

模糊线性回归模型在航空客运周转量预测中的应用研究

□ 周崇华, 苏志哲

(鼎汉沪通工程咨询(上海)有限公司, 上海 200040)

摘要: 为能准确把握航空客运周转量的未来特征, 寻求高精度预测方法是学术界多年致力研究的热点。拟采用模糊线性回归模型预测我国航空客运周转量, 并利用因子分析法对该模型及常见预测方法进行评估。评价结果可知: 模糊线性回归模型的预测精度不仅高于单因数趋势外推的预测方法(灰色模型、增长率统计法和指数平滑法), 也高于多因素综合性预测模型(多元线性回归), 说明此模型在预测航空客运周转量的变化中有一定实用价值。

关键词: 模糊线性回归模型; 航空客运周转量; 因子分析法; 评价

中图分类号: F560

文献标识码: A

文章编号: 1671-3400(2010)07-0031-03

Study on Forecasting of Air Passenger Turnover with Fuzzy Linear Regression Model

ZHOU Chong-hua, SU Zhi-zhe

(THI Shanghai Consultants INC, Shanghai 200040, China)

Abstract: In order to grasp the future characters of air passenger turnover, it is a significant issue how to select a forecast model with higher estimate accuracy at present. The papers build a fuzzy linear regression model between air passenger turnover and relative factors, and forecast air passenger turnover. Then the papers undergo a comprehensive analysis and evaluation to the regression analysis, grey model, growth statistics, exponential smoothing and fuzzy linear regression which predict air passenger turnover by factor analysis. From the evaluative result, it is believed that the prediction precision of fuzzy linear regression method is higher than the other methods, and it can be satisfactorily applied to predict air passenger turnover.

Keyword: Fuzzy linear regression model; Air passenger turnover; Factor analysis; Evaluate

0 引言

模糊数学作为一门新的数学分支, 以“模糊集合论”为基础, 为我们提供了一种处理不确定性和不精确性问题的新方法, 目前正广泛用于工业、农业、环境、医学、机械、电子、管理及交通等领域。采用模糊数学改进后的线性回归(模糊线性回归)模型在电力、用水量、剂—量—反应评定及水文等预测中取得了比较好的预测效果。拟

运用模糊线性回归模型来预测我国航空客运周转量, 并对此方法与其他常用方法进行对比、评价。

1 模糊线性回归模型

日本田中英夫(H·Tanaka)等人^[1]在20世纪80年代不断提出模糊线性回归的概念和方法, 我国学者于九如和杨泽华等人^[2]在20世纪90年代开始对模糊线性回归模型进行研究和应用, 它的理论基础如下:

1.1 模型简介

和经典的线性回归分析类似, 设变量 y 与它的相关因素 $x_1, x_2, \dots, x_j (j = 1, 2, \dots, n)$ 有线性关系: $y = A_1 x_1 + A_2 x_2 + \dots + A_n x_n$ 。回归分析的问题是利用已知的 m 组观测数据 $y_i, x_{i1}, x_{i2}, \dots, x_{ij} (i = 1, 2, \dots, m)$ 去估计回归系数 A_j 。但在

收稿日期: 2010-04-21

作者简介: 周崇华(1977-), 男, 汉族, 湖南衡阳人, 鼎汉沪通工程咨询(上海)有限公司工程师, 规划师, 硕士, 主要研究方向: 智能交通、交通控制、交通规划及交通政策研究。

模糊线性回归(简称FLR)分析中认为模型具有模糊性。具体地说,回归系数 A_j 是模糊数,于是模型的拟合值 \bar{y}_i 与观测值 y_i 之间的偏差就是由这种模糊性引起的。通常取 A_j 三角模糊数 $A_j(\alpha_j, c_j)$,且满足:①必须使各回归系数的模糊幅度之和为最小(即精度最大);②按一定的置信水平 H ,必须能“覆盖”所有的观测数据 y_i 。根据以上原则,求解FLR可转化为求解线性规划:

$$\begin{cases} \min j = c_1 + c_2 + \dots + c_n \\ \text{s.t.} \begin{cases} \sum_j \alpha_j x_{ij} - (1-H) \sum_j c_j |x_{ij}| \leq y_i \\ \sum_j \alpha_j x_{ij} + (1-H) \sum_j c_j |x_{ij}| \geq y_i \end{cases} \end{cases} \quad (1)$$

模型(1)的解中 $A_j(\alpha_j, c_j)$ 可能有下述四种情形:① $\alpha_j \neq 0, c_j > 0$;② $\alpha_j \neq 0, c_j = 0$;③ $\alpha_j = 0, c_j > 0$;④ $\alpha_j = 0, c_j = 0$ 。和经典回归一样,①和②表示 y 与 x_j 相关,但在② A_j 为精确数。③和④表示 y 与 x_j 不相关,当然可以在模型(1)中将 x_j 项剔除。(但在③,如果将 x_j 剔除,可能对模型有微小的影响。)

1.2 模型评价

模型(1)解得后,可按下述标准评价观测值拟合的优劣,即:①各观测值 y_i 对模型的隶属度 $\mu_i(y_i)$,一般认为各 $\mu_i(y_i)$ 的值如都大于0.5,就应认为是比较好的拟合;②各拟合的中心值和观测值 y_i 的相对偏差: $\Omega_1 = \frac{|y_i - \sum_j \alpha_j x_{ij}|}{y_i}$;

模糊幅度对观测值 y_i 的比, $\Omega_2 = \frac{\sum_j c_j |x_{ij}|}{y_i}$;后两个比值如都在30%以内,一般认为拟合是可以接受的。

同时,我们还可通过对配对的 t 检验来考察模拟结果与实际测定值的差异性是否显著,配对 t 检验的 t 值计算公

$$\text{式为: } t = \frac{\bar{d}}{S_{\bar{d}}} = \frac{\bar{x}_1 - \bar{x}_2}{S_{\bar{d}}}, \text{ 式中: } S_{\bar{d}} = \sqrt{\frac{\sum(d - \bar{d})^2}{(n-1)n}} = \sqrt{\frac{\sum d_i^2 - \frac{(\sum d_i)^2}{n}}{(n-1)n}}$$

2 航空客运周转量预测

2.1 预测模型

航空客运周转量不仅与GDP息息相关,还与公路、铁路和水运的客运周转量密切联系,不可分割。本文根据国家统计局公布的1980~2008年《全国年度统计公报》^[3]的历史数据以及模型特征建立FLR模型。鉴于2008年后与航空客运周转量相关因素的预测数据无法取得,本文以1990~2004年共计15年的航空客运周转量及其相关影响因素数据建立FLR模型,以2005~2008年共计4年的航空客运周转量作为验证值(相对于其预测值)来对此模型精度进行研究。令 Y 为航空客运周转量,与 Y 相关的因素有:GDP(x_1)、公路客运周转量(x_2)、铁路客运周转量(x_3)、水运客运周转量(x_4),可得: $Y = A_1 x_1 + A_2 x_2 + \dots + A_4 x_4$,其中回归系数 A_j 为三角模糊数 $A_j(\alpha_j, c_j)$, $c_j \geq 0$, $(j = 1, 2, 3, 4)$ 。

为求出 A_j ,解下列线性规划:

$$\begin{cases} \min J = c_1 + c_2 + \dots + c_4 \\ \text{s.t.} \begin{cases} \sum_j \alpha_j x_{ij} - (1-H) \sum_j c_j |x_{ij}| \leq y_i, i = 1, 2, \dots, 15 \\ \sum_j \alpha_j x_{ij} + (1-H) \sum_j c_j |x_{ij}| \geq y_i, i = 1, 2, \dots, 15 \end{cases} \end{cases} \quad (2)$$

将1990~2004年的相关数据代入式(2),取 $H = 0.5$,利用lindo软件,解得三角模糊数 $A_j(\alpha_j, c_j)$ 为: $A_1 = (0.008\ 617, 0.000\ 223)$, $A_2 = (0, 0)$, $A_3 = (0.061\ 214, 0.013\ 447)$, $A_4 = (0, 0)$ 。与经典回归一样,当 $\alpha_j = 0$ 时,表示 x_j 与 y 不相关。因此,将那些 $\alpha_j = 0$ 的 x_j 从模型中剔除(包括那些 $\alpha_j = 0, c_j > 0$),可得 $Y = A_1 x_1 + A_3 x_3$ 。

2.2 模型评价及预测

模型建立后,将1990~2004年的相关数据代入模型中,求出拟合值,并与实际数据进行比较和评价。

表1 模型评价

年份	评价		
	$\mu_i(y_i)$	Ω_1	Ω_2
1990	0.179	0.115	0.140
1991	0.038	0.135	0.141
1992	0.998	0.001	0.120
1993	0.445	0.058	0.105
1994	-0.143	0.126	0.111
1995	-0.034	0.096	0.093
1996	0.375	0.050	0.080
1997	-0.342	0.111	0.083
1998	-0.673	0.142	0.085
1999	-0.394	0.120	0.086
2000	0.388	0.049	0.081
2001	0.679	0.025	0.078
2002	0.081	0.065	0.071
2003	0.605	0.028	0.072
2004	-1.387	0.144	0.060

由表1可知,除了 $\mu_i(y_i)$ 有几个指标没有达到0.5之外, Ω_1 和 Ω_2 的值都小于0.3,所拟合的模型较好。 T 检验表明,模拟结果与实际测定值的 $t = -0.814$, $t < t_{0.05(14)} = 1.76$,说明模拟结果与实际测定值之间差异性不显著,模拟效果好。

利用模型可预测2005~2008年的航空客运周转量(亿人km)值为:1942.127、2209.819、2566.855、3067.033。

3 与常见预测模型比较

上章预测了2005~2008年航空客运周转量,预测结果的精度如何呢?与其他常用的预测方法相比,预测精度又如何呢?这些都是对一种新的预测方法所关心的问题。本章采用因子分析法并以均方根误差等7个衡量预测方法精度的评价指标为依据对回归分析(简称RA)、

灰色模型 (简称 GM)、增长率统计法 (简称 GS)、指数平滑法 (简称 ES) 以及 FLR 模型进行综合评估。

3.1 4 种常见模型对航空周转量的预测

采用上述 4 种常用方法预测 2005~2008 年我国航空周转量 (亿人 km), 其中 RA 预测值为: 2 338.699, 2 740.173, 3 203.684, 3 912.354; GM 预测结果为: 1 778.706, 1 987.799, 2 221.471, 2 482.611; GS 预测结果为: 2 045.691, 2 347.494, 2 720.586, 3 204.802; ES 预测结果为: 1 857.366, 2 077.290, 2 312.494, 2 562.978。

3.2 预测模型优选

3.2.1 评价模型的建立

测定预测精度的评价指标有^[4]: 最大拟合误差值 (S_1)、平均误差 (S_2)、平均绝对误差 (S_3)、平均相对误差 (S_4)、平均相对误差绝对值 (S_5)、均方根误差 (S_6)、Theil 不等系数 (S_7)、偏方比、方差比及协方差比等。本文以前 7 个评价指标为依据, 对上述 5 种预测模型拟合的好坏程度进行评判。由上述的 7 个评价指标可组成预测模型精度评价集合: $V = \{S_1, S_2, S_3, S_4, S_5, S_6, S_7\}$, 其评价矩阵为:

$$V = \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \\ V_6 \\ V_7 \end{bmatrix} = \begin{bmatrix} S_{11}, S_{12}, S_{13}, S_{14}, S_{15}, S_{16}, S_{17} \\ S_{21}, S_{22}, S_{23}, S_{24}, S_{25}, S_{26}, S_{27} \\ S_{31}, S_{32}, S_{33}, S_{34}, S_{35}, S_{36}, S_{37} \\ S_{41}, S_{42}, S_{43}, S_{44}, S_{45}, S_{46}, S_{47} \\ S_{51}, S_{52}, S_{53}, S_{54}, S_{55}, S_{56}, S_{57} \end{bmatrix}$$

3.2.2 预测模型的优选

根据前文 5 种模型预测值和 2004~2008 年的实际值 (2 044.9, 2 369.9, 2 791.75, 2 882.8), 以及评价矩阵, 可得各个评判模型优劣的指标值, 如表 2 所示:

表 2 5 种预测方法的评价模型指标值

评判指标 预测模型	S_1	S_2	S_3	S_4	S_5	S_6	S_7
RA	1 029.554	-526.403	526.403	-0.201	0.201	602.730	0.107
GM	570.230	404.678	404.678	0.159	0.159	418.979	0.090
GS	322.002	-57.318	104.078	-0.020	0.037	165.261	0.032
ES	479.206	319.793	319.793	0.124	0.124	336.420	0.071
FLR	224.845	75.867	167.983	0.034	0.066	173.699	0.035

依据因子分析法原理, 利用 SPSS 软件对表 2 中各数据进行标准化处理, 以消除量纲的影响, 并计算出各指标变量正交旋转后的因子载荷矩阵, 各因子所对应的特征值、贡献率、累计贡献率等。由此可知, 2 个主因子方差贡献率达到 99.600% (99.600% > 85%), 确定主因子个数为 2 个, 并可求得因子得分矩阵 (表 3)。

根据因子得分矩阵, 计算出各主因子的分值, 然后以主因子的方差贡献率为权数, 计算 5 个模型的综合得分如表 4 所示。鉴于本文是对航空客运周转量预测模型

优劣进行分析, 其分析的指标是预测误差, 属于逆向指标, 因而数值越小, 预测模型的精度越精确。若只从因子得分高低来看的话, RA 得分最高, 综合排名第一, 而 FLR 得分最低, 综合排名最后。但结合实际和本文所研究的对象, 可知所进行的综合排名和理论上的综合排名恰恰是相反的。因此, 可以取模糊线性回归模型为最优拟合模型。

表 3 因子得分矩阵

	主因子	
	1	2
S_1	.990	-.086
S_2	-.528	.849
S_3	.960	.272
S_4	-.533	.846
S_5	.950	.301
S_6	.991	.134
S_7	.946	.321

表 4 因子得分及排名

预测模型	主因子 1	主因子 2	Z	综合排名	实际综合排名
RA	2 400.250	673.738	1 971.777	1	5
GM	1 154.707	435.072	976.113	2	4
GS	612.810	143.066	496.233	4	2
ES	946.076	345.413	797.008	3	3
FLR	516.015	188.446	434.723	5	1

4 结论

(1) 建立航空客运周转量与相关因素的模糊线性回归模型, 并进行预测。结果表明, 该模型对航空客运周转量的预测精度比较高, 为解决具有不确定性因素的航空客运周转量预测提供了一种新的方法。

(2) 采用因子分析法对线性回归、灰色模型、增长率统计法、指数平滑法以及模糊线性回归 5 种预测模型进行评估、优选。结果认为, 模糊线性回归模型的实际排名第一、预测误差最少、精度最高。由此表明, 模糊线性模型在预测航空客运周转量的变化中有一定的实用价值。

参考文献:

- [1] Tanaka, H., S. Uegima and K. Asai, Linear regression analysis with fuzzy model, IEEE Transactions on Systems, Man and Cybernetics[J]. 1982, 12(6): 903-907.
- [2] 于九如, 杨泽华. 模糊线性回归及其应用实例 [J]. 系统工程理论与实践, 1995(4): 32-37.
- [3] 中国国家统计局. 全国年度统计公报 [R/OL]. (2009-02-26) [2010-04-10] <http://www.stats.gov.cn/tjgb/>
- [4] 徐国祥, 胡清友. 统计预测与决策 [M]. 上海: 上海财经大学出版社, 2004.