4可以看到，流动对于结婚的影响较之于结婚对于流动的影响要弱得多。流动稍稍推迟了结婚行为，尤其是对女性而言。但这种推迟只是将开始结婚的年龄推迟了3~4年，23岁以后，流动与否和结婚率之间基本上没有关系了。从代际纵向来看，流动的影响在年龄的代际中要显著些。对于50年代以前出生的女性，流动与否对结婚率基本没有什么影响。事实上，无论是过去还是现在，是农村还是城市，绝大部分青年都在30岁前结婚。

图5为流动后女性的累计结婚率，图5显示，第一代妇女结婚较早，17岁左右就开始结婚。结婚较迟的是

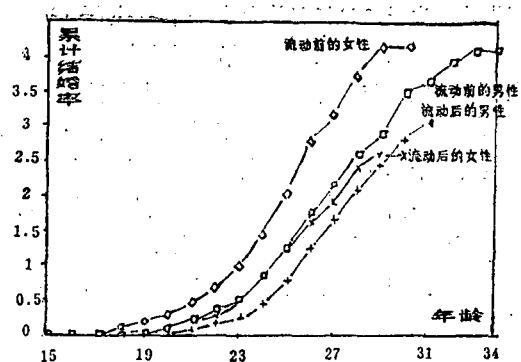


图4 按年龄累计结婚率（生于1950~1959年）

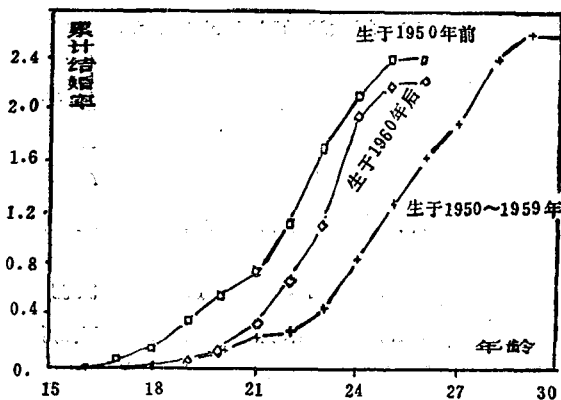


图5 按年龄累计结婚率（流动后的女性）

第二代人，这代人于70年代初步入婚龄，这正是中国大力推行计划生育政策，提倡晚婚晚育的10年，这代妇女的婚龄主要集中在23~30岁。进入80年代后，由于多种复杂的社会、经济因素的影响，如农村实行生产责任制，新婚姻法公布，青年人婚育观的变化等等，婚龄出现了提前的趋势。第三代妇女的婚龄比第二代大致提前了3岁。

（二）其它因素的影响。利用半参数模型，可以考察各种社会、经济特征对流动及结婚两种行为的作用。在此引入4个变量，两个独立变量：文化程度和出生地；两个对应变量：流动前的子女数（对应于结婚行为）和流动后的职业（对应于流动行为）。

表4和表5为文化程度、出生地、流动前子女数对流动行为的回归系数。可以看出，文化程度的效应在各群体中都非常显著。在结婚前，除第三代男性外， $\beta_1$ 的值均为正值，且都在5%的水平下显著。这表明受教育程度越高，其流动倾向愈大。从代际横向上看，在各代人中，女性的 $\beta_1$ 值都高于男性，即文化程度在女性的流动行为中起着更重要的作用，这同时也印证了我们在非参数模型分析中得到的结果：男性比女性具有更主动的迁移倾向。从代际纵向看，从第一代人到第三代人，文化程度对流动行为的作用逐渐减弱，尤其对于第三代男性，基本不起作用。这种现象的产生可能有两个原因：一是由于解放40多年来，全社会的文化程度普遍提高，从而逐渐抵消了文化程度的作用；另一方面，如前所述，大部分流动行为都是经济性或职业性的。近年来，由于各种因素的影响，农村青少年的文化教育受到了一定的冲击，在有些地区，辍学现象较严重，一部分升学无望的青少年放弃学业，过早地投入各种经济活动，如经商、进城做工等，使得在流动人口数量大为增长的同时，其素质较之以往有所下降。结婚以后，文化程度对流动行为依然起作用，有5组群体的 $\beta_1$ 值在5%或10%水平下显著，其中4组为负相关，似说明文化程度越高，流动倾向越低；从代际纵向上看，看不出明显的规律性，有待于进一步作更深入的研究。

现在来看出生地对流动的作用。在结婚前，年轻人和年纪大的人正好相反。年轻一代，农村的流动倾向高于城市，而年纪大者则是城市的流动倾向高于农村。这大致是因为一方面文化程度对流动倾向起着重要作用，而在老一代人中，城乡之间的文化程度差异较大，因此城市居民具有更强的流动倾向。另一方面，在过去很长一段时期内，农村市场经济很不发达，迁移政策对农业劳动力向城镇的流动控制很严，农村的人口流动受到限制。而第三代人由于正赶上经济体制改革，随着劳动力市场的开放，相当一部分农业剩余劳动力流入城镇，转入第二三产业。相形之下，城市居民由于受户口、住房、就业等因素的限制，其流动倾向反低于农村。结婚以后，出生地的效应均变为正向，各代人之间的差异有所缩小，流动倾向农村高于城市。可以说，农村家庭比城市家庭更具有灵活性。

在此引入“流动前的子女数”作为结婚行为的对应变量，从表6可以看到，每个组的 $\beta_2$ 值都有意义，均为负值。这表明子女数始终对流动起牵制作用。另外， $\beta_0$ 的值一般为负值，这符合我们在非参数分析中得到的结婚降低了流动倾向的结论。

现在从相反的方向来看流动行为、文化程度、出生地及流动后的职业诸因素对结婚行为的作用。表6和

表4 各种变量对流动行为的作用(分别引入)

		分 别 引 入		
		婚 前 $\beta_1$	结 婚 $\beta_0$	婚 后 $\beta_2$
文化程度				
男	生于1960年以前	0.419②	-0.481②	-0.181
	生于1950~1959年	0.294②	-0.007	-0.230②
	生于1960年以后	-0.003	-0.944②	-0.232②
女	生于1960年以前	0.460②	-0.061	-0.166①
	生于1950~1959年	0.451②	-0.472①	-0.209②
	生于1960年以后	0.135②	-1.872②	0.338②
出生地				
男	生于1960年以前	-0.317②	-2.285②	0.540②
	生于1950~1959年	0.101	-1.868②	0.458②
	生于1960年以后	0.242②	-1.777②	0.428①
女	生于1960年以前	-0.486②	-1.986②	0.621②
	生于1950~1959年	-0.339②	-2.275②	0.477②
	生于1960年以后	0.258②	-1.843②	0.128
流动前子女数				
男	生于1960年以前		-0.138	-0.219②
	生于1950~1959年		-0.140	-0.206②
	生于1960年以后		0.599②	-0.799②
女	生于1960年以前		0.495①	-0.288②
	生于1950~1959年		-0.199	-0.416②
	生于1960年以后		-0.432①	-0.627②

①结果在10%水平下显著

②结果在5%水平下显著

表7为计算结果,由于数据分布的关系,遗憾未能估计出第一组男性的诸系数。

从表7可知,在流动以前,文化程度与结婚行为呈负相关,文化程度越高,婚龄滞后现象越明显,尤其是青年一代;流动后则呈正相关,文化程度越高,结婚越早,特别是女性。

在流动前,每个组的出生地系数都有意义,均为正值,表明农村的结婚倾向始终高于城市。而且越年轻, $\beta_1$ 值越大,说明随着时间的推移,结婚倾向城乡差别越来越大。在流动后,出生地基本上不起作用了,这是由于相当一部分流动是迁向城镇,出生地在流动后也就失去了重要性。

对应变量(迁移后的职业)对结婚行为的影响效果不很清楚,只有两个组的值有意义,均为正值,表明流动后职业的社会经济地位越高,结婚越早,这与以上文化程度在流动后的效应相类似。

最后,从各组的 $\beta_0$ 值可以证实前面非参数分析中得到的结果:流动对结婚的影响不大,只是略微推迟了结婚年龄,尤其是对较年轻的两代人。

#### 四、结论

生命史的调查固然是人口社会调查中一项强有力的工具,但也正由于它所包含的信息量非常大,而且又

表5 各种变量流动行为的作用(同时引入)

		同 时 引 入		
		婚 前 $\beta_1$	结 婚 $\beta_0$	婚 后 $\beta_2$
男	生于1950年以前			
	文化程度	0.395②		-0.114②
	出生地	-0.099		0.451②
	结婚		-1.058②	
女	生于1950~1959年			
	文化程度	0.360②		-0.206②
	出生地	0.280②	-0.818①	0.485②
	结婚			
性	生于1960年以后			
	文化程度	0.046		0.189①
	出生地	0.260②		0.661②
	结婚		-1.717②	
女	生于1950年以前			
	文化程度	0.394②		-0.104
	出生地	-0.213①		0.514②
	结婚		-0.468	
性	生于1950~1959年			
	文化程度	0.447②		-0.174①
	出生地	-0.015		0.526②
	结婚		-1.100①	
性	生于1950年以后			
	文化程度	0.234②		0.313②
	出生地	0.405②		0.538①
	结婚		-2.561①	
性	流动前子女数			-0.608②

①结果在10%水平下显著

②结果在5%水平下显著

表6 各种变量对结婚行为的作用(分别引入)

		分 别 引 入		
		流动前 $\beta_1$	流 动 $\beta_0$	流动后 $\beta_2$
文化程度				
男	生于1950年以前			
	生于1950~1959年	-0.096②	-0.211	-0.011
	生于1960年以后	-0.224②	-0.498	0.126①
女	生于1950年以前	-0.225②	-0.056	0.113①
	生于1950~1959年	-0.256②	-0.854②	0.243②
	生于1960年以后	-0.358②	-0.867②	0.215①
出生地				
男	生于1950年以前			
	生于1950~1959年	0.217②	-0.203	-0.035
	生于1960年以后	0.473②	-0.031	-0.087
女	生于1950年以前	0.203②	0.089	-0.014
	生于1950~1959年	0.290②	-0.150	-0.100
	生于1960年以后	0.450②	0.249	-0.266②
流动后职业				
男	生于1950年以前			
	生于1950~1959年		-0.388②	0.049
	生于1960年以后		-0.512②	0.127②
女	生于1950年以前		0.123	0.083
	生于1950~1959年		-0.729②	0.142①
	生于1960年以后		-0.424②	-0.011

①结果在10%水平下显著

②结果在5%水平下显著

是回顾性调查,因而对实地调查工作提出了很高的要求。在中国,由于一些社会、技术条件的限制,调查质量往往难以控制,而且数据处理的方法不多,所以很少使用。引进一些国外的最新研究成果,对于推进我们今后的研究工作,具有重要的意义。

以上介绍的两种模型较之全参数模型,具有更大的灵活性,适用范围更广。通过模型的计算,将看来似乎杂乱无章的随机变量归纳成在时空分布上具有规律性的成果。由于模型将样本整体作为时间的函数,使我们能从动态上观察各种行为的发生情况和各种因素的影响。以上只考虑了首次流动和首次结婚,事实上还可将模型连续应用于第二次、第三次以至整个生命史的时间序列,考察各种行为之间更复杂的关系。

在分析中,我们假设各个变量分别独立地发生作用,严格说,这些变量并非完全各自独立,如文化程度、出生地及职业等,其间就存在着内在联系。另外,由于模型考察的是一个较长的时间序列的各种参数之间的相互关系,还必须注意到各个历史时期对这些事件发生的影响。如50年代末至60年代初的干部下放,文革时期的上山下乡,这都属于非正常的人口流动。又如子女数这个变量,就直接受各个时期生育政策的影响。由于这些原因,难免会出现一些不易解释,乃至反常的计算结果。因此在使用上述模型及其它数学模型时,仍必须从历史状况出发,密切联系人口、社会、经济各方面的实际,作深入、细致的分析。

(本文责任编辑:郭汉英)(作者工作单位:武汉大学经济学院人口研究所)

表7 各种变量对结婚行为的作用(同时引入)

		同 时 引 入		
		流动前 $\beta_1$	流 动 $\beta_0$	流动后 $\beta_2$
男	生于1950年以前			
	文化程度	-		-
	出生地	-	-	-
	流动 流动后职业			-
性	生于1950~1959年			
	文化程度	-0.046①		-0.056
	出生地	0.192②	-0.141	-0.051
	流动 流动后职业			0.072
女	生于1960年以后			
	文化程度	-0.133②		0.117①
	出生地	0.419②	-0.769①	-0.032
	流动 流动后职业			0.138②
性	生于1950年以前			
	文化程度	-0.203②		0.143①
	出生地	0.096②	-0.011	0.031
	流动 流动后职业			-0.074
女	生于1950~1959年			
	文化程度	-0.219②		0.270②
	出生地	0.145②	-1.486②	0.116
	流动 流动后职业			-0.177①
性	生于1960年以后			
	文化程度	-0.299②		0.186①
	出生地	0.271②	-0.494	-0.133
	流动 流动后职业			0.028

结果在10%水平下显著

结果在5%水平下显著