

# 中国人口总量预测模型新探

——与赵进文教授商榷

范柏乃 刘超英

**【摘要】** 文章提出了赵进文教授构造的中国人口总量与 GDP 总量关系模型的两个缺陷,并依据 1952~2002 年年度中国人口总量的统计数据,采用逐步回归分析法,构建了预测中国人口总量的自回归模型。模型的参数检验结果表明,自回归模型不仅弥补了赵进文教授模型的缺陷,而且对中国人口总量也有较强的预测能力。

**【关键词】** 中国人口总量 简单回归模型 自回归模型

**【作者】** 范柏乃 浙江大学经济学院公共管理系,副教授;刘超英 浙江大学国家科技园,助理研究员。

## 一、引言

近年来,围绕中国人口增长与经济的关系问题,国内学者已经进行了大量的实际研究,并建立了很多人口预测模型。例如,南开大学李建民等在深入考察不同时期、不同国家,人口增长与经济增长之间表现出不同的相关关系后,通过模拟方法研究了生育率下降及控制人口增长对中国经济增长和人民生活水平提高的贡献,并在特定的假设条件下构建了一个动态的人口——经济运行模型。结果显示,人口模拟结果与实际数据有很大的偏差,说明采用该模型来预测中国人口总量难以获得可靠准确的结果。近期,赵进文教授(2003)采用经济计量学中无约束混合有限多项分布滞后模型,依据 1952~2000 年的 GDP、总人口、出生率与死亡率数据,构造了能反映总人口与 GDP 总量之间内在关系的综合模型(简称综合模型): $y_t = \alpha_0 + \alpha_1 x_t + \alpha_2 x_{t-1} + \alpha_3 x_{t-3} + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + \mu \eta_t + \gamma \rho_t + \varepsilon_t$ 。其中, $y_t$  为第  $t$  年的总人口数, $x_t$  为第  $t$  年的 GDP 值, $\eta_t$  为第  $t$  年的人口出生率, $\rho_t$  为第  $t$  年的人口死亡率,其他为相应变量的不同滞后,并且  $\beta_i = \lambda_0 + \lambda_1 i + \lambda_2 i^2 + \lambda_3 i^3, i = 1, 2, 3$ , 该综合模型预测的 2001 年中国人口总数与《中国统计年鉴(2002)》中公布的数据仅相差 46.3367 万人。

该综合模型对实际数据拟合的效果确实很好,预测精度很高,但仔细推究发现有以下两个明显缺陷:(1)一个优良的回归模型,所有的回归系数都应该通过显著性检验。而在综合模型中只有  $\mu$  和  $\gamma$  两个回归系数通过显著性检验,说明中国人口总量即被解释变量( $y_t$ ) 主要是由出生率( $\eta_t$ ) 和死亡率( $\rho_t$ ) 两个解释变量决定的,其他解释变量对被解释变量的影响在统计上是不显著的。我们依据同样的数据采用逐步回归分析结果表明,综合模型中的 8 个解释变量只有出生率和死亡率两个解释变量进入回归方程,其他解释变量均未能进入回归方程。(2)综合模型中包含了出生率和死亡率两个解释变量,而实际上获得这两个解释变量的实际数据比获得人口总量即被解释变量的数据更为艰难。假如能够获得出生率和死亡率两个解释变量的实际数据,只要采用简单的数学函数公式  $y_t = y_{t-1}(\eta_t - \rho_t)$ ,

就能精确地得到中国总人口的实时数据,没有必要借助于预测模型。事实上,国家统计局就是通过人口变动的抽样调查获得出生率和死亡率的实际数据,由此推测某一年度中国人口的总量。

依据 1952~2002 年的数据计算 Pearson 相关系数显示,GDP 与中国人口总量的相关系数为 0.776,而年份与中国人口总量的相关系数则高达 0.998。很显然,年份与人口总量比 GDP 总量与人口总量有着更为紧密的依存关系。从图 1 和图 2 可以看出,与 GDP 相比,年份与中国人口总量之间的线性关系也更为清晰。因此,从理论上讲可以建立人口总量(被解释变量)对年份(解释变量)的回归模型。

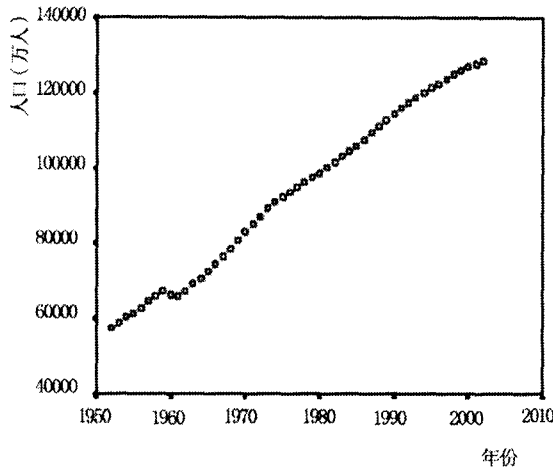


图 1 年份与人口之间依存关系的散点图

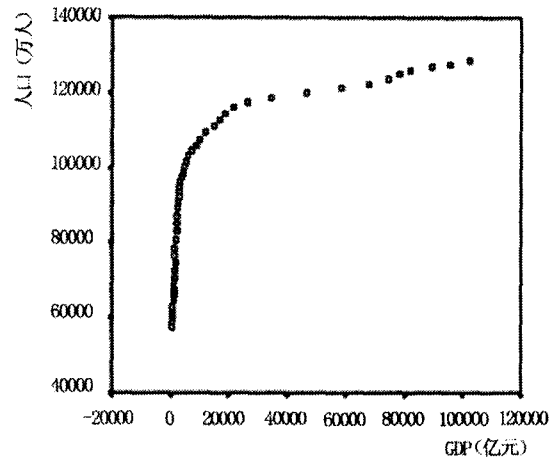


图 2 GDP 与人口之间依存关系的散点图

## 二、中国人口总量预测的简单回归模型

以 SPSS for windows 11.5 为工具,依据 1952~2002 年国家统计局公布的人口总量数据,以人口总量 ( $y_t$ ) 为被解释变量和年份 ( $x_t$ ) 为解释变量进行简单线性回归分析。表 1 给出了简单回归模型的主要评价指标,表 2 给出了回归方程的显著性检验结果,表 3 是关于回归系数的显著性检验结果。

表 1 简单线性回归模型的主要评价指标

| 评价指标 | 简单相关系数 | 复相关系数 | 修正后复相关系数 | 估计标准误差    | DW 统计量 |
|------|--------|-------|----------|-----------|--------|
| 数值   | 0.997  | 0.995 | 0.995    | 1618.4879 | 0.123  |

表 2 简单线性回归模型的方差分析表

| 方差来源 | 平方和             | 自由度 | 均方              | F        | 显著性水平 |
|------|-----------------|-----|-----------------|----------|-------|
| 回归   | 25242033471.948 | 1   | 25242033471.948 | 9636.191 | 0.000 |
| 误差   | 128355653.229   | 49  | 2619503.127     |          |       |
| 总和   | 25370389125.176 | 50  |                 |          |       |

表 3 简单线性回归模型的回归系数分析

|     | 非标准化的回归系数    |            | 标准化的回归系数 | t       | 显著性水平 |
|-----|--------------|------------|----------|---------|-------|
|     | B            | Std. Error | Beta     |         |       |
| 常数项 | -2894497.128 | 30440.167  |          | -95.088 | 0.000 |
| 年份  | 1511.406     | 15.397     | 0.997    | 98.164  | 0.000 |

表 1 显示,年份与人口总量的复相关系数高达 0.995,回归方程的变差解释能力达到了 99.5%。表 2 显示,回归方程的方差分析结果 F 统计值为 9 636.191,达到了高度显著性水平,表明回归模型的拟合效果是比较理想的。表 3 显示,两个回归系数 t 统计值达到了足够大,检验结果呈现高度显著性,说明回归模型中的回归系数显著地不为

零,解释变量对被解释变量的影响是显著的、不可忽略的。由此,可以得到人口总量对年份的回归模型: $y_t = -2894497.128 + 1511.406x_t$ 。

表5给出了1990~2002年间的人口实际总量(国家统计局公布的官方数字)、人口预测总量(用该模型获得的预测值)、人口预测偏差(国家统计局公布的数字与人口预测总量之间的偏差)及官方数字偏差(国家统计局公布的官方数字与户籍管理部门公布的官方数字之间的偏差)。从总体上来看,运用上述简单回归模型预测中国人口总量产生的偏差小于国家统计局公布的官方数字与户籍管理部门公布的官方数字之间的偏差。但简单回归模型的一个重要缺陷是DW统

计值仅为0.123,说明模型的残差存在着严重的序列自相关问题,在一定程度上影响了模型预测结果的可靠性和可信性。因此,尽管简单回归模型对人口实际总量拟合的效果是比较理想的,但也不能运用该模型来预测中国人口的总量。

### 三、中国人口总量预测的自回归模型

为了进一步提高人口总量预测的精度,并解决序列自相关问题对模型的干扰,在回归模型中引入内生变量构造自回归模型。自回归模型的一般表达形式为: $y_t = a + \sum_{i=1}^m b_i y_{t-i} + \sum_{j=0}^n c_j x_{t-j} + u_t$ 。式中, $y_t$ 为第 $t$ 年的人口总量, $y_{t-1}$ 为 $y_t$ 的滞后值, $x_t$ 为第 $t$ 年的年份, $x_{t-j}$ 为 $x_t$ 的滞后值, $u_t$ 为预测残差。因为被解释变量的滞后值作为解释变量出现在模型的右边,所以这样的模型被称为自回归模型。本文采用的自回归模型为: $y_t = a_0 + a_1 x_t + b_1 y_{t-1} + b_2 y_{t-2} + b_3 y_{t-3} + u_t$ 。

以1952~2002年年度的数据为样本,为解决回归模型中解释变量的共线性问题,采用逐步回归分析的方法求解上述自回归模型中的回归系数。逐步回归分析的结果如表5、表6和表7所示。

表5给出了自回归分析的主要评价指标。结果显示,复相关系数和修正后复相关系数都接近于1(由于复相关系数小数位数很多,超出了SPSS软件输出结果的显示范围,其结果显示为1),说明构造的自回归模型中的解释变量的方差解释能力达到了很高的水平,回归方程对实际数据拟合的效果是相当理想的。DW统计值也由简单回归模型的0.123提升到1.825,接近于2,说明自回归模型已经圆满解决了残差项的序列自相关问题。

表6给出了自回归模型的显著性检验结果。结果显示,F统计值为39427.10,达到了高度显著性水平,即自回归模型的总体回归系数显著不为零,说明解释变量对被解释变量的影响是显著的。

表4 使用简单回归模型对中国人口总量的预测结果 万人

| 年份   | 人口实际总量   | 人口预测总量   | 人口预测偏差   | 户籍部门公布的数字 | 官方数字偏差  |
|------|----------|----------|----------|-----------|---------|
| 1990 | 114333.0 | 113200.0 | 1132.962 | 113274    | 1059.00 |
| 1991 | 115823.0 | 114711.4 | 1111.557 | 114511    | 1312.00 |
| 1992 | 117171.0 | 116222.8 | 948.1511 | 115563    | 1608.00 |
| 1993 | 118517.0 | 117734.3 | 782.7455 | 116597    | 1920.00 |
| 1994 | 119850.0 | 119245.7 | 604.3399 | 117674    | 2176.00 |
| 1995 | 121121.0 | 120757.1 | 363.9343 | 118468    | 2653.00 |
| 1996 | 122389.0 | 122268.5 | 120.5287 | 119546    | 2843.00 |
| 1997 | 123626.0 | 123779.9 | -153.877 | 120583    | 3043.00 |
| 1998 | 124761.0 | 125291.3 | -530.283 | 121498    | 3263.00 |
| 1999 | 125786.0 | 126802.7 | -1016.69 | 122492    | 3294.00 |
| 2000 | 126743.0 | 128314.1 | -1571.09 | 123672    | 3071.00 |
| 2001 | 127627.0 | 129825.5 | -2198.50 | 124430    | 3197.00 |
| 2002 | 128453.0 | 131336.9 | -2883.90 |           |         |

资料来源:《中国人口统计年鉴》(1990~2002),中国统计出版社,相应年份;《中华人民共和国全国分县市人口统计资料》(1990~2002),群众出版社,相应年份。

表5 自回归模型的主要评价指标

| 评价指标 | 简单相关系数 | 复相关系数 | 修正后复相关系数 | 估计标准误差   | DW统计量 |
|------|--------|-------|----------|----------|-------|
| 数值   | 1.000  | 1.000 | 1.000    | 426.5396 | 1.825 |

表7给出自回归模型的回归系数检验结果。结果显示,在原先设定的  $x_t, y_{t-1}, y_{t-2}, y_{t-3}$  4个解释变量当中,经逐步回归最终前3个变量进入回归模型。进入回归模型的3个解释变量和常数项的回归系数都通过了显著性检验,即回归系数显著不为零,说明3个解释变量对被解释变量的影响是显著的。

通过上述分析,可以得到中国人口总量预测的自回归模型为:  $y_t = -278824 + 145.742x_t + 1.61y_{t-1} - 0.707y_{t-2}$ 。

表6 自回归模型的方差分析

| 方差来源 | 平方和             | 自由度 | 均方             | F         | 显著性检验 |
|------|-----------------|-----|----------------|-----------|-------|
| 回归   | 21519634833.912 | 3   | 7173211611.304 | 39427.100 | 0.000 |
| 误差   | 8005187.088     | 44  | 181936.070     |           |       |
| 总和   | 21527640021.000 | 47  |                |           |       |

表7 自回归模型的回归系数显著性检验

|            | 非标准化的回归系数   |            | 标准化的回归系数 | t      | 显著性检验 |
|------------|-------------|------------|----------|--------|-------|
|            | B           | Std. Error | Beta     |        |       |
| 常数         | -278824.365 | 121407.384 |          | -2.297 | 0.026 |
| $y_{t-1}$  | 1.610       | 0.111      | 1.614    | 14.480 | 0.000 |
| $y_{t-2}$  | -0.707      | 0.113      | -0.709   | -6.261 | 0.000 |
| $x_t$ (年份) | 145.742     | 63.308     | 0.095    | 2.302  | 0.026 |

表8 使用自回归模型对中国人口总量的预测结果 万人

| 年份   | 人口实际总量   | 人口预测总量   | 人口预测偏差   | 预测相对偏差(%) |
|------|----------|----------|----------|-----------|
| 1990 | 114333.0 | 114193.3 | 139.7412 | 1.22      |
| 1991 | 115823.0 | 115775.8 | 47.19468 | 0.41      |
| 1992 | 117171.0 | 117169.2 | 1.84706  | 0.02      |
| 1993 | 118517.0 | 118432.1 | 84.88754 | 0.72      |
| 1994 | 119850.0 | 119792.3 | 57.73579 | 0.48      |
| 1995 | 121121.0 | 121132.9 | -11.8948 | 0.10      |
| 1996 | 122389.0 | 122382.9 | 6.12804  | 0.05      |
| 1997 | 123626.0 | 123671.9 | -45.8602 | 0.37      |
| 1998 | 124761.0 | 124913.0 | -152.047 | 1.22      |
| 1999 | 125786.0 | 126011.9 | -225.892 | 1.80      |
| 2000 | 126743.0 | 127005.7 | -262.718 | 2.07      |
| 2001 | 127627.0 | 127967.8 | -340.820 | 2.67      |
| 2002 | 128453.0 | 128860.4 | -407.447 | 3.17      |

表8给出了使用自回归模型对中国人口总量的预测结果、预测偏差及其相对偏差(预测偏差与人口实际总量之比),其中2002年年度的预测偏差相对较大为407.447万,相对偏差为3.17%,1992的预测偏差最小为1.84706,相对偏差为0.02%。显然,自回归模型与简单回归模型相比,模型的性能有了极大改进,模型对中国人口总量的预测精度有了很大的提高。本研究构造的自回归模型较好地克服了赵进文综合模型存在的回归系数统计不显著及解释变量数据难以采集两个缺陷,是一个较为理想的中国人口总量预测模型。

参考文献:

1. 赵进文(2003):《中国人口总量与GDP总量关系模型研究》,《中国人口科学》,第3期。
2. 李建民、王金营(2000):《中国生育率下降经济后果的计量分析》,《中国人口科学》,第1期。
3. 袁建文(2002):《经济计量学实验》,科学出版社。
4. 国家统计局(2002):《中国人口统计年鉴(2002)》,中国统计出版社。
5. 国家统计局人口和社会科技统计司、劳动和社会保障部规划财务司(2002):《中国劳动统计年鉴(2002)》,中国统计出版社。
6. 崔红艳(2000):《关于中国人口总量的探讨》,《人口研究》,第5期。
7. 郑祖康(2001):《GDM模型及2000年中国人口生存函数预测》,《应用概率统计》,第2期。
8. 翟振武(2000):《中国人口科学研究的发展与展望》,《中国人民大学学报》,第6期。

(责任编辑: 朱 犁)