

后疫情时期秦皇岛站铁路客运量的统计建模与预测效果分析

马嘉悦

燕山大学理学院, 河北 秦皇岛

收稿日期: 2023年1月23日; 录用日期: 2023年2月13日; 发布日期: 2023年2月27日

摘要

后疫情时期, 铁路客运需求发生巨大改变, 突如其来的疫情中断其逐年增加的趋势。建立合适的模型预测客运量, 对铁路运营工作的顺利展开起着至关重要的作用。论文运用2020年1月24日~2022年4月14日秦皇岛站铁路的逐日客运量数据, 提出了基于季节性差分自回归滑动平均模型(SARIMA)和霍尔特-温特斯模型(Holt-Winters)的铁路客运量组合预测模型。采用方差倒数法, 确定各单项模型在组合模型中的权重系数。分别用两个单一模型和组合模型进行15天客运量的短期预测, 并采用2022年4月15日~4月29日的实际数据验证预测效果。设置不同的样本梯度, 分析样本量对三个模型预测准确度的影响。

关键词

后疫情时期, 客运量预测, SARIMA模型, Holt-Winters模型, 组合预测模型

Statistical Modeling and Prediction Analysis of Passenger Volume of Qinhuangdao Railway Station in Post-Epidemic Period

Jiayue Ma

College of Science, Yanshan University, Qinhuangdao Hebei

Received: Jan. 23rd, 2023; accepted: Feb. 13th, 2023; published: Feb. 27th, 2023

Abstract

In the post-epidemic period, the demand for railway passenger transport changed dramatically, and the sudden epidemic interrupted its increasing trend year by year. The establishment of a

suitable model to predict passenger volume plays a vital role in the smooth development of railway operation. Based on the daily passenger volume data of Qinhuangdao Station from January 24, 2020 to April 14, 2022, this paper proposes a combined forecast model of railway passenger volume based on seasonal autoregressive integrated moving average (SARIMA) and Holt-Winters model (Holt-Winters). The inverse variance method is used to determine the weight coefficient of each single model in the combination model. Two single models and combination models were used to predict the 15-day passenger volume respectively, and the actual data from April 15 to April 29, 2022 were used to verify the prediction effect. Different sample gradients were set to analyze the influence of sample size on the prediction accuracy of the three models.

Keywords

Journals Post-Epidemic Period, Passenger Volume Forecast, SARIMA Model, Holt-Winters Model, Combinatorial Forecasting Model

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

中国铁路串点成线，连线成网，四通八达，覆盖面极广，如今已逐渐形成“八纵八横”的高铁网，交通大动脉越来越畅通。并且随着城市间人口流动性增强和由于经济发展带来的生活节奏的加快，人们对于交通速度也有着一定的追求。而铁路运输具备运量大、方便快捷、安全系数高等特点，通常会成为人们远途出行的首选方式。对铁路未来的客流量进行较为准确的预测，是铁路线路设计、场站规划、铁路运营的工作基础。通过算法模型预测出来的数据，是车辆运输规划和人员安排的决策依据，有助于在能够满足运输需求的前提下，减少资源的浪费，应对节假日期间客流突增的情况。因此，准确预测铁路客流量对于车站的长期发展具有非常重要的意义。尤其是在 2020 年 1 月，突如其来的疫情给铁路客运带来巨大冲击，中断了客流量逐年增加的趋势[1]。根据铁路营销网站的数据，秦皇岛站铁路客运量从 2020 年 1 月 24 日开始大幅下降，1 月客运量累计 39.4233 万人次，较 2019 年同比下降 12.20%，2 月客运量累计 3.6683 万人次，较 2019 年同比下降 90.81%。随着国内疫情有所好转，铁路客流逐渐增加，应用建立好的模型对未来客流进行预测，有助于疫情恢复阶段铁路运营工作的顺利展开。

目前，学者们对客运量的预测进行了深入的研究。如：蔡倒录等提出采用自回归移动平均模型(ARMA)预测新疆铁路客运量，拟合效果理想[2]。赵鹏等构建自回归综合移动平均模型(ARIMA)在对北京地铁站进行客流预测时具有很高的预测精度[3]。段然等考虑到客流量数据的周期性，选用季节性差分自回归滑动平均模型(SARIMA)对某铁路站点非节假日的客流数据进行短期预测，拟合效果良好[4]。王小凡等对青岛市铁路客运量分别建立多元回归模型和 BP 神经网络模型，通过比较证明了 BP 神经网络模型的预测效果更优[5]。桂文毅基于灰色线性回归模型预测哈尔滨铁路枢纽的客运量，经验证预测数据可靠[6]。由于客运量数据的特征较为复杂，仅用单一模型预测会有一些的局限性，对此许多学者构建组合模型，并证明其有效性和实用性。如：Gliovi 等提出了基于神经网络和遗传算法的组合模型，预测塞尔维亚的铁路月客运量[7]。Li 以 IOWHA 算子概念为基础，提出了基于 ARIMA-REGRESSION 的组合预测方法，对中国民航年客运量进行预测，证明比单一模型有更高的预测精度[8]。张蕾将支持向量回归(SVR)、SARIMA 模型和霍尔特 - 温特斯模型(Holt-Winters)进行两两组合，讨论发现每种组合模型的预测效果均

优于单一模型[9]。张玺君等提出了基于 SARIMA-GA-Elman 的组合预测模型, 研究结果表明其预测效果和单一模型相比有明显提升[10]。

目前对铁路客运量预测的研究多为月统计数据 and 年统计数据, 虽保证了较高的预测准确度, 但降低了对相关部门制定短期规划的参考价值。并且疫情在很大程度上影响了客运量数据的变化趋势。据此, 论文采集了后疫情时期秦皇岛站日客运量数据, 根据其具有线性特征和非线性特征, 提出了采用 SARIMA 模型和 Holt-Winters 模型对客运量进行短期预测。随后运用方差倒数法, 建立加权组合模型, 并对三种模型的预测效果进行比较分析。

2. 数据来源及白噪声检验

本文研究使用的秦皇岛市铁路车站客运量数据来源于铁路营销信息网站。以 2020 年 1 月 24 日~2022 年 4 月 15 日共 812 天的逐日客运量数据为样本数据, 利用 Rstudio 软件绘制数据序列, 如图 1 所示。从图中可以看出, 后疫情时期的客运量整体上呈波浪式变化和先增后减的变化趋势, 样本数据的均值、方差波动较大, 初步判断该客运量序列非平稳且受季节效应的影响。

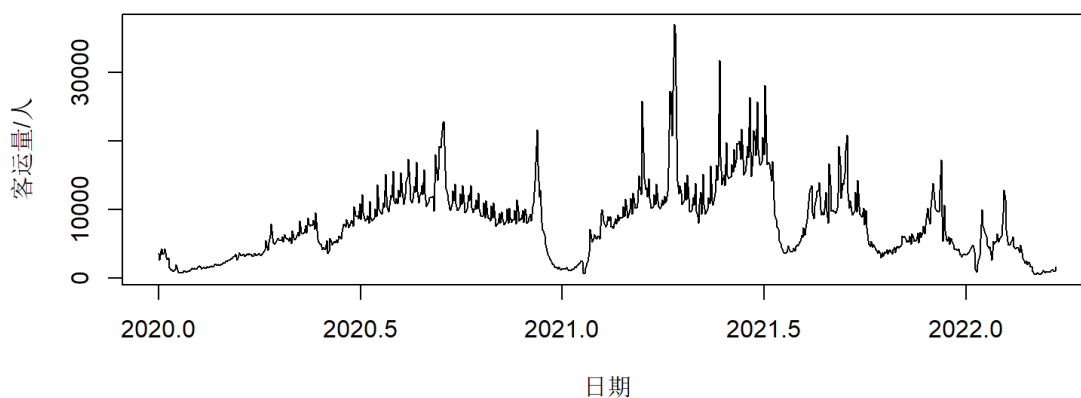


Figure 1. Timing diagram of daily passenger volume from January 24, 2020 to April 14, 2022

图 1. 2020 年 1 月 24 日~2022 年 4 月 14 日逐日客运量时序图

对序列进行因素分解, 来推断出各种因素对序列的综合影响, 结果如图 2。由因素分解图可以发现, 该序列受季节因素和随机因素影响, 有一定的周期性。但由于疫情期间, 铁路相关政策在很大程度上会对客运需求产生影响, 在短时间内使其发生较大变化, 所以没有明显的长期趋势。同时, 从因素分解图中也能看出, 数据中的季节趋势没有取决于数据的变动, 可认为客运量的日度数据具有一定的季节性特征, 即在周日出现高峰, 在周二出现低谷。故本文初步选择 SARIMA 模型和 Holt-Winters 模型分析样本数据, 探索最适合拟合预测秦皇岛站铁路客运量的时间序列模型。

先对时间序列进行白噪声检验。采用 LB 检验, 对该序列做延迟 6 阶、12 阶、18 阶的白噪声检验, 结果如表 1 所示。P 值均小于 0.05, 拒绝原假设, 由此判断该序列非白噪声。

3. 模型建立与比较分析

3.1. SARIMA 模型预测

绘制原始序列的 ACF 和 PACF 图, 如图 3 所示。观察 ACF 图, 发现自相关系数长期位于正值的一边, 同时自相关图还隐约周期为 7 的波动规律。这两个特征常存在于具有周期性的非平稳序列中, 并且与原序列的时序图(图 1)的趋势和特点十分吻合。

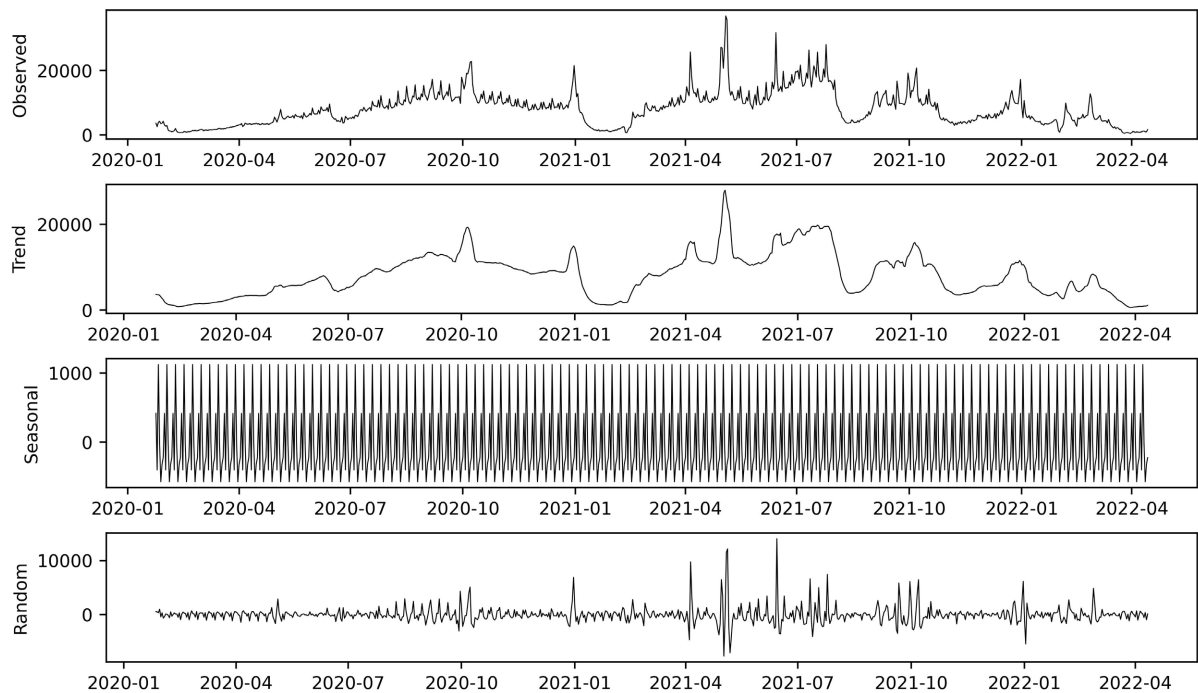


Figure 2. Factor decomposition diagram of sample data time series

图 2. 样本数据时间序列因素分解图

Table 1. LB test results of the time series

表 1. 时间序列的 LB 检验结果

延迟阶数	X-squared	P 值
6	3274	$<2.2e-16$
12	5470.9	$<2.2e-16$
18	7053.8	$<2.2e-16$

由于序列没有明显的长期趋势，所以需要进行一阶 7 步差分来消除季节性。差分后的时序图、ACF 图和 PACF 图如图 4 所示。相比于图 1 原始序列的时序图，变换后序列的均值和方差的波动明显减小，趋势更加平稳。ACF 和 PACF 不呈周期性震荡，说明经过一阶季节差分，季节性已被消除。同时，ACF 图呈现出稳定模型具有的“伪周期”特征。

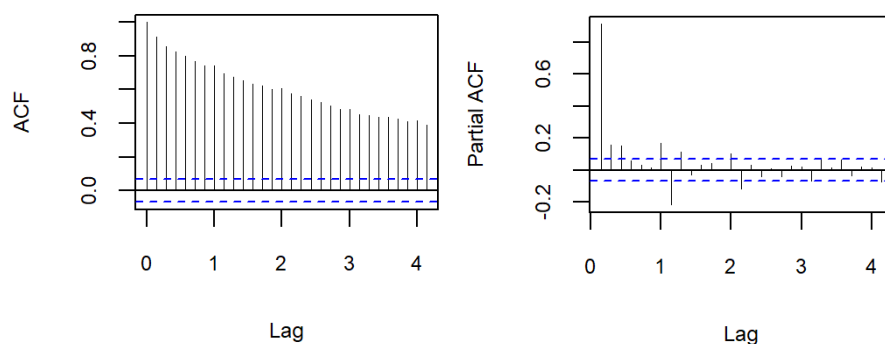


Figure 3. ACF and PACF diagrams of the original time series

图 3. 原始时间序列的 ACF 和 PACF 图

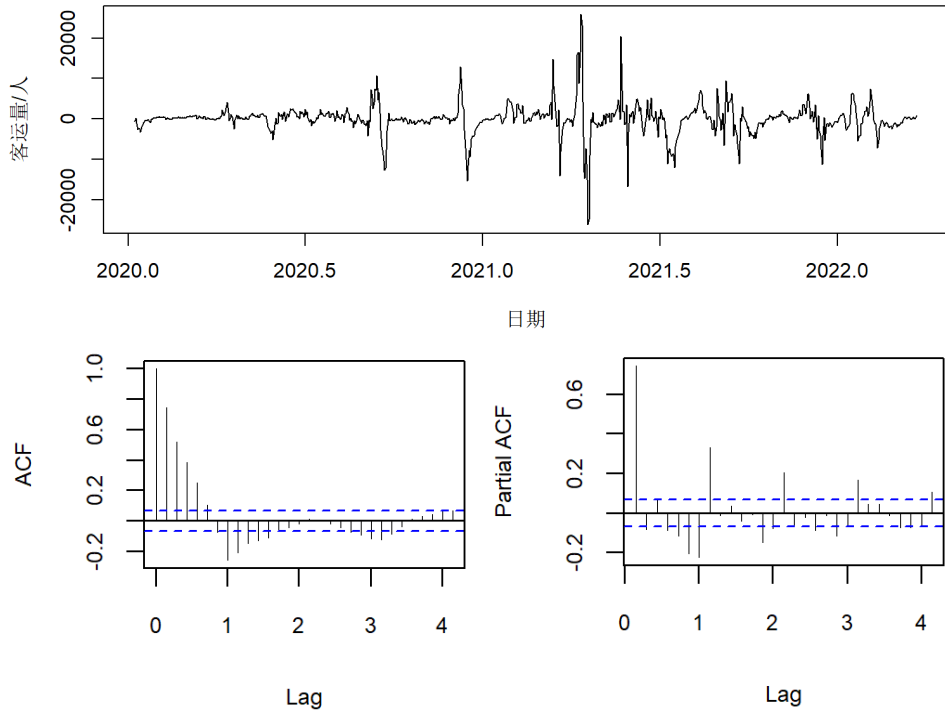


Figure 4. Timing diagram, ACF diagram and PACF diagram after the first order 7-step seasonal difference
图 4. 一阶 7 步季节差分后的时序图、ACF 图和 PACF 图

进一步对季节差分后的序列进行 ADF 检验，结果如表 2 所示。由统计量小于不同显著性水平下的临界值，P 值接近于 0，拒绝非平稳的原假设，可判断出此的序列平稳，可以满足建立 SARIMA 模型的平稳性需求。

Table 2. LB test results of the time series
表 2. 时间序列的 LB 检验结果

统计量	10%水平的临界值	5%水平的临界值	1%水平的临界值	P 值
-26.4414	-1.62	-1.95	-2.58	<2.2e-16

SARIMA 模型的表达形式为 $SARIMA(p,d,q)(P,D,Q)_s$ 。其中，在非季节性模型部分，参数 p 、 d 、 q 分别为自回归阶数、逐期差分阶数和移动平均阶数，在季节性模型部分，参数 P 、 D 、 Q 分别为季节性自回归阶数、季节性差分阶数和季节性移动平均阶数， S 为周期长度。SARIMA 模型的一般表达式为：

$$\Phi(B)\Phi_s(B)(1-B)^d(1-B)^D x_t = \Theta(B)\Theta_s(B)\varepsilon_t$$

其中， B 为后移算子， ε_t 为一个白噪声序列，且

$$\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p; \Phi_s(B) = 1 - \phi_1 B^S - \dots - \phi_p B^{pS}$$

$$\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q; \Theta_s(B) = 1 - \theta_1 B^S - \dots - \theta_q B^{qS}$$

由平稳性检验结果可以确定 $d = 0$ ， $D = 1$ ， $S = 7$ 。图 4 的 ACF 图显示，自相关系数在滞后 5 阶后减为 0，则 q 取 5。其在滞后 7 阶有明显的突起，则 $Q = 1$ 。PACF 图显示，偏自相关系数 1 阶后截尾，故 p 取 1，其在延迟 7 阶、14 阶和 21 阶时在 2 倍标准差范围之外，则 P 取 3。初步对铁路客运量序列建立

SARIMA(1,0,5)(3,1,1)₇ 模型。

但图定阶法有较强的主观性，所以还要通过 AIC 准则进一步较为客观的判断模型阶数的优劣。AIC 准则是将参数个数和拟合精度加权后得到的参数，AIC 函数值越小，参数个数和拟合精度的配置越优。对 SARIMA(1,0,5)(3,1,1)₇ 模型的 p 、 q 、 P 和 Q 在就近范围内适当增减，综合考虑 AIC 值和预测误差的大小，将 SARIMA(1,0,5)(5,1,0)₇ 定为最优模型，此时 AIC 值为 14593.96。对残差序列做延迟 6 阶的白噪声检验，得出 P 值为 0.9994，大于显著性水平 0.05，表明该模型显著成立，SARIMA(1,0,5)(5,1,0)₇ 模型对客运量序列拟合效果良好。SARIMA(1,0,5)(5,1,0)₇ 模型中各参数值如表 3 所示。

Table 3. Parameter values of the SARIMA model

表 3. SARIMA 模型各参数值

参数	ar1	ma1	ma2	ma3	ma4	ma5	sar1	sar2	sar3	sar4	sar5
回归系数	0.8828	-0.0611	-0.1521	0.0504	0.0599	0.0030	-0.7063	-0.5116	-0.4020	-0.2565	-0.1648

综上，可以得到模型口径为：

$$\begin{aligned} & (1-B^7)(1+0.7063B^7+0.5116B^{14}+0.402B^{21}+0.2565B^{28}+0.1648B^{35})x_t \\ & = (1+0.0611B+0.1521B^2-0.0504B^3-0.0599B^4-0.003B^5)\varepsilon_t \end{aligned}$$

使用 SARIMA(1,0,5)(5,1,0)₇ 模型预测 2022 年 4 月 15 日~2022 年 4 月 29 日共 15 天的秦皇岛车站铁路日客运量。在计算出预测值后，使用平均绝对百分比误差(MAPE)来评估模型预测结果的优劣。因为平均绝对百分比误差在避免了正负误差相互抵消的同时，能够直观地体现出误差与真实值间的相对程度，反映预测的准确度，且不依赖于其他指标。MAPE 越小，说明模型的预测效果越好。平均绝对百分比误差(MAPE)的计算公式为：

$$MAPE = \frac{1}{N} \sum_{t=1}^N \frac{|X_t - \hat{X}_t|}{X_t} \times 100\%$$

式中， X_t 为第 t 期的真实值， \hat{X}_t 为第 t 期的预测值， $t=1, \dots, N$ 。真实值和预测结果如表 4 所示。

由表 4 可以看出，基于 SARIMA 模型的铁路日客运量预测值与真实值的最小相对误差为 1.0585%，最大相对误差为 27.4035%，其中 4 月 24 日的误差偏大，其他时间的误差适中，MAPE 为 8.9761%，总体的预测准确度较高。

Table 4. Forecast of daily passenger volume of railway stations by SARIMA model from April 15 to April 29, 2022

表 4. 2022 年 4 月 15 日~4 月 29 日铁路车站日客运量 SARIMA 模型预测

日期	真实值	预测值	误差	相对误差/%	MAPE/%
2022/4/15	1711	1762	51	2.9807	
2022/4/16	1961	1750	-211	10.7598	
2022/4/17	2200	2095	-105	4.7727	
2022/4/18	2174	1811	-363	16.6973	8.9761
2022/4/19	1768	1593	-175	9.8982	
2022/4/20	1853	1559	-294	15.8662	
2022/4/21	1587	1623	36	2.2684	

Continued

2022/4/22	1611	1680	69	4.2831
2022/4/23	1606	1589	-17	1.0585
2022/4/24	1529	1948	419	27.4035
2022/4/25	1572	1690	118	7.5064
2022/4/26	1398	1431	33	2.3605
2022/4/27	1551	1336	-215	13.8620
2022/4/28	1455	1346	-109	7.4914
2022/4/29	1574	1457	-117	7.4333

3.2. Holt-Winters 模型预测

Holt-Winters 三参数指数平滑法是 Holt 两参数指数平滑的延伸，其公式涵盖了时间序列的水平、趋势以及季节性构成，它包括加法模型和乘法模型两种，可以修匀含有季节效应的序列。借助指数平滑法，可以对长期趋势、趋势增量和季节变动做出估计，并能适当过滤掉随机波动的影响。因此，根据客运量序列具有季节效应，但没有明显长期趋势的确定性特征，选择使用 Holt-Winters 三参数指数平滑模型来预测序列未来发展。

由因素分解图 2 可以看出，客运量数据趋势附近的波动程度随时间的变化逐渐减小，所以选择使用乘法模型。Holt-Winters 乘法模型的构造如下：

$$\hat{y}_{t+k} = (a_t + r_t k) s_{t+k}$$

式中 a_t 是稳定成分，通过季节因子的修正消除季节变化的影响，为序列的拟合值； r_t 是趋势部分，为序列变化趋势的指数平滑平均数，将当前一期与上期的趋势增量做加权平均，是对趋势增量的估计； s_t 是季节部分，为季节因子的指数平滑平均数，对当前季节因子和上一个相同季节的季节因子的加权平均，进一步改善季节变化对预测值的影响； k 是向后平滑的期数。 a_t 、 r_t 、 s_t 的计算公式如下：

$$a_t = \alpha(y_t/s_{t-\pi}) + (1-\alpha)(a_{t-1} + r_{t-1})$$

$$r_t = \beta(a_t - a_{t-1}) + (1-\beta)r_{t-1}$$

$$s_t = \gamma(y_t/a_t) + (1-\gamma)s_{t-\pi}$$

式中 y_t 是 t 时刻的真实值， α 、 β 、 γ 为平滑系数，取值范围为(0,1)， π 为一个季节周期长度，对于客运量日度数据， $\pi = 7$ 。

以 2020 年 1 月 24 日~2022 年 4 月 15 日的逐日客运量数据为原始数据，利用 Rstudio 软件对原始数据进行拟合，以平均相对误差最小为目标函数，经过多次的迭代求解，最终在 $\alpha = 0.91$ ， $\beta = 0.06$ ， $\gamma = 0.28$ 时，可以得到最优解。 α 值接近于 1，表明近期观测值的权重较大，历史观测值的权重较小。 β 值接近于 0，表明在整个时间序列上，趋势部分的水平和斜率没有发生太大变化。 γ 值在 0~0.5 之间，表明序列近期的观测值受到季节性的影响较小。使用该模型预测的 2022 年 4 月 15 日~2022 年 4 月 29 日共 15 天的秦皇岛车站铁路日客运量，真实值和预测结果如表 5 所示。

由表 5 可以看出，基于 Holt-Winters 乘法模型的铁路日客运量预测值与真实值的最小相对误差为 1.0520%，最大相对误差为 27.7829%，其中 4 月 16 日~4 月 18 日的误差偏大，其他时间的误差适中，平均相对误差为 9.1912%，总体的预测准确度较高。对比前后两个预测模型的预测结果，可以发现，SARIMA 模型的精度略高于 Holt-Winters 模型。

Table 5. Forecast of daily passenger volume of railway stations by Holt-Winters multiplicative model from April 15 to April 29, 2022**表 5.** 2022 年 4 月 15 日~4 月 29 日铁路车站日客运量 Holt-Winters 乘法模型预测

日期	真实值	预测值	误差	相对误差/%	MAPE/%
2022/4/15	1711	1729	18	1.0520	
2022/4/16	1961	1607	-354	18.0520	
2022/4/17	2200	1777	-423	19.2273	
2022/4/18	2174	1570	-604	27.7829	
2022/4/19	1768	1471	-297	16.7986	
2022/4/20	1853	1548	-305	16.4598	
2022/4/21	1587	1554	-33	2.0794	
2022/4/22	1611	1684	73	4.5313	9.1912
2022/4/23	1606	1565	-41	2.5529	
2022/4/24	1529	1731	202	13.2112	
2022/4/25	1572	1529	-43	2.7354	
2022/4/26	1398	1432	34	2.4320	
2022/4/27	1551	1507	-44	2.8369	
2022/4/28	1455	1513	58	3.9863	
2022/4/29	1574	1639	65	4.1296	

3.3. 组合模型预测

上述研究表明, 通过建立的两个单一的时间序列模型预测均可以进行铁路客运量的短期预测, 但由于不同模型的特点和局限性, 会使得各个预测结果有显著差异。如果采用组合预测模型, 就可以充分利用每个单一模型提取到的序列信息。在构建组合模型时, 选择合适的权值, 可以有效提高模型预测精度, 常用的权值计算方法有等权值法、方差倒数法、方差-协方差法、最优加权法等。由于方差倒数法易于操作, 且效果好, 所以本文使用该方法计算单项模型的权重。一般情况下, 每种单一预测模型的预测精度各不相同, 而误差平方和作为判断模型预测精度的重要指标, 它越大, 说明该单一模型的精度越低, 其在组合模型中的重要性就越低, 应在组合预测中赋予较小的加权系数。

假设观测对象为 X , 对其有 m 种预测方法。 X_t 为真实值序列中第 t 期的实际观测值, x_{it} 为第 i 种预测方法对第 t 期的预测值, w_i 为第 i 个单项预测模型在组合预测模型中的权重系数, 则组合预测模型为

$$\hat{X}_t = \sum_{i=1}^m w_i x_{it} = w_1 x_{1t} + w_2 x_{2t} + \cdots + w_m x_{mt}, \quad t = 1, \cdots, N$$

方差倒数法的权重系数计算公式为

$$w_i = \frac{1}{\sum_{t=1}^N (X_t - x_{it})^2} \bigg/ \left(\sum_{i=1}^m \frac{1}{\sum_{t=1}^N (X_t - x_{it})^2} \right)$$

式中, $w_i \geq 0$, 且 $\sum_{i=1}^m w_i = 1$ 。

通过权重公式, 分别计算出 SARIMA(1,0,5)(5,1,0)₇ 模型和 Holt-Winters 乘法模型的权重, 结果为 $w_1 = 0.61$, $w_2 = 0.39$, 则铁路客运量的组合预测模型为

$$\hat{X}_t = w_1 x_{1t} + w_2 x_{2t} = 0.61x_{1t} + 0.39x_{2t}$$

由该公式计算出组合预测模型的预测结果如表 6 所示。

Table 6. Forecast of daily passenger volume of railway stations by combining model from April 15 to April 29, 2022
表 6. 2022 年 4 月 15 日~4 月 29 日铁路车站日客运量组合模型预测

日期	真实值	预测值	误差	相对误差/%	MAPE/%
2022/4/15	1711	1749	38	2.2285	
2022/4/16	1961	1694	-267	13.6038	
2022/4/17	2200	1971	-229	10.4100	
2022/4/18	2174	1717	-457	21.0207	
2022/4/19	1768	1545	-223	12.5894	
2022/4/20	1853	1555	-298	16.0977	
2022/4/21	1587	1596	9	0.5728	
2022/4/22	1611	1682	71	4.3799	8.3876
2022/4/23	1606	1580	-26	1.6413	
2022/4/24	1529	1863	334	21.8685	
2022/4/25	1572	1627	55	3.5121	
2022/4/26	1398	1431	33	2.3884	
2022/4/27	1551	1403	-148	9.5622	
2022/4/28	1455	1411	-44	3.0151	
2022/4/29	1574	1528	-46	2.9238	

从表 6 中可以看出, 组合模型的最大相对误差、最小相对误差和 MAPE 都比单一模型更小。和 SARIMA 模型相比, 虽然组合模型部分预测值的相对误差更大, 但是后半段组合模型的预测精度有明显提高, MAPE 减小了 0.5885%。和 Holt-Winters 模型相比, 除了 4 个预测值的相对误差更大, 其余预测值的预测精度均为更优, MAPE 减小了 0.8036%。综上, 与单一模型相比, 由方差倒数法得到的组合模型的预测效果最好。

4. 样本量对预测效果的影响

考虑到铁路客运量数据包含的信息量大, 样本量过多, 会导致数据提取时间过长的问题, 以下选择用不同的样本量对客运量进行预测, 分析样本量对预测效果的影响。对 2020 年 1 月 26 日~2022 年 4 月 15 日 810 个样本量的秦皇岛站铁路客运量数据, 以 30 日为间隔, 设置 26 个梯度样本量(60、90、120、150、180、210、240、270、300、330、360、390、420、450、480、510、540、570、600、630、660、690、720、750、780、810)。即从 2022 年 2 月 14 日的数据开始往前递推, 用不同梯度的样本量, 分别使用 SARIMA 模型、Holt-Winters 乘法模型和组合模型预测 2022 年 4 月 15 日~2022 年 4 月 29 日的客运量。

如图 5 所示, 当样本量为 60~480 日时, 随着样本量的增加, 三个模型的 MAPE 的波动都非常大, 且没有明显趋势。MAPE 波动最大的为 Holt-Winters 模型, 变化范围为 9.0233%~36.3749%。MAPE 波动最小的为组合模型, 变化范围为 8.1713%~17.8342%。当样本量大于 480 日时, 三个模型的 MAPE 都趋于稳定, 基本维持在 10%以下, 且组合模型的 MAPE 一直保持为三者中最低。同时可以看出, 在 26 个梯度样本量中, 有 20 个梯度样本量下为组合模型的 MAPE 最小, 约占比 77%, 说明将 SARIMA 模型和

Holt-Winters 模型组合起来具有合理性, 可有效提高预测准确度。

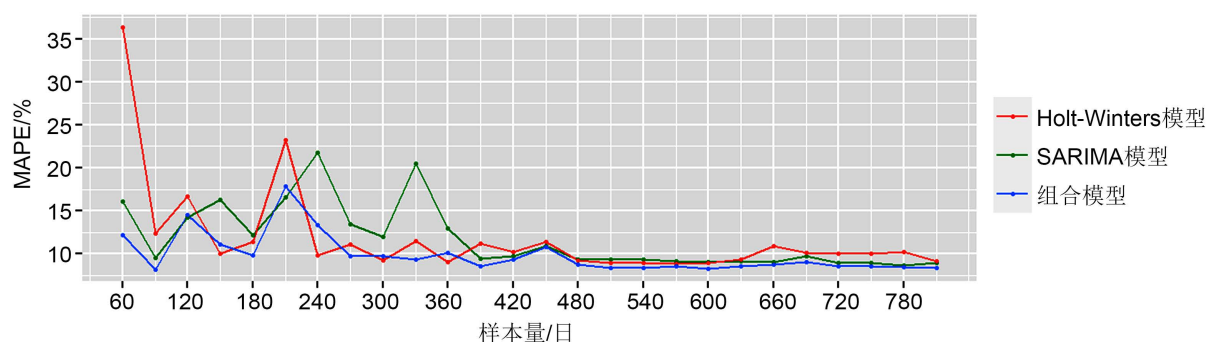


Figure 5. Comparison between sample size and MAPE in three predictive models

图 5. 三种预测模型中样本量与 MAPE 的对比关系

由此可见, 样本量对预测的准确度有明显贡献, 当样本量达到 480 日以上时, 三个模型基本都可以达到 90% 的预测准确度, 并且使用组合模型进行预测, 准确度最高。

5. 结论

本文结合研究背景和现状, 提出了铁路客运量预测存在的问题。随着疫情的出现, 秦皇岛站铁路客运量变化巨大, 使用过去的预测模型显然不合理, 所以针对后疫情时期的铁路客运量建立了新的时间序列模型。由于客运量序列具有明显的季节性和周期性, 所以本研究建立了 SARIMA 模型和 Holt-Winters 模型。经过检验, 两个模型预测出的结果都很接近真实值。且当样本量足够大时, 若进一步使用方差倒数法, 将以上模型组合起来, 通过让正负误差相互抵消, 还可以再提高预测准确度。

由此可见, 对于客运量序列的预测, 进行线性模型和非线性模型的组合, 可以取得良好的预测效果。进一步的研究会考虑在此背景下, 丰富单一模型的种类, 让信息提取更加充分, 为制定客运量的预测方法提供新思路。

参考文献

- [1] 王钰珂, 肖翔, 代庆会. 疫情之下铁路行业的现状与未来发展方向[J]. 中国市场, 2021(6): 76-78.
- [2] 蔡倒录, 王伯礼. 基于 ARMA 模型的新疆铁路客运量预测[J]. 经济研究导刊, 2022(33): 51-53.
- [3] 赵鹏, 李璐. 基于 ARIMA 模型的城市轨道交通进站量预测研究[J]. 重庆交通大学学报: 自然科学版, 2020, 39(1): 40-44.
- [4] 段然, 庞建华, 张良钧. 基于 SARIMA 模型的铁路站点客流量预测研究[J]. 数学的实践与认识, 2019, 49(9): 1-10.
- [5] 王小凡, 朱永强. 铁路客运量预测模型对比研究[J]. 黑龙江交通科技, 2019, 42(6): 184-189.
- [6] 桂文毅. 基于灰色线性回归模型的哈尔滨铁路枢纽客运量预测研究[J]. 中国铁路, 2018(6): 22-27.
- [7] Gliovi, N., Milenkovi, M., Bojovi, N., et al. (2016) A Hybrid Model for Forecasting the Volume of Passenger Flows on Serbian Railways. *Operational Research*, **16**, 271-285. <https://doi.org/10.1007/s12351-015-0198-5>
- [8] Li, C. (2019) Combined Forecasting of Civil Aviation Passenger Volume Based on ARIMA-REGRESSION. *International Journal of System Assurance Engineering and Management*, **10**, 945-952. <https://doi.org/10.1007/s13198-019-00825-6>
- [9] 张蕾. 基于组合模型的民航客运量预测研究[D]: [硕士学位论文]. 大连: 辽宁师范大学, 2019.
- [10] 张玺君, 王晨辉, 陶冶, 等. 基于 SARIMA-GA-Elman 组合模型的短时交通流量预测系统[J]. 兰州理工大学学报, 2022, 48(5): 107-113.