

中国女性婚后离家模型

——模型的建立、检验及估测其主要参数 α 与 β 的解析法

曾毅 李晓丽 马忠东

婚后离家往往意味着新住户的建立。它的变化直接引起家庭、住户数量、规模与结构的变动,关系到住宅、家用消费品市场与社会家庭福利规划。对这一社会现象的了解不但具有学术价值,而且有很大的实际意义。我们曾在另文^①利用中国现有的12省(市)数据,运用生命表分析方法估算了不同省(市)不同时期按结婚年数分的婚后离家率以及其它有关统计指标,对初婚时居住方式与婚后离开公婆(父母)另立门户的现象、过程以及对50~80年代的变动趋势进行了剖析。但是,对于其它并不具备这些详尽数据的地区,能否利用它们已有的或可估算的若干有关主要参数来估计婚后离家率?如果我们希望做家庭、住户变动趋势的预测,为社会经济规划服务,是否可以把主要精力集中在估测几个主要参数上,再依据这几个主要参数估算出详细的按婚后年数分的离家率?而建立一个符合中国实际情况的婚后离家模型正可以达到以上目的。本文第一部分根据中国统计局1985年、1987年组织的7省2市深入生育力调查;北大人口所1988年组织的湖南、山西调查以及台湾省的有关调查数据,首次建立了适合中国情况的婚后离家模型。第二部分对所建立的婚后离家模型进行了统计学检验。第三部分首次提出了估测婚后离家模型及与之类似的其他Brass-Gompertz相关模型的主要参数 α 与 β 的解析法。

一、婚后离家模型的基本原理

一般来说,检验人口学模型的成功与否,有以下几个标准:(1)能找到因数据缺乏而不易直接估算的待估变量与若干主要参数之间的解析或经验统计关系;(2)所建立的模型能较好地拟合实际观察数据;(3)作为模型输入的主要参数应尽可能少;(4)这些输入的主要参数应较易估算(即能从不完全的信息中估算出来),而且必须具有人口学、社会学或经济学意义,即含义明确、通俗易懂,便于根据它们的取值的变化对所研究的人口过程进行定性与定量相结合的分析与预测。最好能体现对人口过程的控制目标。这样,我们就有可能在人口、社会预测或规划中把主要精力集中在估测这几个主要参数上,再根据它们得出预测或规划所必需的详尽的年龄别或婚龄别率。

基于以上标准,在研究了许多前人建立的人口学模型的基础上^②,我们根据Brass-

① 曾毅、李晓丽、贾同金、杨日章:《初婚时居住方式与婚后离开公婆(父母)另立门户的生命表分析》《人口动态》,1991年第二期(专刊)。

② W. Brass, 1978. The Relational Gompertz model of fertility by age of women. London School of Hygiene and Tropical Medicine, 1978. H. Booth, 1984. Transforming Gompertz's function for fertility analysis: the development of a standard for the relational Gompertz function. Population Studies, vol.38, No.3, PP.495-506.

Gompertz相关生育模型的基本思路,建立了中国女性婚后离家模型。

定义:

$d(i)$ ——婚后离家频率或简称婚后离家率,即婚后第 i 年离家人数与该队列(Cohort)妇女总人数之比。

$H(i)$ ——婚后离家频率到婚后年数 i 的累积直,不包括结婚时不与公婆(父母)一起居住的妇女。

$$H(i) = \sum_{x=0}^i d(x)$$

T ——结婚时与公婆(父母)一起居住,但最终将离开公婆(父母)另立门户的妇女占同批(Cohort)妇女总数(包括结婚时不与公婆或父母一起居住的人)的百分比。

$$T = \sum_{x=0}^W d(x)$$

其中 W 是可发生婚后离家事件的最高婚后年数。

将在结婚时不与公婆(父母)一起居住(或婚前离家)的比例作为一个独立的参数 C ($C \geq 0$)。

$$(T+C) \leq 1$$

假定 $H(i)$ 与 T 之比服从Gompertz分布

$$H(i)/T = \exp(A \exp(B \cdot i))$$

其中 A 、 B 为不随 i 变化的待估参数

对上式做双负对数变换,有

$$-1n(-1n(H(i)/T)) = -1n(-A) - B \cdot i$$

$$\text{记为: } Y(H(i)/T) = -1n(-A) - B \cdot i \quad (1)$$

其中 Y 表示双负对数转换

用各省市数据检验: $(H(i)/T)$ 的双负对数转换与 i 之间除很高的婚后年数外有比较好的线性关系,说明关于 $H(i)$ 与 T 之比服从Gompertz分布的假定基本能成立,但离严格的Gompertz分布还有一定的距离。于是,我们需要借助于一个“标准”来建立一个相关模型,以便更好地概括与描述实际的离家过程。

假定我们能找到一个标准模式

$$Y_s(H_s(i)/T_s) = -1n(-A_s) - B_s \cdot i \quad (2)$$

式(2)中的下标“ s ”表示“标准模式”。

(1) - (2) * B/B_s , 有

$$Y(H(i)/T) - \frac{B}{B_s} Y_s(H_s(i)/T_s) = -1n(-A) + \frac{B}{B_s} \cdot 1n(-A_s)$$

或

$$Y(H(i)/T) = \alpha + \beta \cdot Y_s(H_s(i)/T_s) \quad (3)$$

$$\text{其中 } \alpha = -1n(-A) + \frac{B}{B_s} \cdot 1n(-A_s)$$

$$\beta = \frac{B}{B_s}$$

表1 婚后离家的标准模式的双负对数转换值

婚后年数	转换值	婚后年数	转换值
1	-0.775958	9	1.59989
2	-0.233141	10	1.84701
3	0.138111	11	2.13382
4	0.466018	12	2.48215
5	0.718678	13	2.88704
6	0.957671	14	3.37048
7	1.172270	15	4.15817
8	1.379780		

我们将从深入的生育力调查等资料中得到的7省2市和湖南、山西50、60、70年代以及台湾省70年代共21组“Y”转换值的均值作为标准模式（见表1）。前人（包括Brass本人）的研究经验表明，取不同时期不同地区的观察值序列的均值作为“标准模式”，在理论上与实践上都是可行的。

二、婚后离家模型的检验

为了检验我们建立的“标准模式”以及方程（3）表述的关系是否可信，我们利用12个省（市）不同时期观察到的 $H(i)$ 与 T 值数据以及“标准模式”，用最小二乘法分别对每一省（市）每一时期估算出 α 与 β 值，然后根据式（3）利用估算出的 α 、 β 值、观测到的 T 值以及“标准模式”反过来估算出 $H(i)$ 值，从 $H(i+1)$ 与 $H(i)$ 的差值可估算出 $d(i)$ ，再与观测的 $d(i)$ 值比较。我们发现拟合效果相当令人满意。绝大部分拟合曲线与观测曲线基本重合，有的不完全重合的部份，实际上是拟合曲线平滑了观测曲线因少数年龄误报导致的不规则。这是对观测值的一种改进。另外有三条曲线大部份拟合较好，只在起始点（即婚后0~1年）拟合较差（参见图1）。令人满意的拟合效果证明我们建立的“标准模式”以及式（3）表述的观测序列与“标准模式”的双负对数转换值之间的线性关系是可信的。

然而，以上拟合所用“标准模式”包括了被拟合省（市）本身的详细观测数据（离家率）。人们也许要问，假如“标准模式”并未包括被拟合省（市）本身的离家率，拟合能令人满意吗？换一句话说，“标准模式”是12个省（市）的离家率的Y转换值的均值，用它来拟合这12个省的离家率效果很好，但是若用它来拟合其它没有被包括在“标准模式”中的省或地区，其效果是否也能令人满意呢，是否被拟合地区的离家率一定要包括在“标准模式”中才能使模拟效果良好？为了回答这一问题，我们对所有12个省（市）每一时期的离家率都作了一个新的拟合。新的拟合所用的“标准模式”不包括被拟合省（市）本身的数据，而是除被拟合省（市）以外的11省（市）离家率的Y转换值的均值。我们发现，使用不包括被拟合省（市）数据的“标准模式”的新的拟合效果也同样令人十分满意。图1中的两条拟合曲线基本重合，它们与观测曲线都十分吻合。同时，我们还给出了量测这两套拟合效果的定量指标 R^2 （列在表2中）^①。除北京1950~1969年结婚队列的 R^2 值在0.88~0.89之外，所有的 R^2 值都在0.9以上，其中超过一半在0.95以上。而且同一地区同一时期根据包括与不包括本省（市）数据的两个不同的“标准模式”拟合的 R^2 十分接近，一般差异在第三位小数。以上结果有力地证明，我们用12个（或11个）省（市）数据建立的“标准模型”囊括了中国东西南北以及经济水平上中下各种不同类型的婚后离家特征与模式，具有较好的代表性。一般来说，良好的模拟效果不一定要将被模拟地区的数据包括在“标准模式”中。因此，我们可以用根据12个省（市）数据建立的“标准模式”来估算或预测其它没有被包括在“标准模式”中的省（市）或地区的婚后离家率。

① R^2 是一种常用的用来量测拟合效果的统计指标，当观察值与拟合值序列完全绝对相等时， R^2 等于1。 R^2 值离1越近（远），表明拟合效果越好（差）。

三、预测婚后离家模型主要参数 α 与 β 的解析法

至此,我们建立了一个“标准模式”,并得出了某一 $[H(i)/T]$ 序列的双负对数转换与标准模式的双负对数转换的线性关系,即婚后离家模型。这个模型还得到了中国12个省(市)不同时期观察数据的验证。从代表这个婚后离家模型的方程(3)可知,只要估算出了

T 、 α 与 β 值,便可估算出 $H(i)$,进而根据相邻 $H(i)$ 的差估算出 $d(i)$,即在婚后 i 年离家率。 T 实际上是最终将离家另立门户的比例 TT 与初婚时不与公婆(父母)一起居住的比例(c)之差。这里的 TT 实际上是类似总和生育率或总和初婚率的总和离家率。它可以根据过去的或类似地区的调查数据以及对未来家庭变动趋势的预测进行估计。根据Brass以及其它学者的研究,参数 α 反映了完成所研究过程(这里是婚后离家)的快慢(或称速度);参数 β 则反映频率分布曲线的宽窄程度,实际上除决定曲线的形状外,也与过程的速度有关^①。从现有的12省(市)数据估算出的 α 值在 $-0.539 \sim 0.585$ 之间, β 值在 $0.879 \sim 1.122$ 之间。尽管 α 与 β 的统计学意义明确,但它们都不是人口统计调查研究中容易估算或预测的参数。因此,我们有必要找到一种根据较容易估计或预测的、人口学意义明确的参数来估算 α 与 β 的途径。

Denton与Spencer曾提出一种根据中位数(median)、第三与第一个四分位数之差以及总和生育率,利用牛顿—拉斐桑非线性迭代法(Newton-Raphson non-linear iterative procedure)求Gompertz生育模型的主要参数 α 与 β 的方法^②。这种方法不宜照搬到我们的婚后离家模型的参数估计中来。其原因在于我们的模型并非严格的Gompertz模型。它实际上是一种经过改进的相关(Relational)Gompertz模型,通过一个“标准”来更好地描述实际离家过程。而且,Denton与Spencer方法假定生育过程严格服从Gompertz分布也是不符合实际情况的,于是Brass等人创立了Brass-Gompertz相关生育模型。但是迄今为止Brass等人并未提出在缺乏详尽数据时预测Brass-Gompertz相关生育模型的主要参数 α 与 β 的方法。下面,我们将推导出一种计算Gompertz相关离家模型主要参数 α 与 β 的十分简便的解析法。

表2 拟合(A)*与拟合(B)*的 R^2 值

省(市)	1950~1969初婚队列		1970~1979初婚队列	
	拟合(A)	拟合(B)	拟合(A)	拟合(B)
河 北	0.981068	0.982097	0.929218	0.924031
上 海	0.954782	0.953788	0.966813	0.966761
陕 西	0.950568	0.949726	0.948411	0.945067
北 京	0.888852	0.877912	0.948750	0.936259
辽 宁	0.977216	0.971531	0.959415	0.951563
山 东	0.964619	0.958139	0.941076	0.931917
广 东	0.931097	0.924198	0.948547	0.942334
甘 肃	0.915934	0.900360	0.935591	0.924283
贵 州	0.966792	0.963573	0.984930	0.983434
湖南、山西	0.982639	0.979461	0.978858	0.976119
台 湾	—	—	0.975302	0.973104

*拟合(A): 根据包括被拟合省(市)本身的所有12省(市)数据的“标准模式”的拟合值。

拟合(B): 根据除被拟合省(市)本身以外的11省(市)数据的“标准模式”的拟合值。

① United Nations, 1983, Manual X “Indirect Techniques for Demographic Estimation”, United Nations, New York, pp.25-26.

② F.T.Denton and Byron G.Spencer, 1974, Some demographic consequences of changing fertility Patterns: an investigation using the Gompertz function, Population Studies, Vol.28, No.2, pp.309-318.

记M为中位数，即初婚时与公婆（父母）一起居住，但最终将离家的妇女中的50%已离家的婚后确切年数。分别记N与O为第一与第三个四分位数，即分别为初婚时与公婆（父母）一起居住，但最终将离家的妇女中的25%与75%已离家的婚后确切年数（参见图3）。

根据方程（1）可得

$$Y(H(O)/T) = -\ln(-A) - B \cdot O \quad (4)$$

$$Y(H(N)/T) = -\ln(-A) - B \cdot N \quad (5)$$

根据方程（2）可得

$$Y_s(H_s(O_s)/T_s) = -\ln(-A_s) - B_s \cdot O_s \quad (6)$$

$$Y_s(H_s(N_s)/T_s) = -\ln(-A_s) - B_s \cdot N_s \quad (7)$$

（5）式减去（4）式得

$$Y(H(N)/T) - Y(H(O)/T) = B(O - N)$$

（7）式减去（6）式得

$$Y_s(H_s(N_s)/T_s) - Y_s(H_s(O_s)/T_s) = B_s(O_s - N_s)$$

由于 $Y(H(O)/T)$ 与 $Y_s(H_s(O_s)/T_s)$ 相等， $Y(H(N)/T)$ 与 $Y_s(H_s(N_s)/T_s)$ 相等，所以

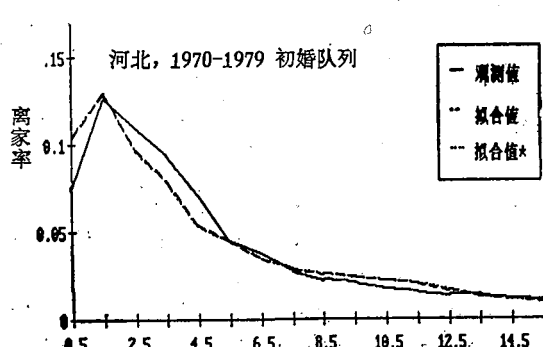
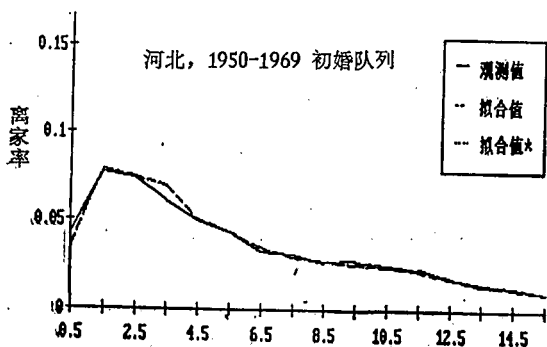
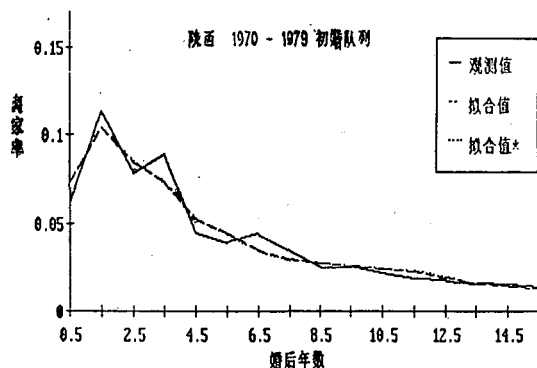
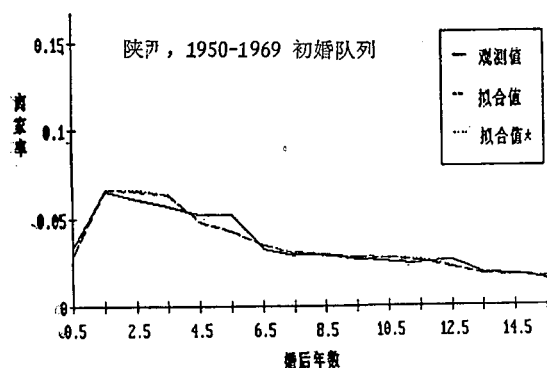
$$B(O - N) = B_s(O_s - N_s)$$

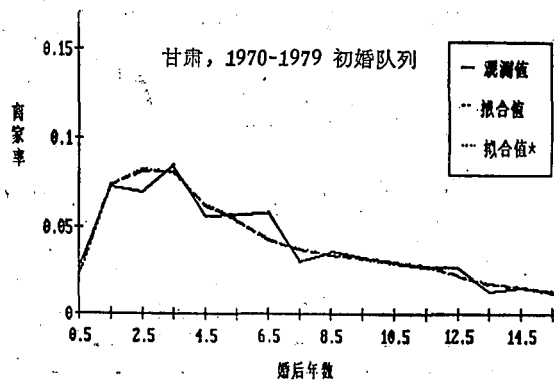
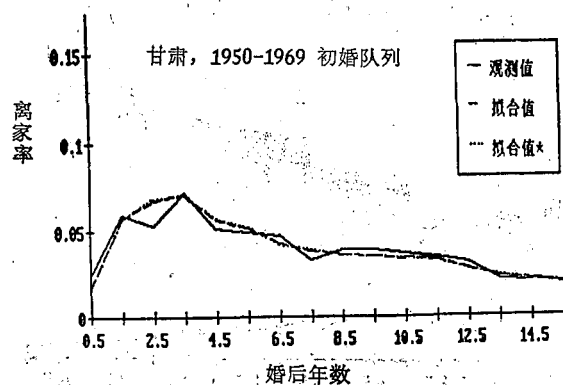
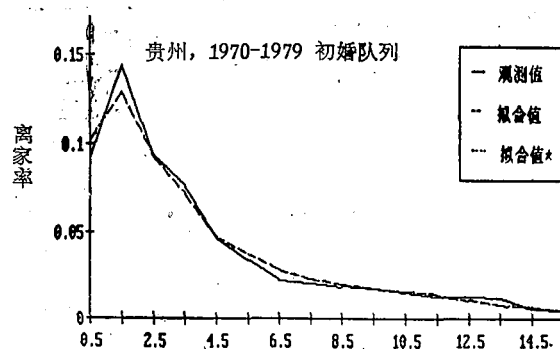
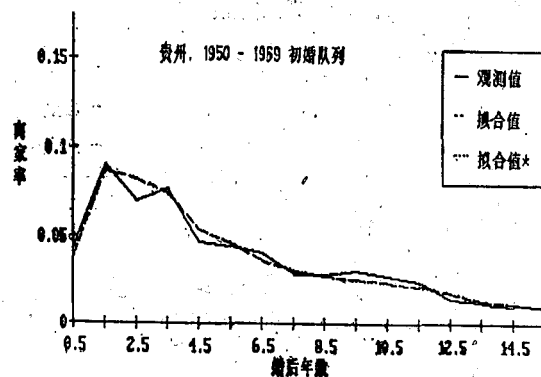
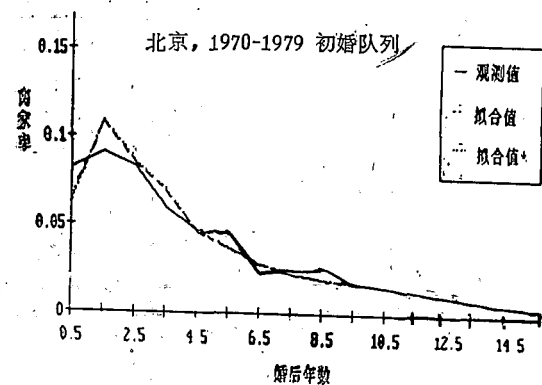
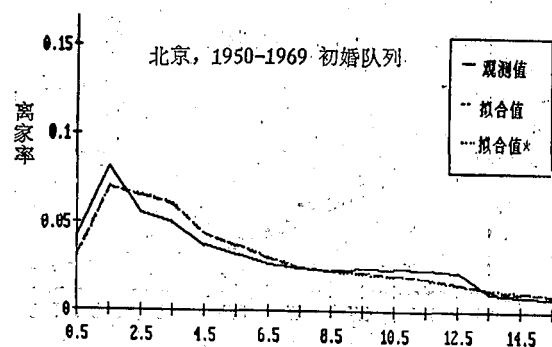
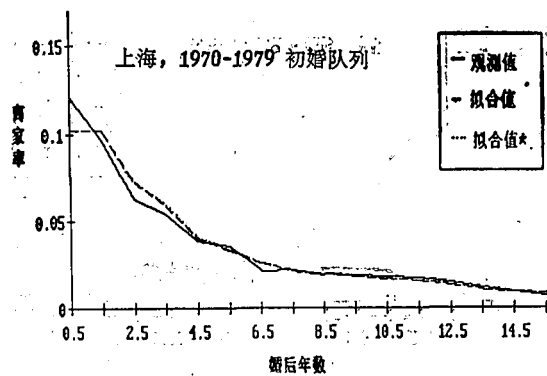
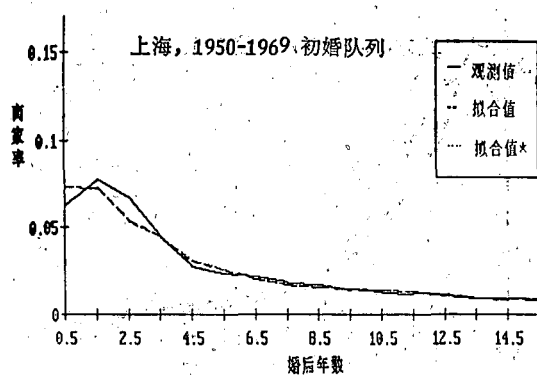
$$B/B_s = (O_s - N_s) / (O - N)$$

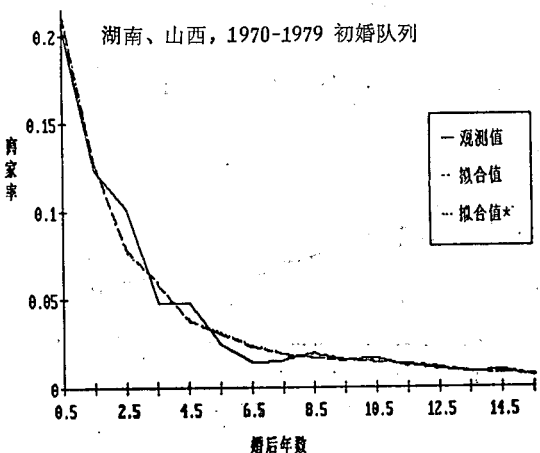
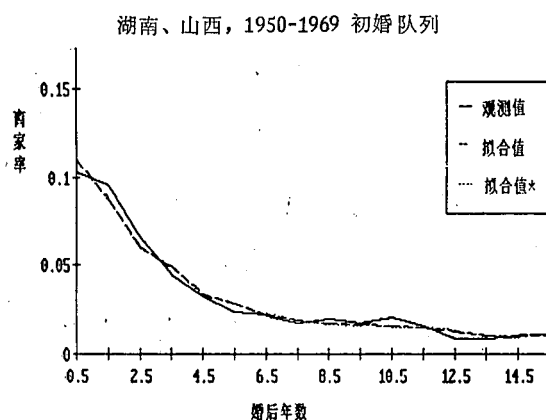
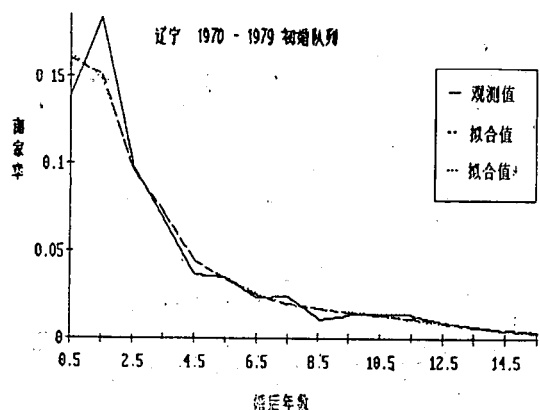
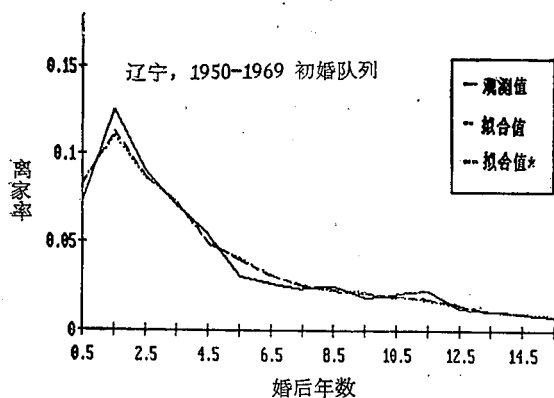
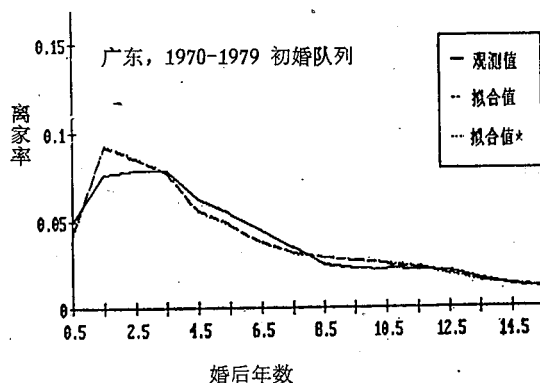
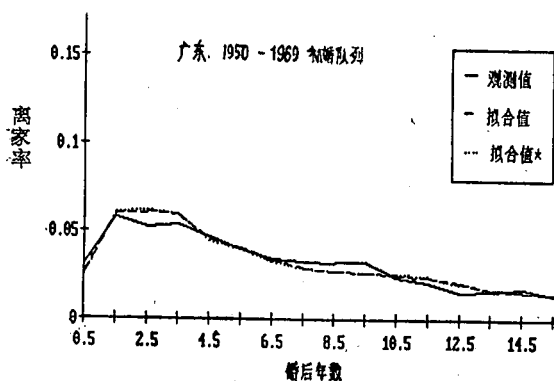
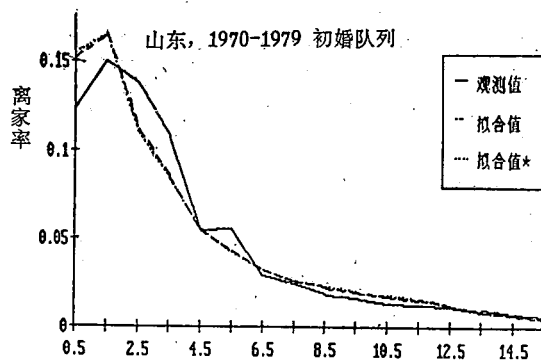
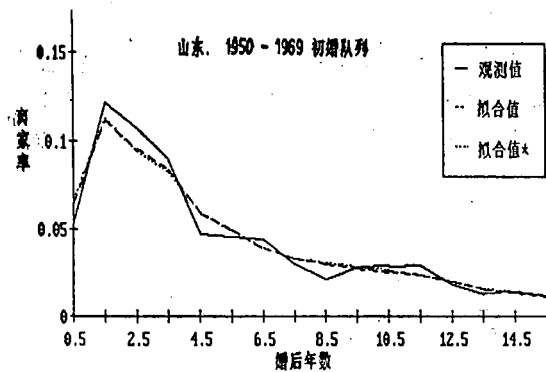
前面已约定 $\beta = B/B_s$ ，因此

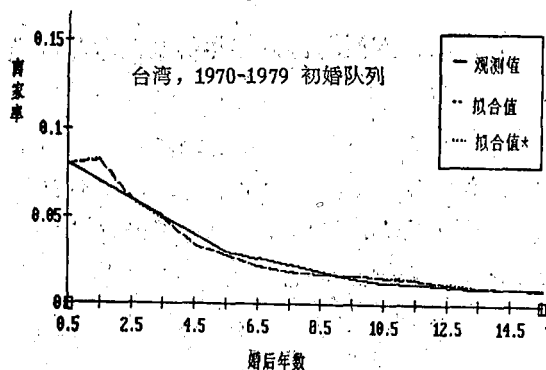
$$\beta = (O_s - N_s) / (O - N) \quad (8)$$

式（8）中的 $(O_s - N_s)$ 是我们已建立的“标准模式”的第三与第一个四分位数之差，其值为 $7.37198 - 3.70599 = 3.66599$ 。式（8）中的 $(O - N)$ 是待估测的离家率的第三与第一









注：——拟合值：根据包括被拟合省(市)本身的所有12省(市)的“标准模式”的拟合值。
拟合值*：根据除被拟合省(市)本身以外的11省(市)数据的“标准模式”的拟合值。

图1 婚后不同年数离家率的模型
拟合值与观察值的比较

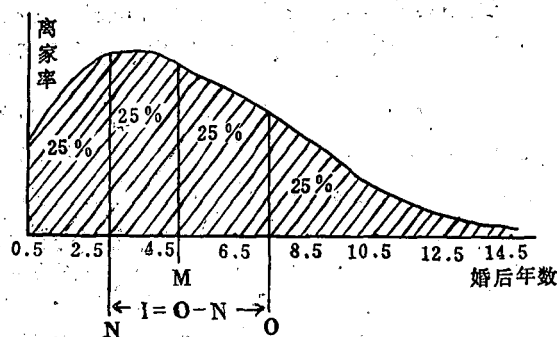


图2 婚后离家率的中位数M, 第一个四分位数N, 第三个四分位数O, 以及第三个与第一个四分位数之差I的定义示意图

个四分数之差, 我们不妨记 $(O - N)$ 为 I (参见图2), 式(8)就成为:

$$\beta = 3.66599 / I \quad (9)$$

于是, 我们只要估测所研究地区妇女婚后离家率的第三与第一个四分数之差 I , 即可很方便地算出 β 。同时, 我们还可以比较容易地估测出所研究地区妇女婚后离家率的中位数 M 。一旦 β 算出, 中位数 M 被估测出, 计算另一个主要数 α 就成为非常简单的代数运算问题。根据方程(3)可得

$$Y(H(M)/T) = \alpha + \beta \cdot Y_s(H_s(M)/T_s)$$

其中 β 已知, $Y(H(M)/T) = -\ln(-\ln(0.5))$, 也是已知的, $Y_s(H_s(M)/T_s)$ 可从表1中通过线性内插求得, 于是 α 可以非常方便地得到。

四 结语

本文首次建立的适合中国情况的婚后离家模型、估测婚后离家模型和与之类似的其它 Brass-Gompertz 相关模型的主要参数 α 与 β 的解析法, 不但具有学术价值, 而且具有较大的实用价值。其实用价值主要在于两个方面: 第一, 用于弥补或修匀婚后离家率。例如, 某地区没有详尽的婚后离家率或虽有这类数据, 但因年龄误报与漏报等原因, 离家率曲线上波动很大, 数据质量欠佳。这时, 我们只要根据已有的不完全信息或参照类似地区的有关数据, 估计出本地区的婚前离家比例 C 与最终离家比例 TT 、离家中位数 M 、以及第三与第一个四分位数之差 I (参见图2)、我们即可估计出详尽的婚后单年年数别离家率, 用于进一步分析与规划。第二, 用于未来家庭、住户规模与结构的预测。在中国核心家庭与三代家庭并存的社会条件下, 任何比较精细的家庭、住户预测都必须考虑并测算婚后离家过程。这就离不开对未来婚后离家率的估测。本文建立的模型与解析法正可以帮助解决这一问题。我们只要根据本地区或类似地区过去的有关数据以及我们对未来初婚时居住方式、婚后离家过程的提前或推迟、最终离家比例的发展趋势的推断^① 估计出因人口学意义很强而较易于预测的主

① 为了帮助读者了解婚后离家率的中位数 M 以及第三个与第一个四分位数之差 I 的取值范围, 我们将12省市不同时期的 M 、 I 值等统计量指标列在附录中, 各省不同时期初婚婚前离家比例与最终离家比例等数据请参阅第1页注①。

要参数C、T、M与I,即可利用我们建立的模型与解析法很方便地得出详尽的婚后单年年数别离家率,以用于家庭、住户规模、结构预测研究与分析。

笔者谨对J.Vaupel, J.Cleland, W.Brass与涂平博士在本文修改过程中提出的宝贵意见,以及王正联工程师对本文制图提供的帮助表示衷心感谢。笔者十分感激国家统计局为本次研究提供了大陆九省(市)有关数据以及张明正,齐力教授提供了台湾省的有关数据。

附录 婚后离开公婆(父母)另立门户过程的统计学指标

省(市)	初婚队列	N ^①	O ^②	I ^③	M ^④	\bar{X} ^⑤	SD ^⑥
河 北	1950~1969	2.329	8.632	6.303	4.650	5.698	0.0207
	1970~1979	1.814	6.359	4.545	3.481	4.629	0.0383
上 海	1950~1969	1.614	7.376	5.762	3.305	4.855	0.0225
	1970~1979	1.225	6.941	5.716	3.097	4.521	0.0327
陕 西	1950~1969	2.712	9.487	6.775	5.298	6.265	0.0163
	1970~1979	1.919	7.734	5.815	3.874	5.250	0.0296
北 京	1950~1969	2.006	8.957	6.951	4.489	5.641	0.0191
	1970~1979	1.635	6.451	4.816	3.392	4.485	0.0296
辽 宁	1950~1969	1.666	6.949	5.283	3.346	4.686	0.0338
	1970~1979	1.185	4.749	3.564	2.225	3.519	0.0531
山 东	1950~1969	2.002	7.619	5.617	3.774	5.178	0.0337
	1970~1979	1.469	5.178	3.709	2.833	3.866	0.0511
广 东	1950~1969	2.801	9.359	6.558	5.510	6.317	0.0142
	1970~1979	2.402	7.868	5.466	4.526	5.528	0.0242
甘 肃	1950~1969	3.296	10.146	6.850	6.121	6.809	0.0152
	1970~1979	2.926	8.912	5.987	5.292	6.101	0.0241
贵 州	1950~1969	2.233	8.296	6.064	4.396	5.491	0.0241
	1970~1979	1.448	5.506	4.057	2.826	4.057	0.0407
湖南、山西	1950~1969	1.267	7.021	5.754	2.886	4.522	0.0301
	1970~1979	0.830	4.670	3.840	2.115	3.520	0.0550
台 湾	1950~1969	1.555	7.017	5.461	3.555	4.762	0.0234

①N表示第一个四分位数,即对于初婚时与公婆(父母)一起居住,但最终将离家的妇女来说,25%已离家的婚后确切年数。

②O表示第三个四分位数,即对于初婚时与公婆(父母)一起居住,但最终将离家的妇女来说,75%已离家的婚后确切年数。

③I表示O与N的差值,即第三与第一个四分位数之差。

④M表示中位数。对于初婚时与公婆(父母)一起居住,但最终将离家的妇女来说,50%已离家的婚后确切年数。

⑤ \bar{X} 表示均值。初婚时与公婆(父母)一起居住,但最终将离家的妇女的平均离家婚后确切年数。

⑥SD表示婚后离家率的标准差。

(本文责任编辑:徐 莉)(作者工作单位:北京大学人口研究所)

(上接第7页)

4.曾牧野:《沿海新潮与广东改革》,广东高等教育出版社,1989年。

5.张敏如等:《中国沿海地区人口流动与管理》,中国广播电视出版社,1989年。

6.Gary S.Becker: An Economic Analysis of Fertility, Princeton University Press, 1960.

7.Juliam L.Simon: The Economics of Population Growth, Princeton University Press, 1977.

8.Lester R.Brown et al: State of the World 1990, A Worldwatch Institute Report on Progress Toward Sustainable Society, W.W. Norton and Company, New York, 1990.

(本文责任编辑:王跃生)

(作者工作单位:中国社会科学院人口研究所)