

南京市商品房价格预测

——基于ARIMA模型和SARIMA模型比较分析

陈 蕾

云南财经大学统计与数学学院, 云南 昆明

收稿日期: 2022年3月13日; 录用日期: 2022年3月30日; 发布日期: 2022年4月12日

摘 要

以南京市商品房房价为研究对象, 选取2014年1月至2021年10月为观测时间, 构建ARIMA和SARIMA时间序列模型进行房价预测。结果表明, ARIMA(1, 1, 4)模型是拟合房价的最优模型, 因此对南京市商品房未来一年的房价做了短期预测, 并结合预测结果和南京市房地产业发展状况提出相关建议。

关键词

南京市商品房, 房价预测, ARIMA模型, SARIMA模型

Forecast of Commercial Housing Prices in Nanjing

—A Comparative Analysis Based on ARIMA Model and SARIMA Model

Lei Chen

School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming Yunnan

Received: Mar. 13th, 2022; accepted: Mar. 30th, 2022; published: Apr. 12th, 2022

Abstract

Taking the price of commercial housing in Nanjing as the research object, the ARIMA and SARIMA time series models are constructed to predict the housing price from January 2014 to October 2021. The results show that ARIMA(1, 1, 4) model is the best model to fit housing prices. So this research made a short-term prediction of Nanjing commercial housing prices in the coming year and put forward some suggestions combined with the forecast results and the development status of the real estate industry in Nanjing.

Keywords

Nanjing Commercial Housing, House Price Forecast, ARIMA Model, SARIMA Model

Copyright © 2022 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

房地产业是我国国民经济体系中的重要组成部分, 作为对中国经济贡献占比较大的第三新兴产业正在不断发展, 房屋居住问题成为了社会主要的民生问题, 为稳定房地产市场供需平衡, 各地纷纷出台了调控政策, 秉持着“坚持住房不炒”的观念, 创建购租并举、多主体供给的住房制度, 使每位公民皆有所可居。虽然新冠肺炎疫情的爆发对国民经济产生了不可忽视的冲击, 但房地产业作为国民经济体系的支柱, 依然能够拉动其他产业链的发展, 养老业、旅游业与地产业相辅相成。但是在房地产业投资收益比住房收益更看好的情况下, 地产企业很容易陷入融资难的困境, 即产生投资泡沫。伴随着不少城市房价居高不下的状况, 大型住房企业认为有利可图, 开始踏入长租公寓市场, 租赁市场供需不均衡, 小企业逐渐边缘化, 不利于房地产业的长久发展。基于此背景, 房地产业发展形势较为严峻, 根据历史状况预测未来房价有利于消费者提前预判, 在合适的时候达到自买自住的目的, 避免加重扰乱房地产业的市场秩序。也有助于政府灵活调控, 进一步稳定房价, 促进产业平衡发展。

对已有研究进行梳理, 付淇伟等人(2021)建立了 ARIMA(1, 2, 0)模型对衡阳市商品房房价进行了短期预测[1]。郑永坤等人(2018)建立 ARIMA(1, 1, 0)模型对广州市和深圳市二手房房价进行滚动预测, 将当月的实际房价加入变量中来预测次月的房价, 并引入 RMSE 判断预测精度, 相较于传统的时间序列预测模型, 该方法预测结果持续性更强[2]。孟盈竹等人(2021)建立 AR(2)模型对沈阳市商品房房价进行短期预测, 研究发现房价的周期波动较明显[3]。聂淑媛(2016)考虑到房价波动具有显著的季节趋势, 因此结合 X-12-ARIMA 模型和 AR(2)-GARCH(1, 1)模型对北京市商品房房价进行了短期预测, 模型系数的显著性结果表明外部因素和预测对象的滞后项对预测对象的影响较大且持续性较强, 体现了房价波动的不确定性[4]。孙悦(2020)为了预测大连市商品房房价, 分别建立了 ARIMA 模型、指数平滑模型、长短期记忆网络模型, 比较发现 ARIMA(0, 2, 1)模型拟合趋势准确, 但具体误差较大, 而 LSTM 网络模型的均方根误差较小, 预测准确度最高[5]。由此可见, 学者们大多通过时间序列模型对城市房价进行了预测, 但主要基于 Stata 软件和 Eviews 软件的实现, 利用 R Studio 软件建立时间序列模型的研究较少, 因此本文基于 R Studio 软件分别拟合了 ARIMA 模型和 SARIMA 模型, 并选择了较为贴合历史房价变化的最优模型 ARIMA(1, 1, 4)对南京市商品房未来一年的房价进行了短期预测, 以为市民和政府提供购房和管控的指导建议。

2. 模型建立

2.1. ARIMA 模型介绍

自回归移动平均 ARIMA 模型是一种时间序列的预测模型, 然而现实中的大部分数据显然是非平稳的, 需要对非平稳数据进行处理, 因此差分自回归移动平均模型 ARIMA(p, d, q)模型旨在通过差分运算将非平稳时间序列转化为差分平稳序列, d 为时间序列成为平稳序列时所做的差分次数。该模型然后利

用因变量的滞后值和随机误差项建立模型，以达到对未来值预测的目的，具体的数学表达式如下：

$$\begin{cases} \Phi(B)\Delta^d x_t = \theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t, \varepsilon_s) = 0, s \neq t \\ E(x_s, \varepsilon_t) = 0, \forall s < t \end{cases}$$

上式中， B 为延迟算子， $\Delta^d = (1-B)^d$ ； $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ 为自回归系数多项式； $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ 为移动平均系数多项式； ε_t 表示均值为零的白噪声序列[6]。

2.2. 数据分析及处理

本文选取南京市商品房 2014 年 1 月至 2021 年 10 月的房价数据进行研究，数据来源于安居客网站。考虑到模型预测具有现实意义，研究对象剔除了南京市周边样本，仅为老城八区(包括江宁区、浦口区、栖霞区、雨花区、建邺区、秦淮区、玄武区和鼓楼区)的平均房价。

对南京市商品房房价绘制时序图，如图 1 所示，过去 118 个月的商品房房价呈现了明显的上升趋势，初步判断原序列不平稳。对原序列做一阶差分处理，一阶差分时序图(如图 2 所示)已消除增长趋势，初步认为一阶差分序列平稳。对一阶差分序列进行单位根检验，无趋势无截距项、无趋势有截距项和有截距项有趋势三种类型下的滞后 2 阶的 p 值均小于 5% 临界水平，认为一阶差分序列是平稳的；对差分序列进行纯随机性检验，检验结果显示各阶数下 LB 统计量的 p 值均小于 5% 临界水平，因此一阶差分序列为非白噪声序列。一阶差分序列通过了单位根检验和纯随机性检验，继续拟合模型是有统计意义的。

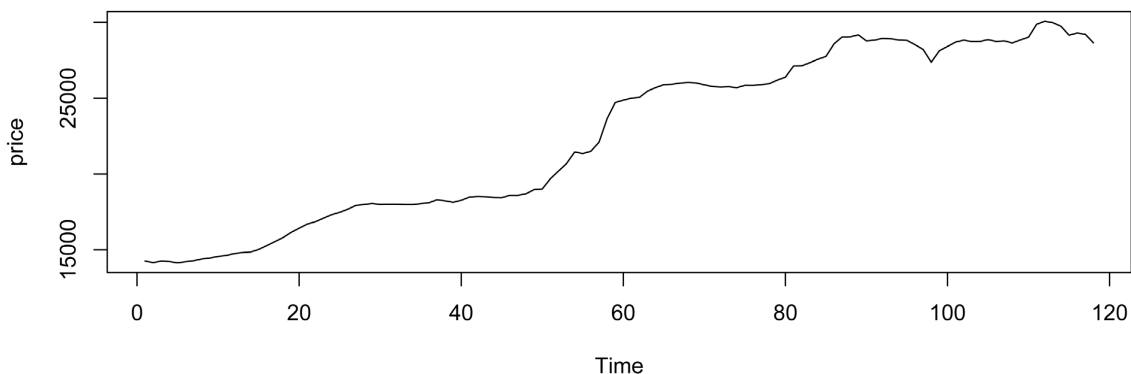


Figure 1. Sequence diagram of the original sequence

图 1. 原序列时序图

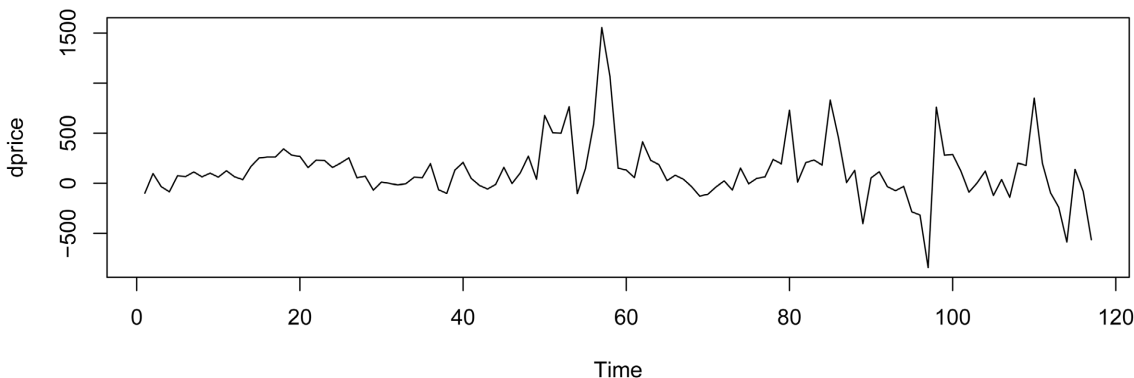


Figure 2. Sequence diagram of the first order difference sequence

图 2. 一阶差分序列时序图

2.3. 模型识别与拟合

为确定模型的阶数，对一阶差分序列进行自相关和偏自相关检验，如图 3 和图 4 所示，偏自相关系数 1 阶截尾，自相关系数无法观察出几阶截尾或拖尾，初步建立 $ARIMA(1, 1, 0)$ 、 $ARIMA(1, 1, 1)$ 、 $ARIMA(1, 1, 2)$ 、 $ARIMA(1, 1, 3)$ 、 $ARIMA(1, 1, 4)$ 、 $ARIMA(1, 1, 5)$ 模型来进行拟合。观察偏自相关图，第四期、第八期、第十八期接近两倍标准差，初步判断可能存在季节效应，因此引入月度 SARIMA 模型进行拟合。

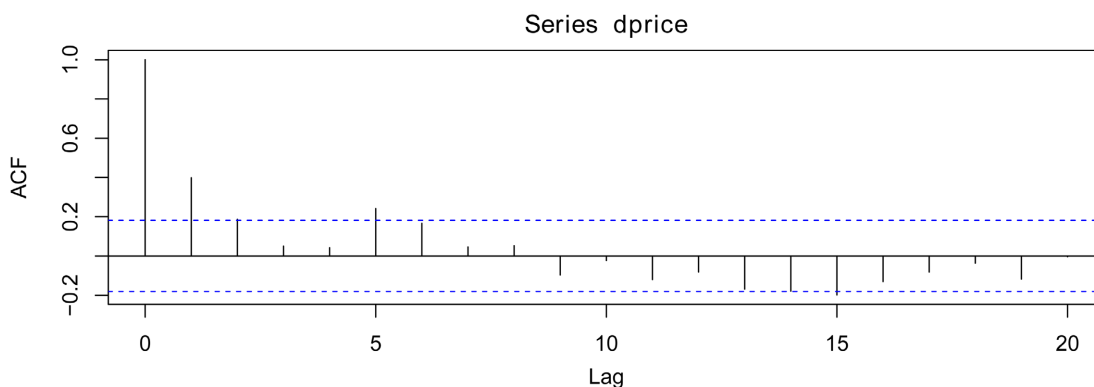


Figure 3. Autocorrelogram of the first order difference sequence

图 3. 一阶差分序列自相关图

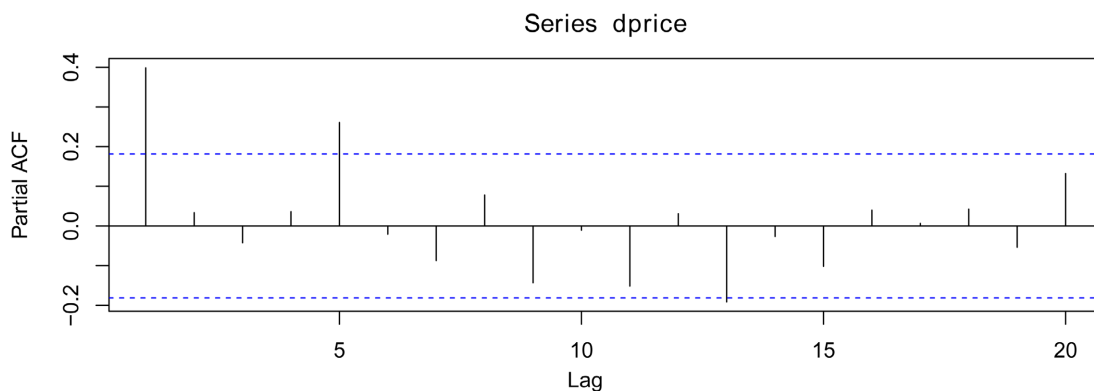


Figure 4. Partial autocorrelation of the first order difference sequence

图 4. 一阶差分序列偏自相关图

对一阶差分序列进行了 SARIMA 拟合，结果表明 $ARIMA(1, 0, 0)(1, 1, 0)$ [12] 通过了白噪声检验(如表 1 所示)，但是模型显著性检验(如图 5 所示)中第二期的值低于临界水平，未通过模型显著性检验。根据残差的自相关图可知，残差的滞前一阶和二阶均具有趋势性，大于两倍标准误，不是随机波动的，因此不选择 $ARIMA(1, 0, 0)(1, 1, 0)$ [12] 模型作为拟合模型。

Table 1. White noise test of SARIMA model

表 1. SARIMA 模型白噪声检验

指标名称	延迟 6 阶	延迟 12 阶
X-squared	473.95	788.52
p-value	<2.2e-16	<2.2e-16

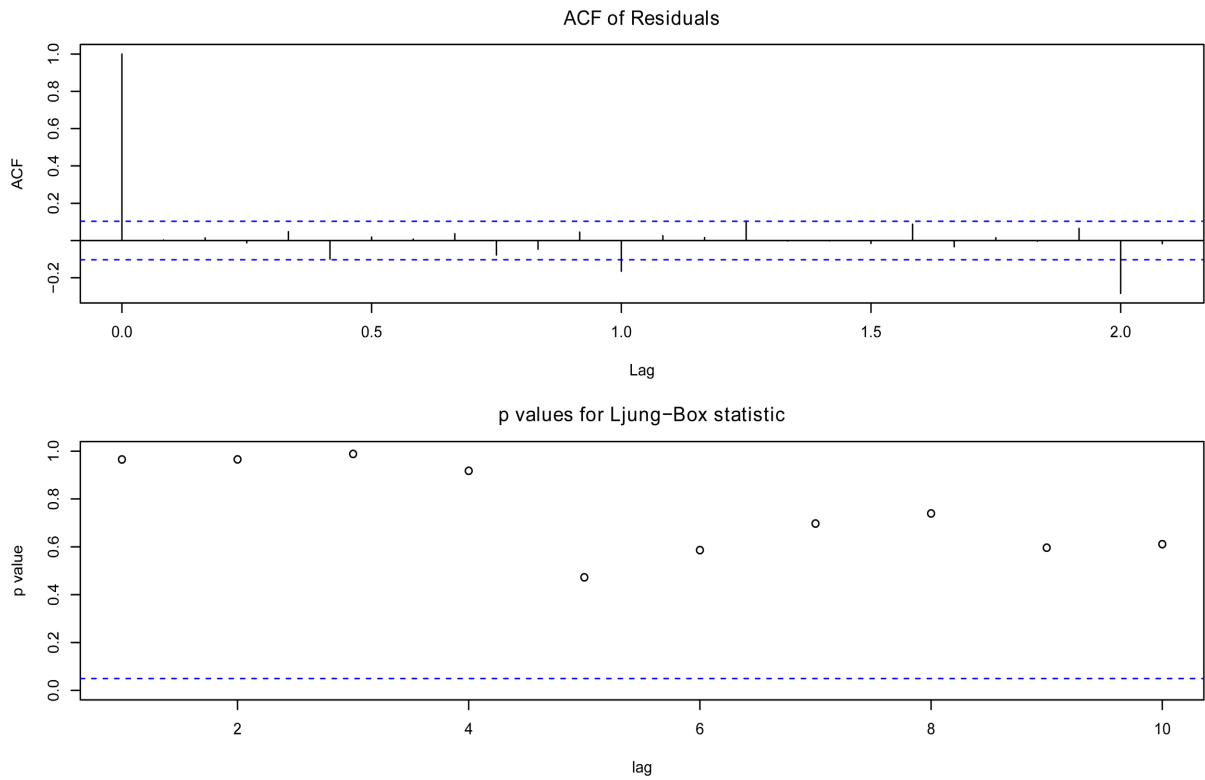


Figure 5. Significance test of SARIMA model

图 5. SARIMA 模型的显著性检验

ARIMA 模型拟合效果如表 2 所示，根据 AIC 信息准则最小化原则确定最优模型为 ARIMA(1, 1, 4) 模型。预测模型的方程表达式为：

$$x_t = 0.7603x_{t-1} - 0.3529\varepsilon_{t-1} - 0.0932\varepsilon_{t-2} - 0.1958\varepsilon_{t-3} + 0.2567\varepsilon_{t-4} + 192.1371$$

Table 2. Fitting effect in ARIMA models

表 2. ARIMA 模型拟合效果

拟合模型	AIC 信息准则
ARIMA(1, 1, 0)	1323.45
ARIMA(1, 1, 1)	1323.28
ARIMA(1, 1, 2)	1322.12
ARIMA(1, 1, 3)	1323.34
ARIMA(1, 1, 4)	1321.64
ARIMA(1, 1, 5)	1323.56

对 ARIMA(1, 1, 4)进行模型显著性检验，如图 6 所示，各阶延迟下白噪声检验统计量的 p 值都显著大于 5%临界水平，拟合模型的残差序列为白噪声序列，且模型残差均在两倍标准误内，即该拟合模型显著。

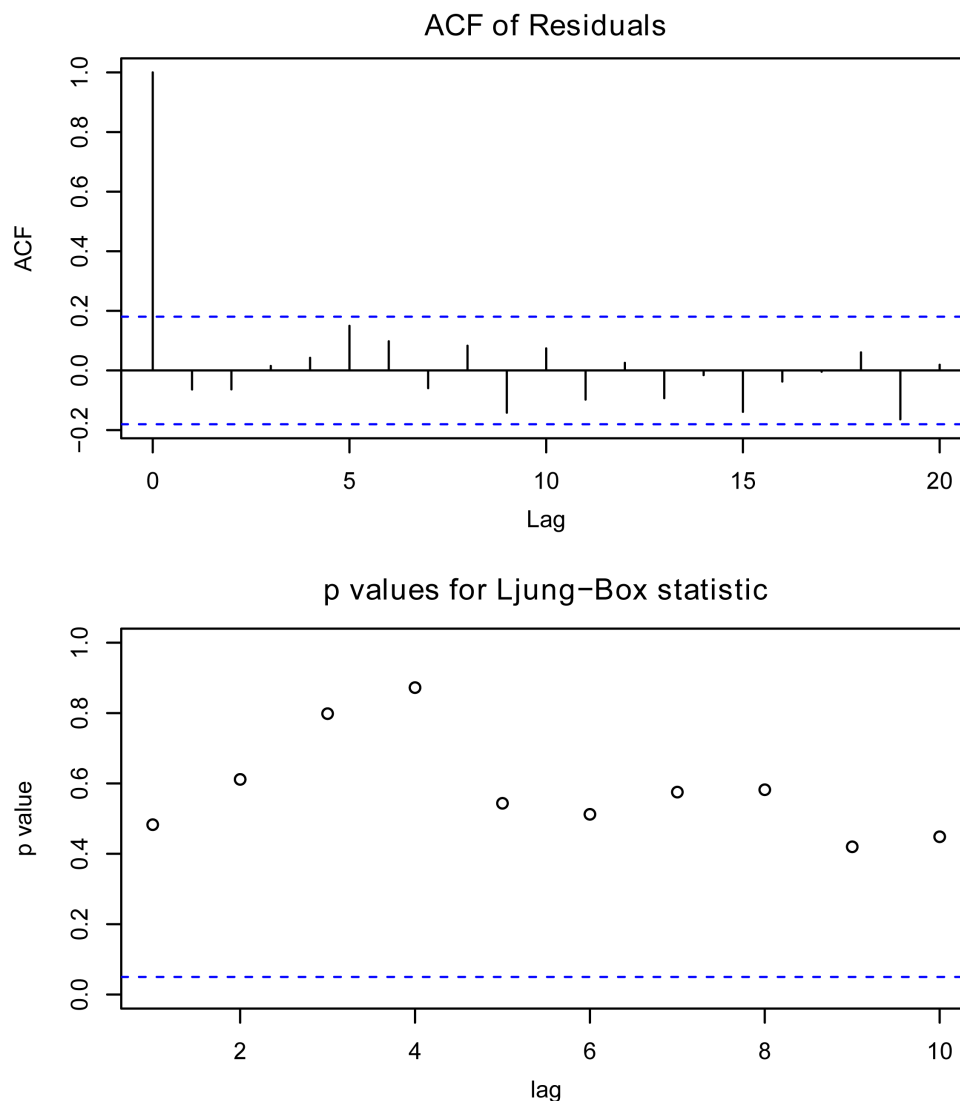


Figure 6. Significance test of ARIMA(1, 1, 4) model
图 6. ARIMA(1, 1, 4)模型的显著性检验

2.4. 模型预测

选取了最近一年半的样本值作为观测值，对其进行了预测，篇幅所限，仅展示 2021 年的房价预测值，如表 3 所示，长期无漂移项的预测值后期涨势放缓，不符合原序列的发展趋势，而有漂移项的预测值呈显著的递增趋势，且有漂移项的相对误差明显小于无漂移项的相对误差，因此有漂移项模型的预测效果明显优于无漂移项模型的预测效果。预测模型得到的预测值和实际值非常接近，平均相对误差为 1.0270%，最大相对误差为 4.1171%，最小相对误差为 0.3870%，该结果说明 ARIMA(1, 1, 4) 作为预测模型具有较高的预测精度。因此选择有漂移项 ARIMA(1, 1, 4)模型对南京市商品房价格进行预测，预测结果如表 4 所示，从模型预测结果上看，预测的房价能较好地反应南京市商品房价格的波动情况，显然南京市政府重视对新冠肺炎疫情传播的管控，目前来看，重大突发事件并没有对南京市房价产生剧烈影响，因此本研究对房价进行短期预测的数据对居民和政府的购房和调控房价行为均具有参考意义。

Table 3. Comparison of predicted values of ARIMA(1, 1, 4) model with or without drift term
表 3. ARIMA(1, 1, 4)模型有无漂移项的预测值比较

Month	test	Forecast (无漂移项)	相对误差	Forecast (有漂移项)	相对误差
2021-01	33,558.13	32,844.58	0.021263	33,805.43	0.007369
2021-02	33,863.38	32,883.38	0.02894	33,994.42	0.003870
2021-03	34,013.63	32,917.65	0.032222	34,184.17	0.005014
2021-04	34,157.88	32,947.92	0.035422	34,374.50	0.006342
2021-05	34,168.75	32,974.65	0.034947	34,565.26	0.011604
2021-06	34,281.38	32,998.26	0.037429	34,756.34	0.013855
2021-07	34,474	33,019.11	0.042203	34,947.68	0.013740
2021-08	34,550.25	33,037.53	0.043783	35,139.21	0.017046
2021-09	34,384.63	33,053.79	0.038704	35,330.89	0.027520
2021-10	34,118	33,068.16	0.030771	35,522.68	0.041171

Table 4. Forecast of Nanjing commercial housing price from November 2021 to October 2022
表 4. 南京市商品房 2021 年 11 月~2022 年 10 月的房价预测

Month	Forecast
2021-11	34,111.47
2021-12	34,194.88
2022-01	34,278.42
2022-02	34,311.86
2022-03	34,377.04
2022-04	34,466.49
2022-05	34,574.47
2022-06	34,696.62
2022-07	34,829.59
2022-08	34,970.84
2022-09	35,118.41
2022-10	35,270.82

3. 结论与建议

3.1. 结论

本文对南京市商品房房价的预测问题建立了 ARIMA 模型和 SARIMA 模型进行比较分析。原序列南京市商品房月均销售额呈现非平稳性，通过一阶差分运算后呈现平稳性，经过纯随机性检验认为差分平稳序列为非白噪声序列，因此建立模型具有统计学意义。观察差分序列的自相关图和偏自相关图，可能存在季节效应，因此拟合了 ARIMA 模型和含有季节趋势的 SARIMA 模型。比较拟合模型的参数估计值

和两倍标准差的值选择了有效模型, 根据 AIC 信息准则最小化原理和模型显著性检验确定最优拟合模型为 ARIMA(1, 1, 4)模型。为了确定拟合模型的预测精度, 对样本值中最近一年半的房价进行了预测, 将预测值与实际值作比较, 选择了预测值和预测趋势更好的有漂移项的拟合模型为预测模型。考虑到现阶段, 我国经济环境仍受疫情反复波动的影响, 不确定未来疫情的影响效果有多大, 因此本文只对未来一年的房价进行了短期预测, 预测结果具有时效性。

3.2. 建议

1) 采取长短期相结合的政策, 短期避免房价波动对经济产生较大影响, 长期若房价依然居高不下, 可以实施租购并举的措施, 提高土地使用率, 加快提供租赁住宅, 从宏观层面看, 房价波动需要保持一个平稳的状态。

2) 鼓励创新创业, 南京作为全国重点院校集聚的城市, 每年都会有大批学子涌入南京, 市政府应对长期留京的外来学子增大租房购房优惠, 减轻年轻人的就业负担, 更有利于拉动南京城市经济。合理引导居民消费, 南京市政府应加大力度监管限购令的实施, 引导居民正确的消费观念, 避免出现热点楼盘哄抬房价的情形。

3) 加大产业资金监管力度, 房价的不断上涨与涌入地产市场的投机性资金息息相关, 过高的房价将会抑制即将购房者的消费, 严重阻碍了房地产市场发挥财富效应, 因此就开发商层面而言, 应严格规范资金与地产的对接渠道, 就个人层面而言, 严查购房首付资金的来源, 严防个人将消费贷款、经营性贷款等不良资金用于购房。

参考文献

- [1] 付淇伟, 易艳春, 张诗佳, 丁嘉, 汤敏岚. 基于 ARIMA 模型的衡阳市房价预测分析[J]. 科技创新与应用, 2021, 11(31): 47-50.
- [2] 郑永坤, 刘春. 基于 ARIMA 模型的二手房价格预测[J]. 计算机与现代化, 2018(4): 122-126.
- [3] 孟盈竹, 孙胜男. 基于 ARIMA 时间序列模型的房地产价格预测——以沈阳市为例[J]. 内江科技, 2021, 42(5): 65-74.
- [4] 聂淑媛. 基于 X-12-ARIMA 和 AR-GARCH 模型的房价波动研究[J]. 河南师范大学学报(自然科学版), 2016, 44(4): 39-44.
- [5] 孙悦. 基于时间序列分析的大连市房价分析与预测[D]: [硕士学位论文]. 大连: 大连理工大学, 2020.
- [6] 王燕. 时间序列分析——基于 R [M]. 第 2 版. 北京: 中国人民大学出版社, 2020.