

文章编号: 1005-0523(2019)06-0064-06

考虑滞后期的机场旅客吞吐量预测

刘月, 朱金福, 陈娴

(南京航空航天大学民航学院, 江苏 南京 211106)

摘要: 机场旅客吞吐量的影响因子通常存在滞后期, 论文在选定了影响因素之后, 通过随机森林重要度指数, 确定各因子的滞后期; 接着, 以 1987—2018 年南京禄口国际机场的数据建立吞吐量预测的逐步回归模型; 最后, 对模型效果进行评价并基于建立的模型对 2019 年和 2020 年南京禄口国际机场的吞吐量进行预测。结果表明: 影响因子对旅客吞吐量的影响存在 1~9 年不等的滞后期; 通过与未考虑滞后期的模型进行对比, 发现考虑滞后期的模型拟合效果更优, MAE 减小显著。因此, 将滞后期引入机场旅客吞吐量预测是很有必要的。

关键词: 滞后期; 预测; 随机森林; 逐步回归; 神经网络

中图分类号: [U8] **文献标志码:** A

DOI: 10.16749/j.cnki.jecjtu.2019.06.010

随着我国民航运输的快速发展, 机场吞吐量每年以 10% 的速度增加^[1]。机场旅客吞吐量预测是机场规划设计的主要参数, 是实现机场资源有效配置的基本根据。其预测的准确程度直接影响着机场新建、改扩建以及机队规模等问题, 决定了机场的需要达到的保障水平。

国内外对于吞吐量预测的研究有很多, Tobias G 和 Wadud Z 分别考虑了机场本身的服务能力以及机场辐射区域内的经济社会发展情况和国家层面的总时间序列信息以及其他机场的赞助, 都使用重力模型对吞吐量进行预测^[2-3], Kim S, Shin D H 利用搜索查询中的大数据来确定短期航空旅客需求的预测模型, 基于 k-交叉验证方法建立预测模型^[4]。焦朋朋, 赵素霞, 张玉, 张倩丽, 杨尚文等人也都对此作了一定研究, 在他们的研究中 GDP 和人口数都是主要影响指标, 另外根据研究省份的不同, 还有第三产业比重、居民可支配收入、社会消费品总额等因素, 亦有着重要影响, 具体指标的选取, 因研究对象不同而有所调整^[5-9]。焦朋朋, 张玉, 张倩丽等人运用多元线性回归方法进行预测^[5,7-8], 赵素霞运用神经网络方法进行预测^[6], 杨尚文, 胡明华将计量经济法和时间序列预测方法相结合, 使用多元线性回归模型确定组合预测方法的权重, 再以计量经济学方法进行预测, 既优于传统的计量经济法, 也优于时间序列预测方法^[9]。

机场旅客吞吐量的影响因素有很多, 有些对于吞吐量的影响在短期内会很快体现, 但有些影响因素对航空市场的带动存在一定的滞后期, 诸如固定资产投资等, 对吞吐量的影响可能无法立刻体现, 但这些因素对航空市场同样有重要作用, 如果不考虑滞后期很可能遗漏这类变量对吞吐量的影响。滞后性是市场的一个重要属性, 是指因变量不仅受到同期各种因素的影响, 而且也受到过去某些时期的各种因素甚至自身过去值的影响。滞后期产生的原因主要有两方面: 一是产品存在一定的生产周期; 二是人们的行为方式往往滞后于经济形势的变化。对于滞后期的研究, 经济、医学等领域已经有较成熟的成果。近年来, 有学者开始将滞后性引入民航领域, 分别以上海、天津为例, 运用格兰杰因果检验方法证实了经济社会指标对机场旅客吞吐量的影响存在滞后期^[10-11]。但目前的研究仅限于在给出某个确定的滞后期时, 判断影响因素对吞吐量的影响

收稿日期: 2019-03-28

基金项目: 南京航空航天大学研究生创新基地(实验室)开放基金项目(kfj20180723)

作者简介: 刘月(1996—), 女, 硕士研究生, 主要研究方向为机场运输规划。

通讯作者: 朱金福(1955—), 男, 教授, 博士, 主要研究方向为机场运输规划。

是否依旧显著,未能确定最佳滞后期,更未将其与预测相结合。

针对以上问题,从市场的滞后性出发,首先利用随机森林方法确定不同影响因素的最佳滞后期,在此基础上,利用逐步回归方法与未考虑滞后期的预测结果进行对比,并利用禄口国际机场1987—2018年数据建模验证。结果表明,考虑了滞后期后,预测精度提高显著。

1 影响因素选取

从人口、居民生活、经济、城市交通4个方面选择年末户籍人口(p_{hj})/万人、城市居民人均可支配收入(inc_city)/元、城镇居民人均消费支出(ex_city)/元、农村居民人均可支配收入(inc_rural)/元、农村居民人均消费支出(ex_rural)/元、社会消费品零售总额($turnover$)/亿元、地区生产总值(GDP)/亿元、第一产业生产总值($GDP1$)/亿元、第二产业生产总值($GDP2$)/亿元、第三产业生产总值($GDP3$)/亿元、固定资产投资($investment$)/亿元、铁路旅客运输量(p_rail)/万人12个指标。

以禄口国际机场为例,选择1987—2018年的指标数据进行研究,影响因素的数据来自历年南京统计年鉴,机场吞吐量值来源于民航局公布数据。

由于时间序列呈指数型增长,首先对数据取对数。以Pearson相关系数衡量变量之间的相关性,在E-views中得到相关系数结果,各因素与旅客吞吐量的相关系数均大于0.8,对吞吐量的影响显著。

作出各影响因素与旅客吞吐量的散点图如图1,按照从上至下,从左至右的顺序,分别表示城镇居民人均消费支出,农村居民人均消费支出,地区生产总值,第一产业生产总值,第二产业生产总值,第三产业生产总值,城市居民人均可支配收入,农村居民人均可支配收入,固定资产投资,年末户籍人口,铁路旅客运输量,社会消费品零售总额。可以看出,除了铁路旅客运输量,其他影响因素与旅客吞吐量基本呈线性关系。

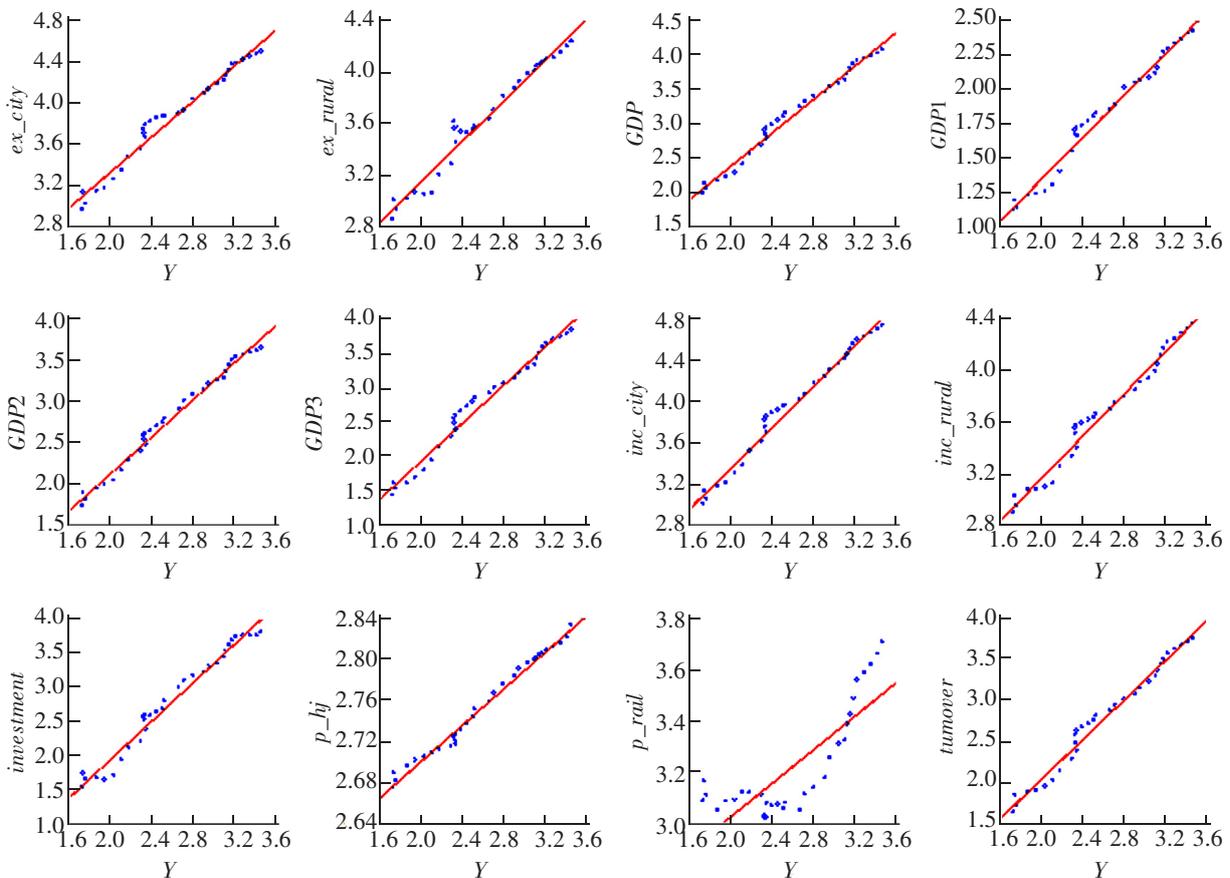


图1 影响因素与旅客吞吐量散点图

Fig.1 Scatter plots of influencing factors and passenger throughput

2 预测模型建立与评价

针对最佳滞后期的数据进行研究,由于研究的问题自变量较多,为了在不遗漏重要变量的同时尽可能减小多重共线性,选择逐步回归方法进行建模。由于铁路旅客运输量与旅客吞吐量明显不呈线性关系,因此在建立回归模型时,剔除这项指标。

2.1 最佳滞后期的确定——基于随机森林重要度

文献[10-11]将滞后性引入民航领域,认为影响因素对旅客吞吐量的影响存在滞后性。根据文献[12]的研究,各行业的滞后期从化工医药行业的9年到科研机构的5年不等。因此,论文搜集1987—2018年禄口国际机场影响因素的数据,规定最大滞后期不得超过9年,借助随机森林中的重要度确定各个影响因素在不同滞后期下对旅客吞吐量影响,重要度指数越大,影响越大。

在R平台运行得到各因素不同滞后期下重要度如表1。

表1 各因素不同滞后期重要度指数表
Tab.1 Importance index of each factor at different lag periods

变量	1	2	3	4	5	6	7	8	9
<i>ex_city</i>	9.86	9.67	8.43	9.15	7.61	8.87	9.14	9.85	8.61
<i>ex_rural</i>	11.12	10.94	10.54	10.02	8.26	8.26	8.52	9.71	8.41
<i>GDP</i>	8.98	9.17	9.11	9.69	8.23	9.95	8.63	8.53	9.07
<i>GDP1</i>	8.56	9.01	8.98	9.77	9.42	8.48	8.79	8.79	10.02
<i>GDP2</i>	8.62	8.91	9.89	8.63	8.95	9.11	9.62	9.39	9.19
<i>GDP3</i>	9.78	9.87	9.12	8.86	10.02	7.66	9.02	8.60	9.05
<i>inc_city</i>	9.28	9.05	10.07	8.76	8.52	8.59	8.56	9.18	9.13
<i>inc_rural</i>	8.93	8.05	9.01	8.80	9.42	9.20	8.51	9.53	9.48
<i>investment</i>	8.68	9.26	9.72	8.18	9.39	8.42	9.05	8.59	9.16
<i>p_hj</i>	9.48	8.79	9.33	9.02	9.00	8.99	8.21	9.12	8.50
<i>turnover</i>	9.23	8.65	8.57	9.22	9.74	9.38	8.31	9.80	9.12

根据表1中的重要度指数,得出各影响因素的最佳滞后期,如表2。

表2 影响因素最佳滞后期表
Tab.2 The best lag period of the influencing factors

指标	滞后期/年
年末户籍人口/万人	7
城市居民人均可支配收入/元	9
农村居民人均可支配收入/元	8
城镇居民人均消费支出/元	6
农村居民人均消费支出/元	2
<i>GDP</i> /亿元	3
第一产业 <i>GDP</i> /亿元	1
第二产业 <i>GDP</i> /亿元	6
第三产业 <i>GDP</i> /亿元	6
社会消费品零售总额/亿元	8
固定资产投资/亿元	3

各因素与旅客吞吐量之间的滞后期在 1~9 年不等,选取表中各因子对应滞后期下的值与旅客吞吐量建立预测模型。

2.2 预测模型建立

利用 Eviews10.0 软件进行逐步回归模型的构建。每得到一个回归结果,分析模型中各变量是否显著(在 95%置信区间下,各变量的 P 值是否小于 0.05);若模型中的某个变量不再具有显著性,则将具有最大 P 值的变量从模型中删除。

选取 1995—2012 年数据作为训练数据,2013—2018 年数据作为测试数据(此处年份均指 Y 的年份),在 Eviews10.0 中运用逐步回归进行建模,最终得到的回归结果如表 3 所示。

表 3 逐步回归结果表(考虑滞后期)
Tab.3 Stepwise regression results (considering lag period)

变量	系数	标准误差	t-Statistic	P
C	2.008 345	0.351 743	5.709 688	0.000 1
inc_rural	1.276 359	0.140 713	9.070 679	0.000 0
$GDP2$	1.128 840	0.204 045	5.532 319	0.000 1
ex_city	-1.581 911	0.173 651	-9.109 713	0.000 0
$investment$	0.763 510	0.207 218	3.684 569	0.003 1
GDP	-0.918 425	0.355 099	-2.586 392	0.023 8

$R^2=0.992\ 742$,接近 1,拟合效果良好。各变量 P 值均小于 0.05,因此在 95%的置信区间下,各变量对 Y 的影响都是显著的。

根据软件给出的结果,旅客吞吐量 Y 与影响因素之间的模型关系为

$$Y=1.276\ 359inc_rural+1.128\ 84GDP2-1.581\ 911ex_city+0.763\ 51investment-0.918\ 425GDP+2.008\ 345 \quad (1)$$

代入测试数据,得到预测结果见表 4,以平均绝对误差(MAE)为衡量指标,计算公式为

$$MAE=\frac{1}{n} \sum_{i=1}^n |Y_i-\hat{y}_i| \quad (2)$$

式中: i 为对应年份下测试组别。将测试数据代入公式得到测试数据预测误差为 6.78%。

表 4 考虑滞后期的真实值与预测值对比表
Tab.4 Comparison of real and predicted values considering the lag period

年份(以 Y 为准)	预测值/万人	真实值/万人
2013	1 687.56	1 501.18
2014	1 760.69	1 628.38
2015	1 973.87	1 916.38
2016	2 216.39	2 235.80
2017	2 456.1	2 582.29
2018	2 531.1	2 858.15

根据上述建立的模型,预测 2019 年,2020 年南京禄口国际机场的旅客吞吐量分别为 3 097.08 万人和 3 369.03 万人。

2.3 预测结果评价

为了证明引入滞后期对提高预测精度的有效性,运用逐步回归方法和BP神经网络方法在未考虑滞后期的情况下同样对禄口国际机场的旅客吞吐量建立预测模型。按照4:1的比例确定训练数据和测试数据,选取1987—2012年数据作为训练数据,2013—2018年数据作为测试数据,同样对2019年,2020年两年禄口国际机场旅客吞吐量做预测。其中逐步回归模型的建立与考虑滞后期的模型方法相同,神经网络模型的构建在matlab2016a上完成。根据选取的指标,设置网络输入层节点数为12,输出层节点数为1。经调试发现,隐含层节点数为4,隐含层传递函数选用logsig,输出层传递函数选用purelin,训练函数选择traingdm,学习率为0.005时预测效果最佳。

两者对于测试数据的预测结果如表5,同样以MAE为衡量指标,测试误差分别为16.73%和12.98%。逐步回归预测2019年,2020年的旅客吞吐量分别为2677.1万人和3007.6万人;神经网络预测2019年,2020年的旅客吞吐量分别为2550.9万人和2677.3万人。

表5 预测结果对比表
Tab.5 Comparison of prediction results

年份	逐步回归预测值	神经网络预测值	实际值
2013	1 420.02	1 578.34	1 501.18
2014	1 478.42	1 656.91	1 628.38
2015	1 560.07	1 768.07	1 916.38
2016	1 684.87	1 851.40	2 235.80
2017	1 908.76	1 971.51	2 582.29
2018	2 388.87	2 217.18	2 858.15

以上3个模型均以2013—2018年数据作为测试数据,得到旅客吞吐量数据预测值的测试误差,同时都对2019年和2020年旅客吞吐量做了预测,时间区间一致,外部环境一致,不对预测产生影响,因此预测结果具有可比性。

作出2013—2020年(2013—2018年为测试集,2019—2020年为预测集)未考虑滞后期的预测结果、考虑滞后期的预测结果以及真实值的图像如图2。其中,2019年和2020年旅客吞吐量真实值根据趋势外推法得出。

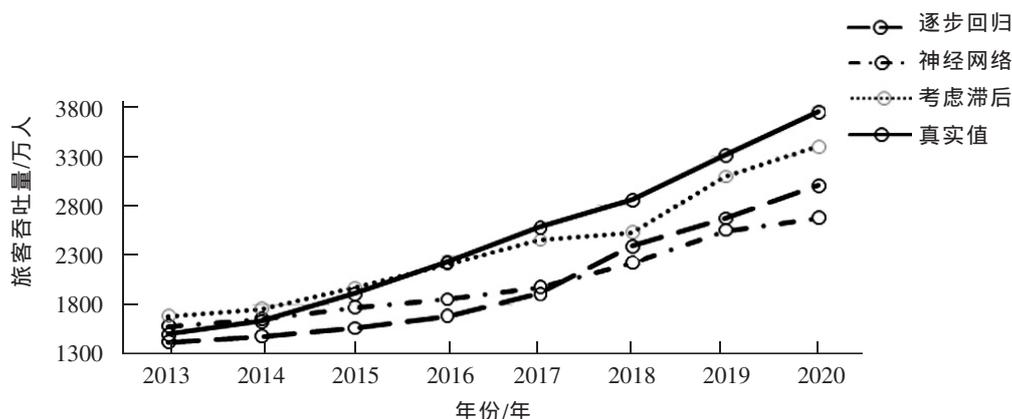


图2 航空旅客吞吐量预测值对比图

Fig.2 Comparison of predicted values of air passenger throughput

由图2可以看出两点:第一,对比未考虑滞后期的两种模型发现,对于短期预测,神经网络拟合更优;对于长期预测,逐步回归预测更接近于真实值。第二,对比考虑滞后期的预测模型和未考虑滞后期的预测模

型,考虑滞后期的预测模型比未考虑滞后期的模型减少了8年训练数据,但根据图像,考虑滞后期的预测结果更加接近真实值,直观证明了考虑滞后期的的重要性。

另外,根据 R^2 可以看出,是否考虑滞后对拟合效果影响不大,3种模型拟合效果均较好,但比较3种模型的 MAE 发现,考虑了滞后期后,尽管训练数据量减少,但考虑滞后期的模型 MAE 仍然减少显著,相较于逐步回归减少了9.95%,较于神经网络,减少了6.2%,精度提高显著,用数据说明了在机场旅客吞吐量预测中引入滞后期的的重要性。

3 结论

通过各因素与旅客吞吐量的滞后关系分析,发现影响因素对旅客吞吐量的影响存在滞后期,且滞后期在1~9年不等。通过对比考虑滞后期的模型和未考虑滞后期的模型的预测图像和 MAE 值,证明了将滞后性引入吞吐量预测的重要性。根据模型预测结果,禄口国际机场旅客吞吐量在2019年,2020年仍将保持高速增长态势,且2019年将突破三千万大关。

参考文献:

- [1] 王俊丹,曾小舟,冯琳. 基于DEA方法的我国机场运行效率评价[J]. 华东交通大学学报,2017(6):82-89.
- [2] TOBIAS G. Gravity models for airline passenger volume estimation[J]. Journal of Air Transport Management,2007,13(4):1-183.
- [3] WADUD Z. Modeling and forecasting passenger demand for a new domestic airport with limited data[J]. Transportation Research Record;Journal of the Transportation Research Board,2011,2214:59-68.
- [4] KIM S,SHIN D H. Forecasting short-term air passenger demand using big data from search engine queries[J]. Automation in Construction,2016:S0926580516301303.
- [5] 焦朋朋. 机场旅客吞吐量的影响机理与预测方法研究[J]. 交通运输系统工程与信息,2005,1(1):107-110.
- [6] 赵素霞. 基于因果反馈分析的机场旅客吞吐量预测方法研究——以北京首都国际机场为例[J]. 价值工程,2015,34(16):35-38.
- [7] 张玉,欧家杰. 海南省航空旅客需求量预测研究[J]. 中国外资,2013(6):267-268.
- [8] 张倩丽. 厦门高崎国际机场旅客吞吐量预测[J]. 科学技术创新,2017(23):21-22.
- [9] 杨尚文,胡明华. 机场旅客吞吐量预测的组合方法研究[J]. 武汉理工大学学报(交通科学与工程版),2009,33(2):231-233.
- [10] 龙继林,师萍. 我国民航运输服务与国民经济增长关系研究[J]. 生产力研究,2013(9):113-116.
- [11] 邢诗雨,董楚,熊钦. 天津民航业与地方经济协调发展实证研究[J]. 人力资源管理,2017(4):268-270.
- [12] BARUCH LEV. The capitalization, amortization, and value-relevance of R&D[J]. Industrial Health, 1996, 53(6):498-504.

Airport Passenger Throughput Forecast Considering Lag Period

Liu Yue, Zhu Jinfu, Chen Xian

(College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: The impact factor of airport passenger throughput usually has a lag period. After selecting the influencing factors, the paper determines the lag period of each factor by the random forest importance index. Then, a stepwise regression model for throughput prediction is established using the data of Nanjing Lukou International Airport from 1987 to 2018. Finally, the model effect is evaluated and the throughput of Nanjing Lukou International Airport in 2019 and 2020 is predicted based on the established model. The results show that the influence of impact factors on passenger throughput has a lag period ranging from 1 to 9 years. By comparing with the model without considering the lag period, it is found that the model fitting effect considering the lag period is better and the reduction is significant. Therefore, it is necessary to introduce the lag period into the airport passenger throughput forecast.

Key words: lag period; prediction; random forest; stepwise regression; neural network