

中国女性初婚年龄与 不婚比例的参数模型估计*

封婷

【摘要】文章使用中国分城乡 1946~1980 年女性出生队列初婚数据,分析中国女性队列初婚年龄模式、初婚年龄推迟的趋势,以及晚婚转化为不婚的可能性。同时,对比了广义对数伽马模型、Hermes 扩散模型和改进的广义对数逻辑斯蒂模型对中国女性初婚年龄模式的拟合效果。结果显示,中国女性初婚年龄分布集中、对称性强;1970 年之后出生队列初婚率达到峰值的年龄逐步推后、初婚年龄分布变得分散、对称性和不同群体间异质性增强,初婚推迟。城市女性初婚晚,分布分散,农村女性异质性较强。初婚延迟和队列初婚年龄模式使 1970 年之后出生的女性终身未婚比例上升,预计 1980 年出生队列达 1.48%~6.39%,其后出生的队列趋势变动加快,中国女性普婚的传统或将被打破。

【关键词】初婚年龄 模式 参数模型 终身未婚比例

【作者】封婷 中国社会科学院人口与劳动经济研究所,助理研究员。

一、研究背景

婚姻是个人生命历程中的重要事件,也往往标志着新家庭的开始。社会中的成员,尤其是女性,结婚时间的早晚、不婚比例的高低对整个社会的家庭结构变化、生育行为、迁移流动、死亡等重要人口事件有着深刻的影响。20 世纪中期以来,发达国家普遍出现了初婚推迟和不婚比例上升的现象。尽管东亚女性传统上普遍结婚,但相关研究表明这一传统正在消退(Jones 等,2009;Jones,2018)。近年来,中国女性进入婚姻的年龄普遍推迟,结婚率呈下降趋势,30~34 岁女性未婚率从 2000 年“五普”的 1.35% 升至 2010 年“六普”的 5.35%,再升至 2015 年 1% 人口抽样调查的 6.95%。女性结婚率的下降在一定程度上造成了一孩生育率的下降(郭志刚,2017)和其他社会问题。因此,分析中国女性初婚年龄与终身未婚比例的特征与变化趋势已成为人口与婚姻家庭领域的重要课题。

* 本文为国家社科基金重大项目“人口统计调查的国际前沿理论及其在中国的应用”(编号:16ZDA090)的阶段性成果。

目前探讨初婚推迟的影响因素的研究(郭志刚、段成荣,1999;王鹏、吴愈晓,2013;於嘉、谢宇,2013;李建新、王小龙,2017),大多运用计量模型考察影响婚姻事件的社会经济因素、个体特征等,但对于婚姻状况的变化规律研究不足。首先,在快速变迁的社会、经济、文化背景下,特定时期具有突出影响的社会经济特征很可能在下一时期不再显著,并且这些变量自身的含义也在发生变化,由此导致既有的研究结论难以外推。其次,在个体层面,吸引力、个性与信念、情感需求等个体特征难以测度,无法纳入计量分析。择偶偏好、适婚两性群体特征和婚配时机的匹配,难以重构和量化,导致这些分析虽使用微观模型却失之于对个体决策机制的把握。在这方面,人口统计学中初婚年龄分布的参数模型提炼了可被反复验证的初婚普遍规律,具有独到优势。已有研究表明,使用这些模型可以很好地刻画不同人口的初婚进程,更加准确地预测终身未婚比例(黄荣清、亓昕,1990;Liang,2000;Bloom等,1990;Brüderl等,1995;Goldstein等,2001;Kaneko,2003)。

有学者基于普遍结婚背景和假设,发现 Coale-McNeil 模型、Hernes 扩散模型和对数逻辑斯蒂模型均能较为准确地拟合中国女性初婚年龄的时期数据(黄荣清、亓昕,1990)。鉴于国内已有研究大多未涵盖终生未婚的情况,本文对比不同参数模型,分析中国城乡女性队列初婚年龄模式及变动特征,通过拟合和外推等方式来检验模型适用性,进而估计终身未婚比例。

二、初婚年龄分布的参数模型

个体的初婚行为虽然取决于个体特征、偏好、适婚对象等方面的特殊性,但由于受社会制度和文化的规范调整,汇总至宏观层面会表现出规律性——队列初婚开始后,随着年龄增长在各年龄有不同比例成员进入初婚、已婚比例不断增长直到增长逐渐停止的动态过程,可以用年龄分布和累积分布来反映,表现为特定的分布模式。然而,初婚年龄分布通常具有“钟”形、峰度高和右偏的特征,常见参数模型不适用,为此,人口统计学界从经验数据、社会学假设和模型形式改进等方面发展了各具特色的参数模型。

(一) Coale-McNeil 模型及其等价形式

寇尔曾提出初婚年龄模式具有一致性(Coale,1971;Coale等,1972),使用初婚年龄分布标准模式能很好地拟合荷兰、美国和中国台湾地区等婚姻制度迥异的人口,既包括传统上平均初婚年龄低于15岁、99%进入婚姻的东亚女性,也包括平均初婚年龄高于25岁、50岁曾婚比例低于80%的某些欧洲国家的初婚模式。这说明初婚行为在不同人口中服从特定的年龄别进程规律(Rodriguez等,1980)。Coale-McNeil模型在广泛应用中拟合效果较好(Bloom等,1990;Liang,2000)。Kaneko(2003)提出一种广义对数伽马(GLG)分布,即:

$$g(a) = \frac{|\lambda|}{b\Gamma(\lambda^{-2})} (\lambda^{-2})^{\lambda^{-2}} \exp \left\{ \lambda^{-1} \left(\frac{a-u}{b} \right) - \lambda^{-2} \exp \left[\lambda \left(\frac{a-u}{b} \right) \right] \right\} \quad (1)$$

其中, $-\infty < u < \infty$ 为位置参数, $b > 0$ 为尺度参数, λ 为形状参数。当 $\lambda < 0$ 时, GLG 分布右偏, 且与 Coale-McNeil 模型等价。

总人口与初婚人口 a 岁概率分布存在以下比例关系:

$$f(a) = Cg(a) \quad (2)$$

其中, C 为终身曾婚比例。标准 Coale-McNeil 模型是 GLG 模型 $\lambda = -1.287$ 时的特例, $\lambda = -1$ 时模型退化成极值分布, 很多常见分布也是固定某些参数后的特例, 可以认为 GLG 模型拟合效果会优于这些分布。其优点包括: u 为分布众数, 图示直观, 分布单峰右偏使众数在 30 岁前就能观察到; λ 很好地概括了初婚分布曲线形态特点, 可用于队列间的比较; 通过分析固定 λ 预测能有效降低误差, 30 岁左右外推预测误差将小于 1%, 预测准确度较高 (Kaneko, 2003)。因此, 本研究使用 GLG 模型分析中国女性初婚数据。

(二) 扩散模型

扩散模型常用于刻画创新或疾病的扩散过程, Brüderl 等 (1995) 将其一般形式记为:

$$\frac{dF(a)}{da} = f(a) = s(a)F(a)^m [1 - F(a)]^n \quad (3)$$

其中, $F(a)$ 为人口中事件发生的累积分布。初婚事件的扩散表现为队列进入初婚的速度 $dF(a)/da$ 既与已扩散的已婚比例 $F(a)$ 有关, 也与待扩散的未婚比例 $[1 - F(a)]$ 有关。

1. Hernes 非同质扩散模型

Hernes (1972) 认为, 随着队列已婚比例升高, 未婚成员进入初婚的社会压力增强, 但初婚能力随年龄增长持续下降。一是从结构上看, 早结婚群体一般初婚能力更强; 二是动态来看, 随着队列初婚进程持续, 未婚群体特征发生了变化, 如随着年龄的增大, 对异性的吸引力减小、适龄未婚异性规模缩减及行为固化很难为结婚做出相应的调整。因此在式 (3) 中引入随年龄持续降低的 $s(a) = Ab^a$ 。其中, A 是初始的平均初婚能力, $b < 1$ 反映初婚能力随年龄增长下降的速率, $m = 1, n = 1$ 。积分求得的已婚比例 $F(a)$ 随年龄增长趋近于终身曾婚比例, Hernes 模型形式隐含了终身曾婚比例的估计, 不需要单独引入参数, 比 Coale-McNeil 模型 4 个参数减少 2 个。

Hernes 模型形式简洁, 应用效果至少与 Coale-McNeil 模型相近 (Burch, 2018; Diekmann, 1989; Coale 等, 1996; Goldstein 等, 2001), 参数含义直观, 连续队列参数的变动能捕捉初婚模式变化趋势 (Hastings 等, 1973)。Goldstein 等 (2001) 发现, 使用 30 岁前的初婚数据, 累积初婚比例预测误差小于 2%, 仅使用 25 岁前的数据, 误差仍低于 3.5%。

2. 对数逻辑斯蒂模型的改进

式 (3) 中取 $s(a) = \frac{p}{a}$, $m = n = 1$, 或 $s(a) = \lambda p (\lambda a)^{p-1}$, $m = 0, n = 2$, 即为对数逻辑斯蒂 (LL) 模型, 据此形式所有成员都将进入初婚, 不存在终身未婚的情况。此外, 若形状参数 $p > 1$, 比例参数 λ 既决定尺度又决定位置, 灵活性较差。Brüderl 等 (1995) 针对这两个

缺陷各提出一种改进方法,并分别应用于拟合美国和德国的初婚数据。一种是广义对数逻辑斯蒂(GLL)模型,通过在风险函数中加入强度参数 b 使风险曲线上下移动,参数 λ 主要控制年龄轴上的位置,缺点在于仍不能容纳不婚的情况。另一种是对 LL 分布函数按式(2)增加一个终身曾婚比例系数 C , λ 仍同时影响曲线的尺度和位置^①。本研究将两种改进方法结合,即按式(2),将 GLL 分布密度函数乘以终身曾婚比例作为最终模型(仍称为 GLL 模型):

$$f(a)=C \frac{bp(\lambda a)^{p-1}}{[1+(\lambda a)^p]^{(1+\frac{b}{\lambda})}} \quad (4)$$

三、数据、模型估计方法和软件实现

(一) 拟合数据和处理

初婚年龄分布参数模型使用分年龄初婚率进行参数估计,实现对初婚经验分布的参数拟合。社会调查数据很难达到推断各出生队列年龄别初婚率所需要的样本量,本研究使用“六普”长表回顾性数据,重构各队列初婚年龄分布,在此基础上对各队列分别建模并追踪队列间变动。数据采取以下假设和处理:(1)为得到相对完整的女性初婚状况资料,对2010年全国人口普查长表5-5和5-4进行处理和拼接。将长表5-5调查时点年龄转化为出生队列进行队列分析,如无特别说明下文中队列均指出生队列。“六普”标准时点为2010年11月1日零时,普查时点30岁人口出生于1979年11月1日至1980年10月31日,该队列人口大部分出生于1980年,简便起见称为1980年出生队列,其他队列类同。根据长表5-4补充40~49岁初婚率。由于初婚年龄 a 岁是周岁年龄 $[a, a+1)$,初婚年份 y 代表 y 年1~12月,而本研究使用的出生队列 b 是指生于 $b-1$ 年11月至 b 年10月,为此,假设队列内各月人口规模均匀分布,且初婚在同一年队列不同月份子队列及单岁年龄的不同月份中均匀发生,调整长表5-4得到1946~1969年队列40~49岁初婚率。(2)将队列女性人数作为分母,以该队列女性在各年龄初婚人数为分子,计算出该队列女性的年龄别初婚率。各年龄累积初婚比例以初婚开始年龄至该年龄初婚率累积求和得到。(3)汇总数据及周岁年龄记法下,各队列在普查当年的年龄受到截尾影响,初婚风险暴露不足一年,使该年龄初婚率明显偏低,为避免引入误差,这一岁的数据不纳入分析。因此1980年队列只能使用29岁及之前的初婚率,其他队列采用类似处理。(4)考虑到15岁以下初婚率极低,在拟合时假设初婚从14岁开始,15岁以下初婚均发生在14岁(虽

^① 本研究用第二种方式改进的 LL 模型分析中国女性初婚数据时,发现对20世纪70年代后期出生队列拟合不佳,使用改进的 GLL 模型时,参数 b 的拟合值是 λ 的数倍,说明需要设置强度参数以把握初婚年龄分布集中(峰度高)的程度。

然《婚姻法》规定法定最低初婚年龄为 20 岁,但事实上各地存在结婚早于 20 岁的情况)。

(二) 参数估计方法

由于本研究使用普查汇总数据,适用区间删失数据的极大似然估计,即:

$$L(\boldsymbol{\theta}) = \frac{N!}{m_{a_0}! m_{a_0+1}! \cdots m_{x-1}! n_x!} \left\{ \prod_{a=a_0}^{x-1} [F(a+1; \boldsymbol{\theta}) - F(a; \boldsymbol{\theta})]^{m_a} \right\} [1 - F(x; \boldsymbol{\theta})]^{n_x} \quad (5)$$

其中, $\boldsymbol{\theta}$ 是需估计的参数向量,如 GLG 模型的 λ 、 b 、 u 和 C , m_a 使用 a 岁初婚率, n_x 使用 x 岁未婚比例^①,此时 $N=1$ (Kaneko, 2003)。

(三) 统计软件实现

本研究选用的 GLG 模型、扩散模型中的 Hermes 模型和改进的 GLL 模型并非常见统计软件的内置分布函数,并且 Coale-McNeil 模型与 GLG 模型分布函数包括上不完全伽马函数项(从大于零的变量到正无穷的积分)。本研究使用 JMP 14.1 软件,通过自定义损失函数和数值算法,对全国、分城乡女性 1946~1980 年出生队列初婚年龄、年龄别初婚率分别建模拟合和估算。

四、中国女性初婚年龄模式的参数拟合

(一) 中国女性初婚年龄的基本状况

1. 未婚比例的分布与变动

1990 年以来,全国 30 岁左右未婚比例增长较快,晚婚趋势明显(见表 1)。1990 年 25~29 岁男性未婚比例为 16.71%,2010 年为 36.29%;1990 年女性 25~29 岁未婚比例为 4.29%,2010 年达到 21.62%。2010 年城市及受教育程度较高的女性未婚比例更高,城市大专及以上学历女性 25~29 岁未婚比例超过 40%,30~34 岁高于 10%。

中国 50 岁左右人口未婚比例处于较低的水平。2010 年,全国男性 50 岁左右未婚比例高于 3%,在 20 年间有所降低;全国女性 45~54 岁未婚比例低于 0.5%,比 1990 和

表 1 1990、2000、2010 年中国两性未婚比例

%

| 年龄组 (岁) | 1990 年 | | 2000 年 | | 2010 年 | | | | | |
|------------|--------|------|--------|------|--------|-------|-------------|-------|-------|-------------|
| | 男 | 女 | 男 | 女 | 男 | | | 女 | | |
| | 全国 | 全国 | 全国 | 全国 | 全国 | 乡村 | 乡村小学 及以下 | 全国 | 城市 | 城市大专 及以上 |
| 25~29 | 16.71 | 4.29 | 24.68 | 8.67 | 36.29 | 31.11 | 38.23 | 21.62 | 29.13 | 40.94 |
| 30~34 | 7.16 | 0.64 | 7.45 | 1.35 | 12.62 | 13.67 | 23.89 | 5.35 | 7.35 | 10.74 |
| 45~49 | 5.07 | 0.18 | 3.96 | 0.21 | 3.12 | 4.29 | 10.20 | 0.44 | 0.82 | 1.30 |
| 50~54 | 4.48 | 0.17 | 4.05 | 0.19 | 3.21 | 4.56 | 8.64 | 0.30 | 0.57 | 1.18 |

注:根据“四普”、“五普”、“六普”数据计算。

① 这些比例的广义阶乘依常例使用伽马函数计算。

2000年略有提高。2010年50岁左右的城乡和不同受教育程度人口之间差异明显,乡村男性未婚比例更高,其中受教育程度为小学及以下的男性未婚比例为10%左右;女性中城市未婚比例高于镇和乡村,其中受教育程度为大专及以上的城市女性未婚比例高于1%。

2. 女性初婚年龄分布出现推迟和分散

由图1可见,初婚年龄的频率分布兼具“钟”形、峰度高、右偏的形态特点。较晚出生的队列初婚年龄出现分散和推迟,峰值附近初婚率明显下降,且随着众数年龄逐渐推后,分布偏度也在降低。1981~1985年队列不仅初婚率峰值大幅下降,峰值年龄之后下降的速度也更快,晚婚势头明显。

累积分布显示,1961~1965年和1966~1970年队列较为相似,结婚较早,24岁时约85%已经结婚,其后队列初婚不断推迟。1971~1975年队列24岁累积初婚比例降至79%以下,1976~1980年和1981~1985年队列进一步降至71%和64%。29岁累积初婚比例在1961~1965年及1966~1970年队列超过97%,1976~1980年队列已降至92%(见图1)。2015年1%人口抽样调查中1986年队列29岁的曾婚比例为86%。

3. 城乡与队列的初婚年龄模式分化

城乡分布是影响女性初婚年龄的重要因素。“六普”数据中,初婚数据来自回顾性问题,城乡分布取自调查时点,不能完全反映初婚时的情况,但考虑到调查时点的城乡分布仍然可以将总人口划分为不同的初婚群体,因此,本研究分城乡进行比较(见图2)。1961~1965年和1971~1975年队列,城市和镇女性比乡村初婚晚,即分布曲线形态类似,几乎依次向右平移1岁左右。1981~1985年队列,镇与乡仍然保持前述差别,城市女性不仅初婚年龄众数大幅提高,且分布更平缓,累积初婚比例远低于镇和乡。由于城乡女性初婚年龄模式不同,总人口的参数模型便成为城乡子人口的混合模型,为分析城乡差异,本研究将对城乡女性分别建模。

城乡人口中队列间初婚年龄的变动主要表现为较晚出生的队列分布曲线峰值不断

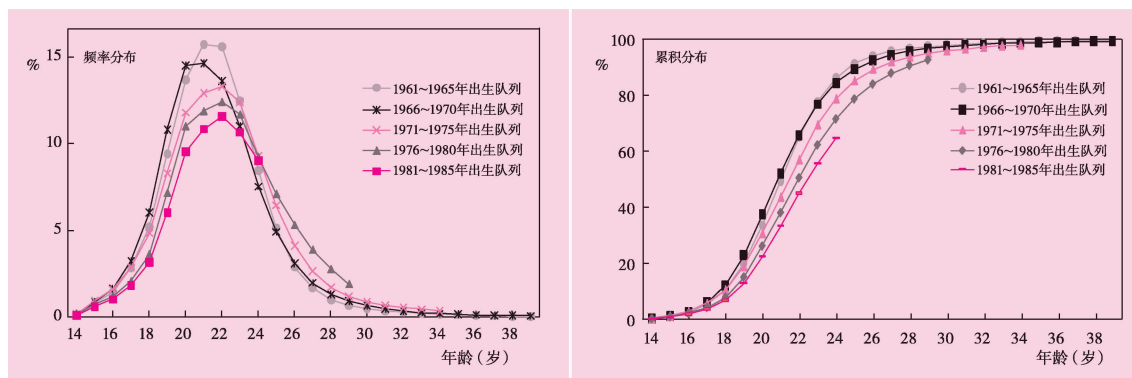


图1 中国女性各出生队列初婚年龄频率分布和累积分布

注:根据“六普”数据计算。

降低(见图3)。乡和镇不同队列间初婚众数年龄变化不大,而峰值不断下降,1961~1965年队列至1981~1985年队列下降约3.5%,同时,峰值两侧的变化更平缓。事实上初婚率峰值区域附近及之前初婚率降低,年龄均值仍会提高,即初婚在推迟。队列间的变化可能意味着理想初婚年龄仍然存在,但约束力变小。城市1981~1985年队列仍表现出不同的特点,不仅峰值大幅降低,且分布形态变化较大,众数年龄加速推迟,整体形态仍很难确定,可能会出现不同于以往的特征。

综合城乡与队列间的差异,城乡之间理想初婚年龄不同,1980年之前队列城市女性初婚年龄众数比镇和乡女性分别大1~2岁,1980年之后出生的队列城乡差异扩大;城乡子人口中,晚出生的队列理想初婚年龄的约束力变弱,初婚率峰值不断下降,导致初婚年龄分散和推迟。

(二) 参数模型拟合和外推检验

1. 拟合效果

本文使用3种参数模型对全国、乡、镇和市四类女性人口1946~1980年每一年出生队列的年龄别初婚率分别建模。总体上看,3个模型拟合效果均较好,都体现出初婚年龄分布“钟”形、峰度高、右偏的形态特征,累积初婚比例吻合程度较高(见图4、图5)。

从整体拟合效果看(见表2),负对数似然相近,表明3个模型拟合优度接近,但AIC和BIC准则在负对数似然基础上惩罚参数个数,Hernes模型因参数较少在简洁与准确之间更为平衡。从拟合效果图(见图4)看,GLG模型的年龄别误差分布与Kaneko(2003)拟合日本女性时的结果很相似,低估了峰值附近初婚率。Kaneko(2003)用较早队列估计误差的经验分布(调整的三次样条函数)对GLG模型进行调整,以提高拟合吻合度及预测准确性,但影响中国女性婚姻进程的法律、政策、社会经济环境变动不易把握,难以做出合理的调整假设。拟合效果较好的是改进的GLL模型,众数年龄附近比GLG模型接近,又不至于像Hernes模型,拟合形态受峰值点较强的影响。众数右侧准确度优于GLG函

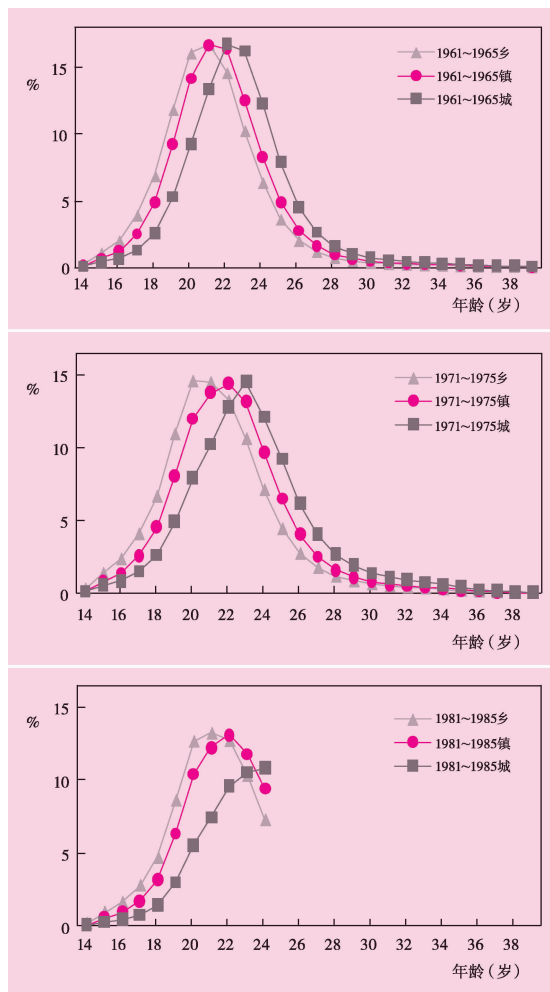


图2 各出生队列中城乡间初婚年龄分布
注:同图1。

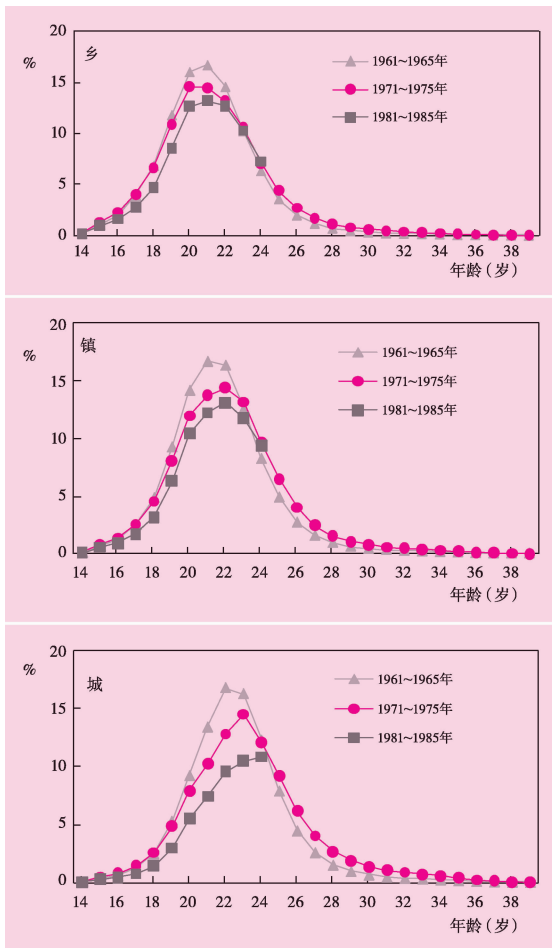


图3 城乡女性中不同出生队列初婚年龄分布

3种模型使用部分数据拟合和外推的年龄别初婚率与真实值相比仍较为接近(见图6),说明使用29岁之前的数据可以较好地捕捉年龄分布模式。然而,右侧误差图显

数,也由于参数个数更多,容易贴近原始分布的形态,右尾部的表现优于Hernes函数。

值得注意的是,只有两个参数的Hernes扩散模型拟合效果与GLL模型接近,优于4个参数的GLG模型。对初婚率峰值区域的估计优于GLG模型,对于初婚进程受外界政策环境严重影响的队列,也表现出良好的估计效果。这也许说明,不仅初婚行为在队列中扩散的进程符合Hernes模型的假定与特征,当出现强有力的外部干预时,影响被内化,在群体中蔓延的过程也能较好地非同质扩散模型刻画。

Hernes模型与GLG和GLL模型估计值在较晚出生队列右尾部差别较大(见图5),会影响终身未婚比例的预测结果,需要对预测误差进行检验和对比。

2. 部分数据外推检验

对基本完成初婚过程的队列,仅使用29岁及之前人为右截断的部分初婚率数据进行拟合,将外推结果与29岁之后真实值比较,检验3种模型的外推能力。

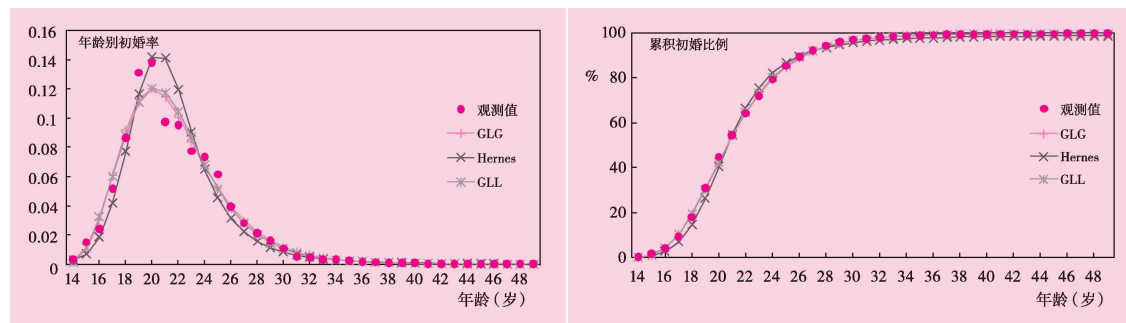


图4 全国1950年出生队列的拟合效果

注:1950年队列初婚率在20岁堆积,3种模型以不同方式进行了平滑,事实上参数模型拟合本身也是修匀的常用方法。

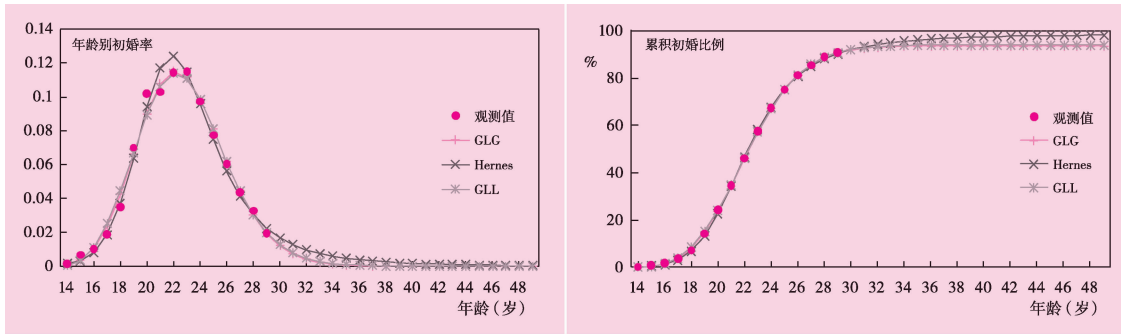


图 5 全国 1980 年出生队列的拟合效果

表 2 3 种参数模型拟合整体效果

| 参数模型 | 1950 年队列 | | | 1980 年队列 | | |
|--------|----------|---------|---------|----------|---------|---------|
| | 负对数似然 | AIC | BIC | 负对数似然 | AIC | BIC |
| GLG | 38.2062 | 84.4124 | 90.7464 | 18.0869 | 44.1737 | 47.2641 |
| Hernes | 38.2070 | 80.4141 | 83.5248 | 18.0778 | 40.1556 | 41.5717 |
| GLL | 38.2056 | 84.4112 | 90.7453 | 18.0890 | 44.1780 | 47.2684 |

注:这些准则评价的是模型整体拟合效果,鉴于 3 个模型机理、特点和适用情况各有不同,与用某种准则评判优劣相比,实际运用中更适合根据具体模型性质和研究目的判断和取舍。如本研究应用时更关注对分布尾部的拟合准确性和稳定性。

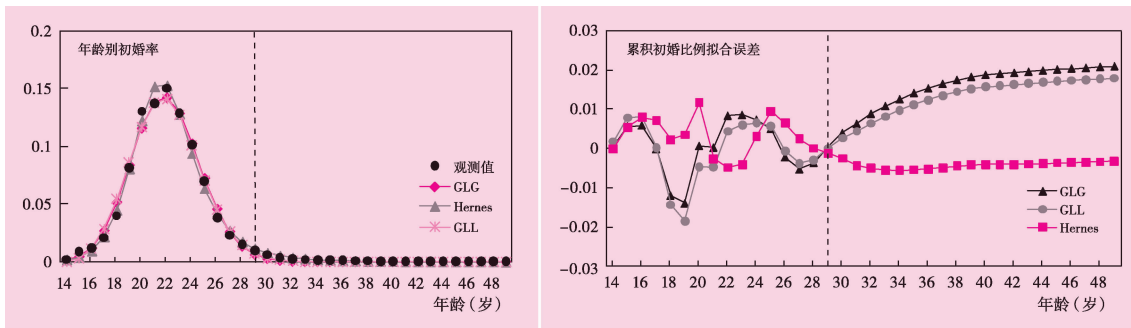


图 6 使用部分数据的拟合和外推效果(1960 年队列)

注:图中虚线均为 29 岁。在 29 岁及之前为拟合误差,29 岁后为外推误差。

示外推结果出现分化,GLG 和 GLL 模型会以约 2%的比例低估较高年龄的累积初婚比例(即高估未婚比例),GLL 略优于 GLG 模型,效果最好的是 Hernes 模型,误差接近零,倾向于低估未婚比例。

(三) 多队列参数变动分析

将多队列拟合参数汇总分析,可以综合初婚模式在队列间的变动特征,参数的变动应能反映历史重要事实且合乎直觉,从而能验证模型的应用效果并揭示初婚模式变化的机制。此外,本研究对城乡女性每一年出生队列分别建模拟合,各参数随队列推移的变

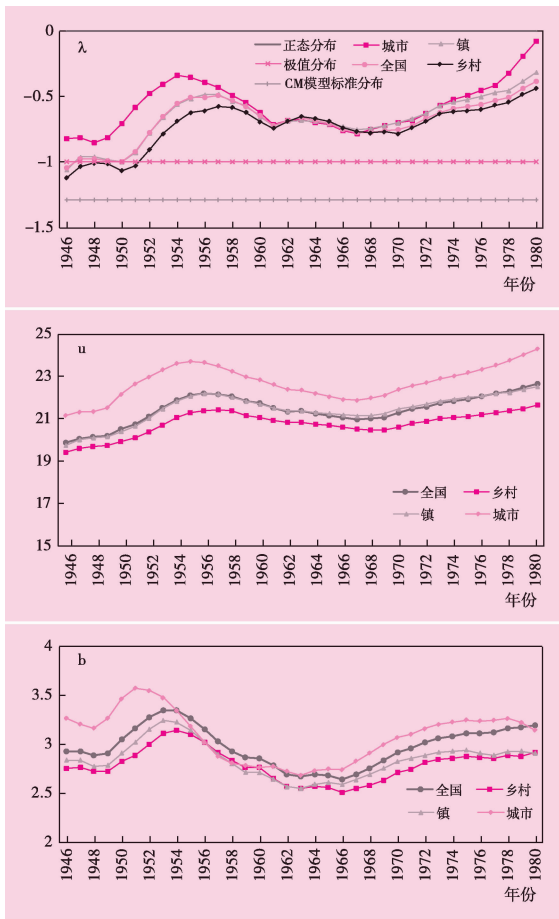


图7 GLG模型拟合中国女性1946~1980年队列初婚年龄分布参数变动

动均表现出连续性和规律性,说明模型拟合稳健、可靠。

计划生育政策和新《婚姻法》的实施曾对中国女性初婚年龄模式产生过较大影响,城乡女性初婚模式也有差别(见图7)。20世纪70年代计划生育政策下,很多地方规定女性登记结婚不得早于23岁,初婚出现推迟(Liang, 2000)。随着新《婚姻法》的实施,女性法定最低初婚年龄定为20岁,1956年之后队列初婚年龄开始缓慢下降(郭志刚、段成荣, 1999)。不同年份出生的队列在事件发生时年龄不同,初婚受到了不同的影响,参数所反映的初婚模式特征应有所区别。

GLG模型除终身曾婚比例之外的3个参数对应分布形态位置(众数年龄 u)、对称性(偏度,形状参数 λ)和集中性(尺度参数 b),清晰地分解了分布特征。除了1946~1952年队列接近极值分布,中国城乡女性各队列 λ 估计值始终高于Coale-McNeil标准分布(见图7),也逐渐高于极值分布,表现出较强的对称性。这可能源于中国女性理想初婚年龄的规范较强,不仅限制早婚,也包括对晚婚的限制,使女性集中在初婚“黄金年龄”前后1~2岁结婚,分布右侧初婚率快速降低,陡峭程度接近初婚过程起步时的左侧曲线,使分布峰度高、偏度低。

Hernes扩散模型参数 A 和参数 b 将初婚能力变动分解为同质和异质两部分。值得注意的是,1975年队列之后农村女性参数 b 下降更快,意味着群体异质性更强,且趋势变动急剧,未来农村未婚女性可能会进一步分化。

改进的GLL模型拟合城市女性1978~1980年队列时参数大幅偏离正常值域,可能是由于参数个数较多,数据点过少,且城市这些队列年龄分布平坦、截尾处未婚比例高、曲线特征不易把握,而出现过度拟合,因此GLL模型的分析只包括1946~1977年队列。在多个历史阶段和城乡比较中,参数 b 和 p 的作用方向相反,互相抵消,不便于解释。尽管GLL模型从拟合效果来说在峰值区域和右侧尾部精度最高,但以上缺陷的存在,使

其更适用于固定终身未婚比例推断初婚年龄分布或修匀初婚年龄堆积。

对队列间参数变动的分析表明,受历史事件影响的队列范围、参数变动特点和含义符合预期,说明初婚年龄参数模型不仅在社会环境变化中适用,而且能够动态反映出历史事实的影响。Hernes 模型的参数变动有助于透视初婚模式变动的原因。

五、年龄别累积初婚比例和终身未婚比例的估计及检验

对尚未完成初婚进程的队列,使用全部观测数据进行参数模型拟合,并将模型外推至未能观测的年龄,由于队列成员年龄增长与日历年份推移同步,因此各队列年龄外推结果也成为时期维度上部分年龄初婚水平的预测。为检验估计的准确性,先使用全部数据外推预测的年龄别累积初婚比例与 2015 年 1% 人口抽样调查资料进行比较。

(一) 估计结果与 2015 年 1% 人口抽样调查比较

2015 年 1% 人口抽样调查与“六普”标准时点一致,使两套数据的年龄与队列可比。相比城乡人口会发生乡城迁移,全国女性各队列可视为封闭人口。2015 年 1% 人口抽样调查只含截面的年龄别未婚比例数据,可计算出 49 岁及以下周岁年龄的曾婚比例,与本研究 1966~1980 年全国女性队列外推至 2015 年的累积初婚比例进行比较。1% 人口抽样调查周岁年龄曾婚比例在当前年龄暴露不足一年,本研究定义的累积初婚比例为下一生日之前的情况,前者相比后者偏低约当前年龄初婚率的一半,由于进行估算和比较的队列 2015 年时的年龄均在 35 岁以上,在本研究使用的全部经验数据和模型估计结果中,这一年龄段年初婚率的一半低至 0.24% 以下,差异有限。

GLG 和 GLL 模型外推结果出现一定程度的低估,其中晚出生、截断在较低年龄、累积初婚比例较低的队列低估更严重,Hernes 模型的预测与 1% 人口抽样调查较为接近(见图 8)。如能校正 2015 年 1% 人口抽样调查最高年龄暴露不足带来的差异,3 条曲线均会向下方移动,并且一般而言在左侧较低年龄下移会多于右侧较高年龄,这样图中 1% 人口抽样调查与 GLG 模型和 GLL 模型的差值会略为缩小。

(二) 终身未婚比例的预测

累积初婚比例的终身水平即终身曾婚比例,对 GLG 和 GLL 模型来说是式(2)的参数 C ,Hernes 模型则是已婚比例表达式中年龄趋于无穷时的极限值。终身未婚比例用 1 减去终身曾婚比例求得。

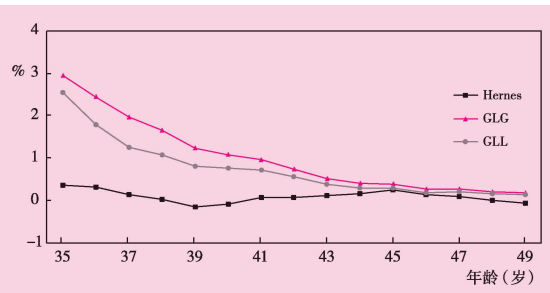


图 8 2015 年 1% 人口抽样调查全国年龄别曾婚比例与累积初婚比例模型预测结果的差值

注:由于“六普”与 1% 人口抽样调查间隔 5 年、以不同方式和比例抽样,因此图中差值除预测误差和 1% 人口抽样调查最高年龄不完全暴露带来的差异,也包含两次普查的调查总误差。根据“六普”和 2015 年 1% 人口抽样调查数据计算。

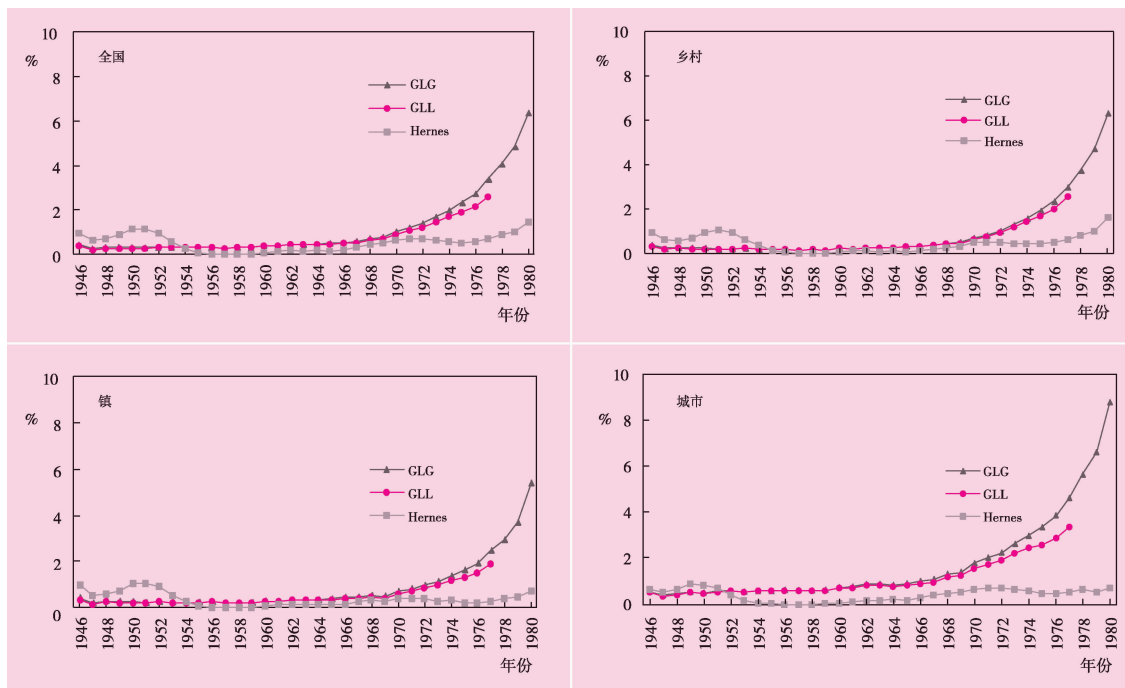


图9 中国城乡女性1946~1980年队列终身未婚比例的测算

3种模型对全国和城乡女性终身未婚比例的测算如图9所示。需要注意的是,城乡分布是“六普”调查时点的状况,测算中假设城乡分布保持不变。长期以来,中国女性终身未婚比例接近于零,但1970年之后队列的终身未婚比例开始升高,不同模型的测算结果差异较大。最高的是GLG模型,1980年队列终身未婚比例将达到6.39%,原因在于队列初婚众数年龄持续升高,表明初婚不断推迟;分布更对称,即众数年龄之后初婚可能性迅速下降;概率密度向下压缩,进入初婚的速度变慢。城市女性初婚推迟和分散更严重,因此终身未婚更多(见图7)。Hernes模型估计终身未婚比例虽然有所升高,但整体仍处于较低水平,1980年队列将增至1.48%,较低的增长幅度源于模型中同质性和异质性变动互相消长,初婚能力经历了由起步阶段无差别的下降到较晚队列随年龄放大的差异性下降,群体间异质性增强逐渐成为主导力量导致不婚的出现,且乡村女性初婚能力随年龄下降更快,因此预计将有更多乡村女性不婚。GLL模型估计结果处于二者之间,更接近GLG模型,预计1977年队列升至2.61%,原因是随着初婚年龄推迟,概率分布曲线向右下压缩,制约累积初婚比例的增长,带来终身未婚比例升高,且城市不婚更严重。虽然模型估计结果差异较大,但在较晚队列均已呈现明显的变动趋势。

六、结论与讨论

本研究使用初婚年龄别参数模型刻画各年出生队列初婚进程特征及变动,进而外

推预测终身未婚比例的变化,可归纳为两个层次的分解:(1)将截面的变动趋势分解为队列间变动。时期指标混杂了各个队列在不同年龄阶段的初婚状况,队列分析更能把握内在的变动机制。本研究分析结果表明,目前观察到的初婚推迟和初婚率下降,实质上是不同队列初婚年龄模式改变的队列效应,参数模型刻画了中国女性队列初婚模式的特点、城乡差异和变动。累积初婚比例随年龄的增长受到队列初婚模式的制约,从而使较晚出生队列的初婚延迟有转化为不婚的潜能。(2)将各队列初婚推迟进一步分解为晚婚和不婚。对参数模型拟合出的进度模式进行年龄外推,实现对未来初婚和终身未婚情况的估计,检验中部分和全部数据外推较为准确。综合3种模型的估计结果,中国女性普遍结婚的传统或将发生改变,在1970年之后出生的队列,终身未婚比例将升高,且有加剧的态势,全国1980年队列终身未婚比例预计为1.48%~6.39%。Hernes模型对不婚比例的估计结果最低,仅为1.48%,这主要是因为Hernes模型假设初婚扩散速度与已婚比例正相关,逐渐增大的社会压力会促使多数未婚女性最终进入初婚。GLG模型对不婚比例的预测结果为6.39%,反映在严格的理想初婚年龄限制下,年龄规范增加了大龄未婚女性进入初婚的阻力,适婚年龄之后初婚可能性快速下降,将出现较多不婚。本研究改进的GLL模型放松了Hernes模型中进入初婚的速度与已婚比例正相关的假定,众数右侧的初婚率处于Hernes模型和GLG模型之间,终身不婚比例的预测结果与GLG模型接近。1980年之后出生队列较高年龄累积初婚比例加速下降,转化为终身不婚的潜能更大,亟须进一步监测。

本研究结果表明,中国女性初婚年龄存在一定的进度模式,参数模型拟合和预测准确稳健,多队列的参数变化具有连续性和规律性,且变动符合预期,提示可采取一定假设,用时间序列方法对关键参数外推,以提高对更晚队列预测的把握性,如GLG模型参数 λ 和Hernes模型的参数 b 。然而,模型起作用的机制和理论意蕴仍有待挖掘,可从分析层次与应用范围两方面着手。受数据限制,本研究使用汇总数据进行了单性别的分析,分析层次有待拓展,与其他分析方法的结合使用尚显不足。事实上在已被验证的初婚模型及规律基础上结合使用新的统计分析方法与数据资料,可以提高认识和把握微观层面个体行动机制的能力,分析大样本微观数据时,由于GLG和GLL模型为常见风险模型,Hernes模型也是常见风险函数逻辑斯蒂模型的一个扩展,都便于容纳协变量,如GLG模型可对位置参数 u 加入协变量,GLL模型可对位置参数 λ 和强度参数 b 加入协变量等。此外,以模型刻画的群体初婚规律为指引,研究微观层面如何聚合形成宏观层面的已知规律,宏观和微观的对照与印证可能有助于透视初婚事件发生的机制。更多情境下的应用和检验有助于认识模型的作用原理和限制,如在当前中国同居和离婚增多、适婚年龄男多女少的情况下,初婚进度模式将如何变化、模型未来能否继续适用、对男性是否适用,都有待探索和总结。

参考文献:

1. 郭志刚(2017):《中国低生育进程的主要特征——2015年1%人口抽样调查结果的启示》,《中国人口科学》,第4期。
2. 郭志刚、段成荣(1999):《北京市人口平均初婚年龄的研究》,《南京人口管理干部学院学报》,第2期。
3. 黄荣清、亓昕(1990):《中国的初婚初育模型研究》,《中国人口科学》,第4期。
4. 李建新、王小龙(2017):《人口生育政策变迁与初婚风险——基于CFPS2010年调查数据》,《人口学刊》,第2期。
5. 王鹏、吴愈晓(2013):《初婚年龄的影响因素分析——基于CGSS2006的研究》,《社会》,第3期。
6. 於嘉、谢宇(2013):《社会变迁与初婚影响因素的变化》,《社会学研究》,第4期。
7. Bloom D.E., Bennett N.G.(1990), Modeling American Marriage Patterns. *Journal of the American Statistical Association*. 85(412):1009-1017.
8. Brüderl J., Diekmann A.(1995), The Log-Logistic Rate Model: Two Generalizations with an Application to Demographic Data. *Sociological Methods & Research*. 24(2):158-186.
9. Burch T.K.(2018), Theory, Computers and the Parameterization of Demographic Behavior. In *Model-Based Demography: Essays on Integrating Data, Technique and Theory*. Springer.
10. Coale A.J.(1971), Age Patterns of Marriage. *Population Studies*. 25(2):193-214.
11. Coale A.J., McNeil D.R.(1972), The Distribution by Age of the Frequency of First Marriage in a Female Cohort. *Journal of the American Statistical Association*. 67(340):743-749.
12. Coale A.J., Trussell J.(1996), The Development and Use of Demographic Models. *Population Studies*. 50(3):469-484.
13. Diekmann A.(1989), Diffusion and Survival Models for the Process of Entry into Marriage. *Journal of Mathematical Sociology*. 14(1):31-44.
14. Goldstein J.R., Kenney C.T.(2001), Marriage Delayed or Marriage Forgone? New Cohort Forecasts of First Marriage for U.S. Women. *American Sociological Review*. 66(4):506-519.
15. Hernes G.(1972), The Process of Entry into First Marriage. *American Sociological Review*. 37(2):173-182.
16. Hastings D.W., Robinson J.G.(1973), A Re-Examination of Hernes's Model on the Process of Entry into First Marriage for United States Women, Cohorts 1891-1945. *American Sociological Review*. 38(1):138-142.
17. Jones G.(2018), What is Driving Marriage and Cohabitation in Low Fertility Countries?. In Poston D.(eds) *Low Fertility Regimes and Demographic and Societal Change*. Springer.
18. Jones G.W., Gubhaju B.(2009), Factors Influencing Changes in Mean Age at First Marriage and Proportions Never Marrying in the Low-Fertility Countries of East and Southeast Asia. *Asian Population Studies*. 5(3):237-265.
19. Kaneko R.(2003), Elaboration of the Coale-McNeil Nuptiality Model as the Generalized Log Gamma Distribution: A New Identity and Empirical Enhancements. *Demographic Research*. 9(10):223-262.
20. Liang Z.(2000), The Coale-McNeil Model: Theory, Generalisation and Application. University of Groningen. The Netherlands: Thela Thesis.
21. Rodriguez G., Trussell J.(1980), Maximum Likelihood Estimation of Coale's Model Nuptiality Schedule from Survey Data. *World Fertility Survey Technical Bulletin*. (7).

(责任编辑:李玉柱)