

文章编号: 1007-6735(2018)01-0040-06

DOI: 10.13255/j.cnki.jusst.2018.01.008

谨防数学方法预测常住人口出生率的偏误

罗国芬, 邢 青

(上海理工大学 管理学院, 上海 200093)

摘要: 对一项运用数学方法进行上海市 2050 年前常住人口出生率预测所得结果进行比较分析, 提出更优的预测方法. 通过线性回归和 10 种曲线回归模型, 使用低生育率阶段的数据, 结合人口学方法进行新的预测, 总结出可靠性更高的预测方法. 预测结果表明, 缩小数据范围, 并使用人口预测软件, 平均预测精度有较大提升, 预测准确性更高. 研究结果说明: 单纯依靠数学方法预测常住人口出生率存在较大局限, 不宜多加使用, 尤其不适合对人口出生率进行长期预测; 回归人口本身的发展规律, 依据人口学方法可有效提高预测精度.

关键词: 常住人口出生率; 人口预测; 幂函数回归模型; 上海市

中图分类号: O 213.9 **文献标志码:** A

Beware of Forecast Errors of the Permanent Population Birth Rate by Using Unapplicable Mathematical Methods

LUO Guofen, XING Qing

(Business School, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Better forecasting methods were proposed through examining the prediction results of the birth rate of permanent population in Shanghai by 2050. A method of higher reliability was presented by adopting the power function fitting and the corresponding forecasting method is of the least error by comparing with the population birth rate data obtained by the linear regression model and ten kinds of curve regression models. It is found that the average prediction accuracy can be improved by re-forecasting combined with reducing the data scope and using a population prediction software. It is shown that there is a significant limitation to the prediction of permanent population birth rate by using mathematical methods, and it is not suitable for many cases, especially for the prediction of long-term birth rate. In the light of the demography method, recurring to the rule of population development itself can improve the prediction accuracy greatly.

Keywords: permanent population birth rate; population forecast; power function regression model; Shanghai

收稿日期: 2017-06-11

基金项目: 上海理工大学人文社科培育基金项目(SK18YB19)

第一作者: 罗国芬(1975-), 男, 讲师. 研究方向: 人口学、城市社会学、公共政策. E-mail: dxflgf@126.com

人口出生率及出生人口数量的预测,是人口预测中最基本的内容之一,其重要性毋庸置疑.以往的研究表明,人口预测有数学方法(特指选用某种适当的直线或曲线函数,拟合区域人口过去的发展变化,建立回归模型,预测区域人口将来发展趋势的方法)、队列预测法、多区域矩阵法、社会经济模型法及其他预测方法等^[1].由于所使用的数据较少,且数据易于获取,人口预测的数学方法得到部分研究者的青睐.但如果使用不当的话,数学预测方法也可能使人口预测结果出现较大的偏误.本研究以2014年发表的《上海市人口结构变化预测研究》^[2]一文中的相关论述为例,比较分析该文预测的相关问题,探索更为准确的人口出生率预测方法.

1 数学方法预测上海市常住人口出生率案例

1.1 基础数据

上海市人口出生率的数据主要来源于上海市人口和计划生育委员会网站、上海市人口和计划生育年鉴^[3],以及上海市国民经济和社会发展统计公报中户籍人口出生率和常住人口出生率资料^[4].文献^[2]使用的上海市历年户籍人口出生率数据如表1所示.

表1 上海市历年人口出生率

Tab.1 Birth rate in the city of Shanghai

年份	出生率/‰	年份	出生率/‰	年份	出生率/‰	年份	出生率/‰
1952	39.1	1967	12.5	1982	18.5	1997	4.92
1953	43.1	1968	14.9	1983	15.4	1998	4.73
1954	52.7	1969	14.8	1984	13.7	1999	5.01
1955	42.3	1970	13.9	1985	12.7	2000	5.27
1956	39.5	1971	12.2	1986	14.5	2001	4.34
1957	46.0	1972	10.8	1987	15.3	2002	4.66
1958	37.9	1973	10.2	1988	13.2	2003	4.28
1959	31.6	1974	9.2	1989	12.5	2004	6.0
1960	27.7	1975	9.4	1990	10.25	2005	6.08
1961	22.4	1976	10.2	1991	7.8	2006	5.95
1962	26.1	1977	10.8	1992	7.3	2007	7.34
1963	30.3	1978	11.3	1993	6.5	2008	6.98
1964	20.6	1979	12.3	1994	5.9	2009	6.62
1965	17.0	1980	11.8	1995	5.47	2010	7.13
1966	14.5	1981	16.8	1996	5.21	2011	7.17

2.2 预测方法中对年度数据的处理有差错

按照文献^[2]所依据的基础数据和列出的研究结果,用SPSS统计软件对数据分析过程进行检验,

1.2 数学方法预测上海市常住人口出生率结果

文献^[2]使用双变量回归分析(自变量 X 为年份,因变量 Y 为出生率)的方式,依据表1数据,假定生育政策不进行调整的基础上,采用SPSS预测分析2013—2050年的上海市常住人口出生率^[2],该团队得出上海市常住人口出生率的预测公式为

$$Y = 100.458\ 097\ 773\ 243\ 5X^a$$

式中, $a = -0.675\ 222\ 788\ 553\ 966\ 1$.

据这个常住人口出生率的预测公式,他们也得到了上海市2013—2050年常住人口出生率预测值结果^[2].

2 数学方法预测上海市常住人口出生率存在的问题

2.1 基础数据中混淆户籍人口出生率与常住人口出生率概念

经查证文献^[2]所使用的数据仅为上海市户籍人口的出生率数据,但在我国超大城市中,其常住人口构成既包含户籍人口,又包含大量外来常住人口.而且,户籍常住人口的年龄结构与外来常住人口的年龄结构往往有较大差别,户籍人口中老龄化较严重,而外来常住人口中主要以15~49岁的青壮年人口为主(见下页表2),年龄结构的不同对出生率的影响较大.

发现其采用的是曲线回归方法中的幂函数回归模型,但提供的预测结果在年份上略有差错,即“上海市2013—2050年出生率预测结果表”中所呈现的应

是2012—2049年上海市常住人口出生率的预测数据。

表2 “六普”上海市户籍常住人口与
外省市常住人口年龄结构比较

Tab.2 Comparison of the age structures between the registered
resident permanent population in Shanghai and the
permenant population coming from other provinces
and cities according to the sixth national census

年龄结构	外省市常住人口		户籍常住人口	
	人口数/人	百分率/%	人口数/人	百分率/%
0~14岁	765 938	8.53	1 216 921	8.67
15~49岁	7 477 607	83.30	6 562 943	46.74
50岁以上	733 455	8.17	6 262 332	44.60
总计	8 977 000	100	14 042 196	100

2.3 近年实际数据检验表明数学方法预测误差大

由于文献[2]得出的上海市常住人口出生率的预测结果是2013—2050年(实际应为2012—2049年),可将预测值与对应年份人口出生率实际发生值进行比较,以评估预测误差大小。根据近年《上海市国民经济和社会发展统计公报》,将常住人口、户籍人口出生率数值与文献[2]的常住人口出生率预测值进行比较,结果如表3和表4所示。

表3 依照常住人口口径计算的出生率预测
误差及出生人数预测误差率

Tab.3 Forecast errors of the calculated birth
rate and the calculated number of births
according to the permanent population

年份	年中人口 数/万人 (a)	出生数/ 万人(b)	出生率 预测误 差/%(c)	预测误差 人数/万人 (d = ac)	出生人数预测 误差率/% (e = $\frac{d}{b} \times 100$)
2012	2 363.95	22.61	3.30	7.80	34.50
2013	2 397.79	19.62	1.99	4.77	24.32
2014	2 420.42	20.20	2.23	5.40	26.72
2015	2 420.48	18.19	1.46	3.53	19.43
2016	2 417.49	21.84	3.004	7.26	33.25

由表3可见,依照常住人口口径计算,文献[2]的出生率预测误差不小,最近5年实际出生率数据与预测数据之差为1.5‰~3.3‰。根据最近5年的年中人口数推算,对应出生人数为3.5~7.8万人,其出生人数预测误差率最低为19.43%,最高达到34.50%。

由表4可见,依照户籍人口口径计算,文献[2]的出生率预测误差也不小,最近5年实际户籍人口出生率数据与预测数据之差为1.2‰~3‰,预测误

差所对应出生人数为1.70~4.32万人,其出生人数预测误差率最低为16.27%,最高达33.40%。

表4 依照户籍人口口径计算的出生率预测
误差及出生人数预测误差率

Tab.4 Forecast errors of the calculated birth
rate and the calculated number of births
according to the registered population

年份	年中人口 数/万人 (a)	出生数/ 万人(b)	出生率 预测误 差/%(c)	预测误差 人数/万人 (d = ac)	出生人数预测 误差率/% ($\frac{d}{b} \times 100$)
2012	1 423.15	12.11	2.25	3.20	26.44
2013	1 426.04	10.52	1.20	1.71	16.27
2014	1 427.20	11.90	2.22	3.17	26.63
2015	1 431.44	10.38	1.19	1.70	16.41
2016	1 436.56	12.92	3.004	4.32	33.40

由表3和表4可以看出,不管使用哪个口径计算,实际值与预测值均存在数个千分比的偏差,户籍人口和常住人口的出生率预测偏差均较大,其中常住人口出生率预测偏差更明显,远超一般预测可以容忍的10%。显然,这样的数学方法对人口出生率的预测是不太科学的。

2.4 其他回归方法预测结果比较

那么,是否采用其他回归方法预测效果会更好一些呢?本文采用10种曲线回归模型及线性模型进行分析,并对模型结果进行检验。表5中各回归方法下的预测误差指采用某种回归方法进行拟合,根据回归方程预测某年的出生率数据与实际该年的人口出生率数据进行比较的差值。平均误差是指根据回归方程预测的5年数据与迄今为止5年的实际人口出生率数据进行比较,计算5年中平均每年预测数据的误差值。由于回归方程预测的结果可能高于,也可能低于实际人口出生率,在计算平均误差时先根据年度预测误差值取绝对值后再求和、求平均值。

由表5可以看出,采用数学方法,预测误差最大为线性模型方法,曲线回归模型方法预测误差均比线性模型方法预测误差要小。而在曲线回归方法中,使用幂函数回归模型预测得到的5年平均预测误差最小,只有1.97‰。

3 人口出生率预测方法的改进

3.1 使用较近期时段数据进行回归方法预测所得 结果精度会有所提高

我国的人口发展经历过从高出生率到低出生率

的人口转型^[5].以 20 世纪 70 年代为界,上海也实现了高出生率到低出生率的人口转换.如果不考虑这种人口转型而贸然使用转型前后的全部数据进行回归分析,自然会造成前述预测误差较大的情况.为此,本文分别使用 1978—2011 年和 2000—2011 年

的户籍人口出生率数据,再同样使用 10 种曲线回归方法和线性模型来预测上海市人口出生率变化趋势.预测结果与 2012—2016 年实际的上海市人口出生率数据进行比较,测算不同回归方法拟合的预测误差,结果如表 6 和下页表 7 所示.

表 5 10 种曲线回归预测方法的预测误差及线性模型预测误差

Tab.5 Prediction errors of the ten curve regression prediction methods and the linear model

‰

年份	实际值	线性模型	二次项	复合	增长函数	对数
2012	8.51	-10.21	1.45	-4.36	-4.36	-4.94
2013	7.39	-9.65	3.16	-3.38	-3.38	-4.01
2014	8.34	-11.16	2.83	-4.47	-4.47	-5.16
2015	7.25	-10.63	4.58	-3.51	-3.51	-4.26
2016	9.00	-12.94	3.53	-5.39	-5.39	-6.20
平均误差	—	10.91	3.11	4.22	4.22	4.91
年份	立方模型	S 模型	指数分布	逆模型	逻辑斯蒂	幂函数
2012	-5.93	1.64	-4.36	3.45	-4.36	-2.25
2013	-5.67	2.75	-3.38	4.55	-3.38	-1.20
2014	-7.57	1.79	-4.47	3.59	-4.47	-2.22
2015	-7.51	2.88	-3.51	4.66	-3.51	-1.19
2016	-10.37	1.12	-5.39	2.90	-5.39	-3.00
平均误差	7.41	2.04	4.22	3.83	4.22	1.97

表 6 基于 1978—2011 年出生率数据的曲线回归预测、线性模型预测误差

Tab.6 Prediction errors of the curve regression method and the linear model based on the birth rate data from 1978 to 2011 ‰

年份	实际值	线性模型	二次项	复合	增长函数	对数
2012	8.51	-5.24	-5.24	-4.08	-4.08	-2.93
2013	7.39	-4.44	-4.44	-3.11	-3.11	-1.90
2014	8.34	-5.71	-5.71	-4.21	-4.21	-2.95
2015	7.25	-4.95	-4.95	-3.26	-3.26	-1.95
2016	9.00	-7.02	-7.02	-5.14	-5.14	-3.79
平均误差	—	5.47	5.47	3.96	3.96	2.71
年份	立方模型	S 模型	指数分布	逆模型	逻辑思迪	幂函数
2012	2.83	-1.22	-4.08	-0.47	-4.08	-2.89
2013	5.85	-0.10	-3.11	0.65	-3.11	-1.83
2014	7.07	-1.06	-4.21	-0.31	-4.21	-2.84
2015	10.59	0.03	-3.26	0.77	-3.26	-1.81
2016	11.57	-1.73	-5.14	-0.98	-5.14	-3.61
平均误差	7.58	0.83	3.96	0.64	3.96	2.60

由表 6 可见,基于 1978—2011 年的出生率数据,逆模型回归方法及 S 模型预测,每年预测误差仅为 0.64‰~0.83‰.而表 7 中,基于 2000—2011 年的出生率数据,7 种曲线回归模型预测方法每年预测误差均小于 1‰.考虑到上海曾经历人口转换过程,将预测基础数据范围适当调整后,拟合方程预测

的结果与近年实际值之间的差值均有所缩小.

可见,总体来看,依据几乎全部可获数据年份、时间跨度最长的数据所作的预测,其结果与近年实际值比较,平均误差值最大;而依据近期年份、时间跨度较小的数据所作的预测分析,其平均误差值反而最小.

表 8 中(见下页), c_1 是基于 1978—2011 年数

据,采用拟模型方法进行拟合得到的预测的结果与实际年份户籍人口出生率之差. d_1 是据此计算的预测出生人数与当年实际出生人数之差, e_1 是该方法下出生人数预测误差率.同理, c_2 是基于 2000—

2011 年数据,采用二次项模型方法进行拟合得到的预测的结果与实际年份户籍人口出生率之差, d_2 是据此计算的预测出生人数与当年实际出生人数之差, e_2 是该方法下出生人数预测误差率.

表 7 基于 2000—2011 年出生率数据的曲线回归预测、线性模型预测误差

Tab.7 Prediction errors of the curve regression method and the linear model based on the birth rate data from 2000 to 2011

年份	实际值	线性模型	二次项	复合	增长函数	对数
2012	8.51	-0.77	-0.77	-0.54	-0.54	-1.48
2013	7.39	0.62	0.62	0.96	0.96	-0.27
2014	8.34	-0.06	-0.06	0.40	0.40	-1.14
2015	7.25	1.30	1.30	1.91	1.91	0.02
2016	9.00	-0.18	-0.18	0.60	0.60	-1.65
平均误差	—	0.59	0.59	0.88	0.88	0.91

年份	立方模型	S 模型	指数分布	逆模型	逻辑思迪	幂函数
2012	-2.30	-2.17	-0.54	-2.09	-0.54	-1.46
2013	-2.25	-1.03	0.96	-0.95	0.96	-0.23
2014	-4.75	-1.97	0.40	-1.89	0.40	-1.08
2015	-5.73	-0.87	1.91	-0.79	1.91	0.10
2016	-10.15	-2.61	0.60	-2.53	0.60	-1.56
平均误差	5.04	1.73	0.88	1.65	0.88	0.89

表 8 基于 1978—2011 年数据与 2000—2011 年数据的两种出生率预测误差及出生人数预测误差率

Tab.8 Forecast errors of the calculated birth rate and the calculated number of births according to the registered population based on the data from 1978 to 2011 and from 2000 to 2011

年份	年中人口数/ 万人(a)	实际数/ 万人(b)	出生率预测 误差/ $\%$ (c)		预测误差人数/万人 ($d = ac/1000$)		出生人数预测误差率/ $\%$ ($e = \frac{d}{b} \times 100$)	
			c_1	c_2	d_1	d_2	e_1	e_2
2012	1423.15	12.11	-0.47	-0.77	0.67	1.10	5.52	9.05
2013	1426.04	10.52	0.65	0.62	0.93	0.88	8.81	8.40
2014	1427.20	11.90	-0.31	-0.06	0.44	0.09	3.72	0.72
2015	1431.44	10.38	0.77	1.30	1.10	1.86	10.62	17.93
2016	1436.56	12.92	-0.98	-0.18	1.41	0.26	10.90	2.00

由表 8 可见,依据新的拟合与预测,其初始预测结果与最近 5 年实际的人口出生率较为接近,其预测出生人数误差最小的仅为 900 人左右,最多也没有超过 2 万人.最近 5 年预测出生人数与实际相比,预测误差率绝大部分在 11% 以下,甚至有 2014 年的预测出生人数与实际相比,预测误差率在 5% 以内.不过,总体来看,这两种方法的预测数据与实际数据相比时高时低,不太稳定,用来作为人口规划的基础,仍有一些不足.

3.2 使用人口统计软件进行预测的结果精度更高一些

依据前文复盘检验发现,对于人口出生率的预

测,其准确性不仅与回归模型的选择有关,也与所选择的数据有关.考虑人口出生率历史性转变的规律,只基于众所周知的政治时间(1978 年以来)或瞩目时间(2000 年跨入新世纪以来)选择基础数据的范围,预测精度比文献[2]的要更好一些.但这些方法依然不能确定选用哪一时间段的数据.采用哪一种回归模型进行数学拟合预测结果会更好.因此,有必要抛开数学方法预测人口出生率的做法^[2,6],更多回归到依托人口学本身的基本规律来进行相关预测.以往研究表明,人口的出生率主要受人口的性别、年龄分布等因素的影响^[7-10].出生率是生育率(一年中新生婴儿数同育龄妇女人数的比率)和育龄

妇女在总人口数中比重的乘积,后两个因素都直接影响着出生率的变动.因此,可以据此重新设定,开发出更具可靠性的预测人口出生率的研究方法.

中国人口与发展研究中心开发的国际化人口预测软件 PADIS-INT_v1.2.2.5,具有功能强大、技术先进、方便快捷、准确率高、可视化效果好、输入简单、输出结果丰富等特点.依据上海市 2010 年“六普”常住人口分性别年龄数据,再根据上海市 2010 年左右人口发展态势的基础上设定基本参数(其关键点就是对死亡水平、生育水平、迁移水平、性别比等参数的设定)后,进行初步预测,主要预测参数是:起始人口为 2010 年“六普”分性别年龄的上海市常住人口(共计 23 019 196 人),死亡水平设定为男性人均预期寿命由 80 岁渐增至 83 岁,女性由 84 岁渐增至 87 岁,模型生命表为联合国生命表(一般模型),生育水平由 1 左右渐增至 1.3 左右,生育模式为“六普”时户籍人口生育模式,出生性别比由 112 渐降至 108,迁移水平为先增后降渐趋平稳增长的态势,迁移模式以“六普”时迁移情况为准.在这样仍显粗略的条件设定下,进行 2011—2050 年的人口预测,其预测结果中包含人口粗出生率的预测数据.将此预测数据与 2011—2016 年已知的常住人口出生率实际数据进行比较,其 6 年平均预测误差值为 1.34%.本次预测中,对基期人口直接采用六普人口分性别年龄数据,未按 2010 年年终人口数据进行调整.另外,死亡生命表未依据上海市实际情况进行专门计算,而是套用联合国生命表(一般模型)作为模型生命表,且迁移模式的设定上未依年龄进行精确设计,这些都在一定程度上损害了预测的精确度.但即便如此,使用人口学软件获得的预测结果与前述纯数学方法的预测结果相比,平均误差还是要小很多,如表 9 所示.

表 9 上海市常住人口出生率实际值与预测值比较
Tab.9 Comparison between the actual value and the predicted value of the birth rate in Shanghai

年份	出生率实际值	利用 Padis-int 预测值	预测误差 ‰
2011	6.97	9.66	2.69
2012	9.56	9.63	0.07
2013	8.18	9.68	1.5
2014	8.35	9.66	1.31
2015	7.52	9.58	2.06
2016	9.00	9.40	0.40

4 结 论

人口出生率的影响因素较多^[11],具有较大的不确定性.本研究发现,单纯的数学回归预测难以科学把握现实人口出生率走向.对类似的拟合与预测结果,需要引起高度警惕.当然,虽然有些研究对上海市常住人口出生率的预测因误差较大而难有更大的参考价值,但这并不否认这些研究对上海市将长期处于 10‰以下的超低出生率起到了警示作用.实际上,考虑到人口因素,将基础数据范围缩短至人口出生率历史性转变后的两个时间段,采用纯数学方法,可以一定程度上改善预测精度较差的缺陷,但是,这种预测方法的普适性还没有经过更多的验证,目前只能存疑.然而,考虑人口的性别、年龄结构以及生育率等因素,用人口软件进行出生率的预测,可以获得预测精度更高、更稳健的结果.这也说明,更多地考虑人口发展的一些规律性因素,将使人口出生率的预测精度更高,学者对自身研究结果也会有更大的自信.

参考文献:

- [1] 王桂新.区域人口预测方法及应用[M].上海:华东师范大学出版社,2000.
- [2] 王裕明,吉祥,刘彩云.上海市人口结构变化预测研究[J].上海经济研究,2014(3):89-98.
- [3] 上海市统计局.上海统计年鉴[M].北京:中国统计出版社,2016.
- [4] 国家统计局上海调查总队.上海市国民经济和社会发展统计公报[R].上海:上海市统计局,2016.
- [5] 张本飞.中国人口出生率的转变与西方的比较[J].西北人口,2006,27(1):57-60.
- [6] 王成科.灰色数列预测模型在人口出生率研究中的应用[J].数理医药学杂志,1994,7(4):353-355.
- [7] 袁小平,梁海艳.中国人口年龄结构变动对出生率的影响研究[J].西北人口,2014,35(6):49-53.
- [8] 杨慧诗,孟晗.关于人口出生率影响因素的逐步回归分析[J].现代商业,2010(5):177-178.
- [9] 李松臣,张世英.基于逐步回归法的人口出生率影响因素分析[J].统计与决策,2008(4):7-9.
- [10] 王国军,赵小静,周新发.我国人口出生率影响因素实证研究[J].经济问题,2016(2):7-11.
- [11] 李建伟.我国人口出生率的影响因素及其发展趋势[J].发展研究,2014(9):71-78.

(编辑:丁红艺)