Statistics and Application 统计学与应用, **2021, 10(4), 583-592**
Published Online August 2021 in Hans. http://www.hanspub.org/journal/sa
https://doi.org/10.12677/sa.2021.104060

基于时间序列的天津市房价的建模与预测

李正才，牛开兴，伍彩云，苏睿楠，安子兰，白晓东

大连民族大学，辽宁 大连
Email: 1561069175@qq.com

收稿日期：2021年6月28日；录用日期：2021年7月21日；发布日期：2021年7月28日

摘 要

我国房地产近几年发展势头强盛，虽然经历了2020年的新冠疫情，但全国房价从整体上来讲还是呈现出增长的趋势。可见，国内居民对住房需求势头不减。本文利用了2012年1月到2021年1月近十年的月度房价数据，采用时间序列模型来对天津市的房价未来一段时间的价格指数进行了预测，结果显示在未来一段时间内天津市的房地产市场价格将保持一种下降的趋势，且房地产在短时间内回暖的可能性较小。

关键词

时间序列模型，房价指数，预测

Modeling and Forecasting of Housing Prices in Tianjin Based on Time Series

Zhengcai Li, Kaixin Niu, Caiyun Wu, Ruinan Su, Zilan An, Xiaodong Bai

Dalian Minzu University, Dalian Liaoning
Email: 1561069175@qq.com

Received: Jun. 28th, 2021; accepted: Jul. 21st, 2021; published: Jul. 28th, 2021

Abstract

China's real estate industry has enjoyed a strong development momentum in recent years. Despite the COVID-19 outbreak in 2020, the national housing price has shown an overall trend of growth. Visible, domestic residents for housing demand momentum is not reduced. In this paper, using the in January 2012 to January 2021, after nearly a decade of monthly house price data, by using the time series model to Tianjin price index of house prices in the future for a period of time, the re-

文章引用: 李正才, 牛开兴, 伍彩云, 苏睿楠, 安子兰, 白晓东. 基于时间序列的天津市房价的建模与预测[J]. 统计学与应用, 2021, 10(4): 583-592. DOI: 10.12677/sa.2021.104060

sults show that in the future, Tianjin real estate market prices over a period of time will maintain a downward trend, and the real estate is less likely to recover in a short time.

Keywords

Time Series Model, Housing Price Index, Forecast

1. 引言

房地产市场是反映一个国家经济状况的具体表现，是我国国民经济发展的支柱行业，是人民安居乐业的载体。房地产市场长期的低迷不利于刺激发展过程中投资者的热情，不利于国民经济的健康稳定。房价也是影响居民购买力、理财、教育文化和制定决策的重要考虑因素，但近年来的房价波动，使得人们在是否购房上面犹豫不决。本文意在利用 2012 年 1 月~2021 年 1 月的月度数据，采用 ARIMA 模型对天津市新建商品房未来一段时间的价格指数进行预测，希望这个结果能够对天津市想要买房的人员有一个参考作用，也希望能够帮助到天津市政府，对改良房价提供帮助[1]。

2. 数据来源

附录 1 为所搜集到的 2012 年 1 月到 2021 年 1 月天津市的房价数据。

3. 研究方法

1) ARIMA 模型介绍

ARIMA 模型(英语：Autoregressive Integrated Moving Average model)，差分整合移动平均自回归模型，又称整合移动平均自回归模型(移动也可称作滑动)，是时间序列预测分析方法之一。ARIMA(p, d, q)中，AR 是“自回归”，p 为自回归项数；MA 为“滑动平均”，q 为滑动平均项数，d 为使之成为平稳序列所做的差分次数(阶数)。是美国统计学家 G.E.P.Box 和英国统计学家 G.M.Jenkins 在前人时间序列研究的基础上新创立的一种对于时间模型识别、估计、检验及预测的分析方法[2]。

一般来说，很多时间序列本身是非平稳的，在进行适当差分之后就会变成一个平稳时间序列，对差分平稳时间序列可以使用 ARIMA 模型进行拟合，ARIMA(p, d, q)模型的结构如下：

$$\Phi(B)\nabla^d x_t = \theta(B)\varepsilon_t$$

其中，ε_t 为均值为零，方差为 σ_ε 的白噪声，且 $E(x_s\varepsilon_t)=0$，$\forall s<t$；

$\nabla^d=(1-B)^d$；$\Phi(B)=1-\Phi_1(B)-\cdots-\Phi_p B^p$ 为平稳可逆的 ARMA(p, q)模型的自回归系数多项式；$\theta(B)=1-\theta_1 B-\cdots-\theta_q B^p$ 为平稳可逆的 ARMA(p, q)模型的移动平滑系数多项式。

2) 差分运算：

为了将非平稳时间序列转化为平稳时间序列，我们需要对该序列做差分运算。对于一个序列 $\{x_t\}$ 来说，相邻两时刻序列值之差称为一阶差分表示为 $\nabla x_t = x_t - x_{t-1}$，对一阶向后差分一阶所得序列 $\{\nabla x_t\}$ 在进行一阶向后差分运算就得到 2 阶向后差分，记为 $\nabla^2 x_t$，即 $\nabla^2 x_t=\nabla x_t-\nabla x_{t-1}$，本文中只用到 2 阶差分[3]。

4. 实际序列模型建立

天津市住房指数未来走势实证分析

1) 序列平稳性检验

本文数据来源于安居客(https://tj.fang.anjuke.com/fangjia/)，数据内容为2012年1月~2021年1月的月度时间序列数据，数据内容是天津市新建商品住房的价格，这里我们首先对这个数据进行时间序列分析，先画出其时序图，如下图所示：

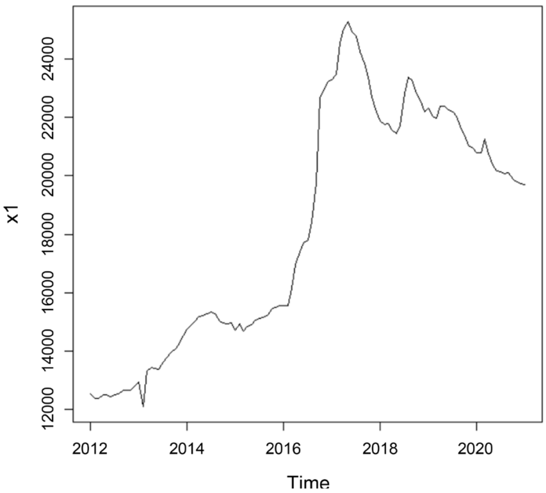

Figure 1. Sequence diagram of stationary test
图1. 平稳检验时序图

图1为2012年~2020年天津市房价时序图，从图1可以看出，自2012年~2020年天津市房价前期总体呈上升趋势，在2017年之后有一个下降，且下降趋势过于明显且降幅过大，所以它很有可能不是一个平稳时间序列。

图2，图3为序列的自相关图和偏自相关图，从图2我们可以看到，自相关图具有很强的拖尾性，并且自相关系数长期大于0，表现出很强的自相关性。图3可以看出偏自相关1阶截尾。

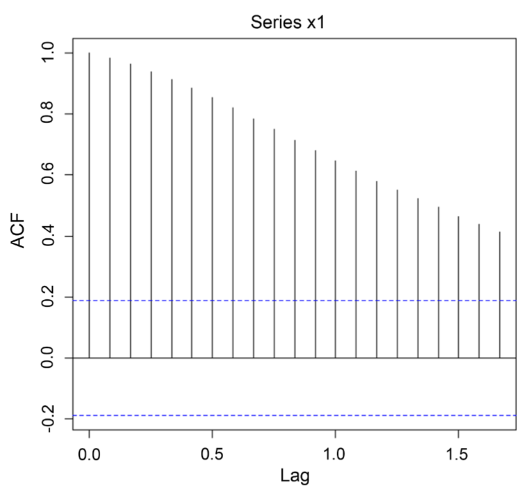

Figure 2. Autocorrelation graph
图2. 自相关图

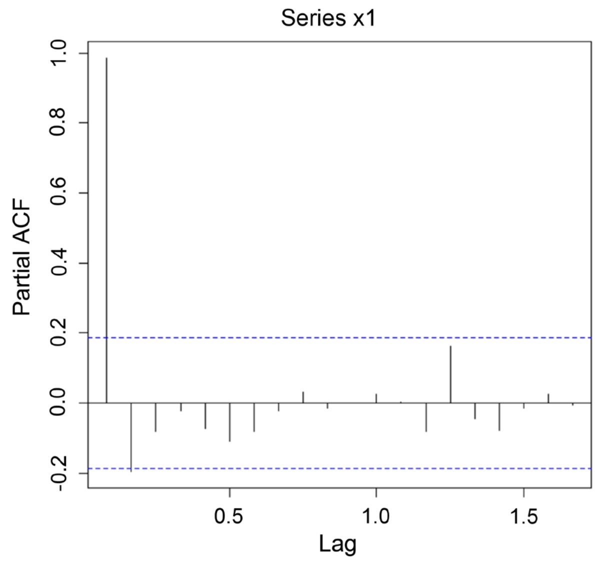

Figure 3. Partial autocorrelation graph
图 3. 偏自相关图

由于该序列为非平稳的时间序列，所以我们对其进行差分处理，依据数据最优以及准确性原则，我们对其进行 2 阶差分。图 4 为差分之后的时序图，从图 1 可以看出从 2012 年 1 月到 2021 年 1 月天津市的新建商品房的销售价格的月度价格始终围绕在 21000 附近随即波动，没有什么明显的周期线特征和趋势，所以该序列可以基本视为平稳时间序列。

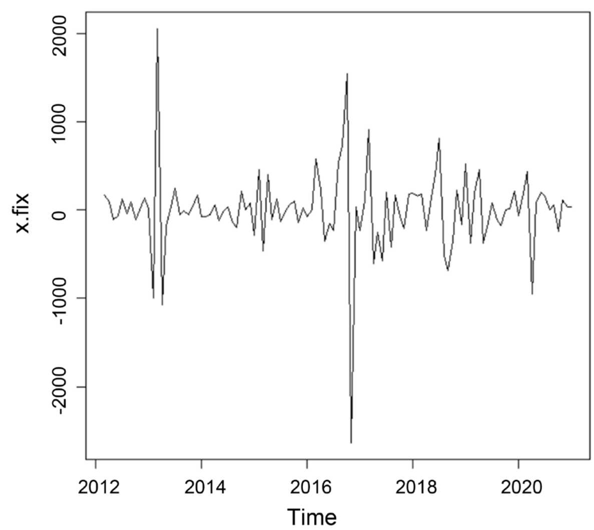

Figure 4. 2 Sequence diagram after order difference
图 4. 2 阶差分之后的时序图

但为了使结论更为严谨，我们做出其自相关图进行进一步分析，图 5 为销售价格的时间序列自相关图。从平稳序列的自相关图中可以看出，在新建商品住房的价格指数 ACF 在延迟 2 阶之后，全部衰减到 2 倍标准差之内，且自相关图在延迟 2 阶之后具有明显的截尾特征，所以可以认为该序列为非常强的平稳序列。下面我们在所给的自相关图的基础上做出它的偏自相关图，图 6 所示，来进一步确认该序列的截尾关图如下所示：

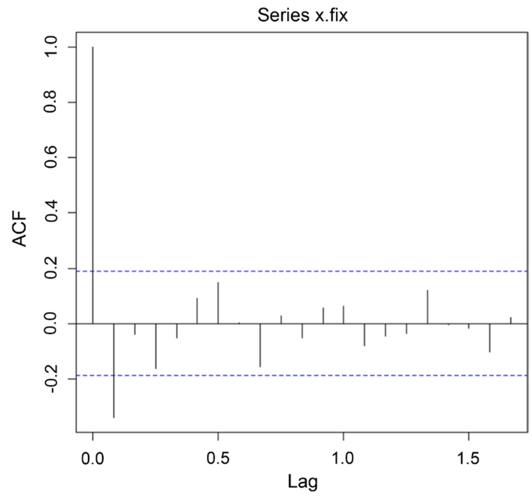

Figure 5. Second order difference sequence autocorrelation graph
图 5. 2 阶差分序列自相关图

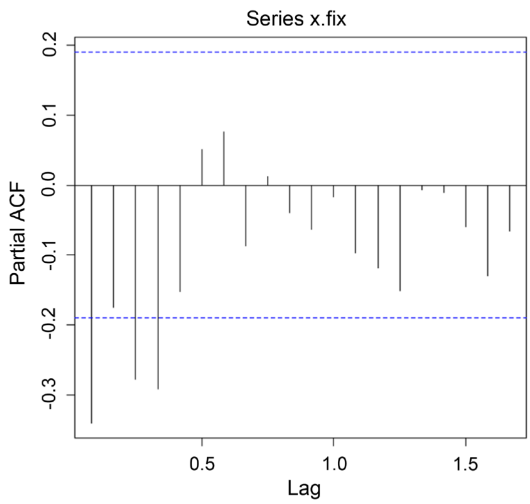

Figure 6. Second partial autocorrelation diagram of order difference sequence
图 6. 2 阶差分序列偏自相关图

2) 模型选择与参数估计

此时选择 ARIMA(p, d, q)模型进行预测时，参数选择根据 0，1，2 从低阶到高阶选择，根据 AIC 准则作为选择最优值模型。根据附录 2 程序运算可得如下结果，根据 AIC 最下即最优原则，我们选择 ARIMA(1, 2, 1)模型。

Table 1. Model selection
表 1. 模型选择

模型	Log likelihood	aic
ARIMA(1, 2, 0)	−805.12	1614.24
ARIMA(1, 2, 1)	−795.41	1596.83
ARIMA(2, 2, 0)	−803.48	1612.95
ARIMA(2, 2, 1)	−795.11	1598.23

3) 白噪声检验

对残差序列进行白噪声检验，得出 p 值 $=0.2835>0.05$，表 2。残差序列白噪声检验说明模型显著。即 ARIMA(1, 2, 1)模型对时间序列拟合成功。

Table 2. White noise test values
表 2. 白噪声检验值

延迟阶数	χ2 值	P 值
6 阶	7.4228	0.2835
12 阶	10.051	0.6115

根据估计结果，确定了该模型的拟合结果为：

$$x_t = 0.412x_{t-1} + \varepsilon_t - 0.9667\varepsilon_{t-1}，\quad \sigma_\varepsilon^2 = 164712$$

4) 模型预测

运用上述得到的 ARIMA(1, 2, 1)模型对天津市 2021 年房价分别进行置信水平为 80%和 90%双层置信区间进行预测，并给出预测表(表 3)及预测图(图 7)。

Table 3. Comparison between predicted and actual housing prices in Tianjin
表 3. 天津市房价预测值与实际值对比

2021 年 2 月~2021 年 6 月天津市房价预测值				
时间	预测值	80%置信区间	90 置信区间	实际值
2021 年 2 月	19,644.71	(19,150, 20,223)	(18,760, 20,530)	19,687
2021 年 3 月	19,599.00	(18,684, 20,514)	(18,200, 20,998)	19,600
2021 年 4 月	19,558.26	(18,327, 20,789)	(17,676, 21,441)	19,146
2021 年 5 月	19,521.95	(17,981, 21,063)	(17,167, 21,879)	18,954
2021 年 6 月	19,489.59	(17,642, 21,337)	(16,665, 22,314)	18,917

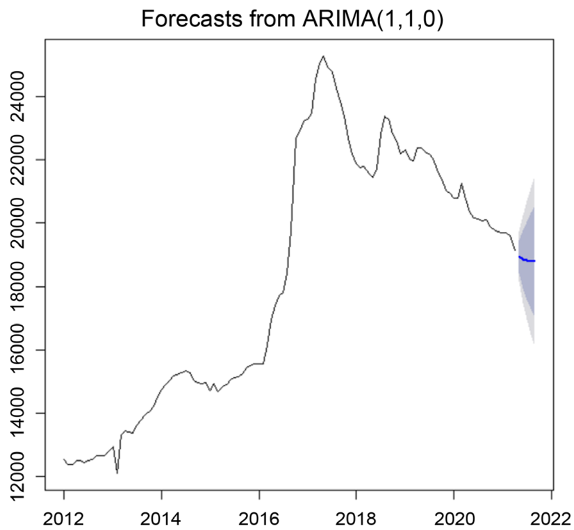

Figure 7. Model prediction diagram
图 7. 模型预测图

从预测图中可以看出，预测数据与实际数据的误差区域狭窄，说明预测与实际情况直接基本吻合，由此说明 ARIMA(1, 2, 1)模型拟合效果良好，以此模型来得出的预测结果基本上是可信的，从图中也可以看出在接下来 5 个月天津市的房价将会持续下降，且这个预测的 95%的置信区间为(16,665, 22,314)。这说明天津市在未来一段时间内房价市场将会持续低迷，也不会有明显的好转，从天津目前的房价情况来看，可以折射出我国房价在 2021 年也基本上看不到有回暖的趋势[4]。图 8 为天津市房价模型的个性化预测图，散点图为观测值序列，实线为拟合值，虚线为置信度为 95%的置信线。

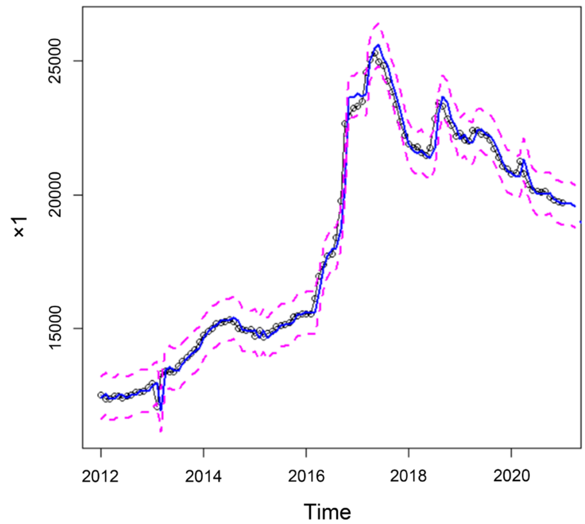

Figure 8. Personalized prediction graph
图 8. 个性化预测图

5. 模型评价

模型优点：

① 该模型是经历一系列的差分以及参数估计来确定的，准确度较高。

② 根据所得到的房价预测情况，可对居民购房决策提供一定的参考。

③ 通过预测值与实际值的对比，我们发现实际值在预测范围之内，我们有理由相信该模型具有很强的可信度。

模型缺点：

虽然实际值整体在预测范围之内，但是我们发现预测值整体偏高，可能是由于在数据采集过程中有数据缺失，或者在差分时处理完善，也表明该模型还具有改良空间。

6. 结论

本文搜集了 2012 年 1 月~2021 年 1 月天津市的房价情况，分别采用了差分、参数估计、模型预测的方法建立模型，在此基础上对序列进行分析，并预测了 5 期序列值。该模型也较好总结了天津市近 10 年房价的整体情况，以及总体发展趋势，序列结果具有较高可信度。这就是政府必须积极采取宏观措施，使得房地产市场走出目前的窘境，当然目前房地产出现如此情况，也并非政府的不作为而引发，而是“炒房”所引起的。房地产经济对于国民经济的发展具有重要作用，房地产的良性发展，对我们国家经济的发展有一个带动作用，因此调整好产业结构，对提高人民美好生活的向往有着重要作用[5]。由于房地产价格的高低不定的原因，使得居民购买时内心也是飘忽不定，由此造成买房热情浓缩，在买房问题上犹

豫不决。在这里我们利用指数平滑模型对天津市房价进行预测，从中发现了房价随时间变化的过程。揭示了天津市房地产价格发展变化规律，并通过对数据的合理分析对数据进行了预测，为近期人们在天津市购房提供了理论性的保障。我们还通过对实际问题的调查，对其依据房价构成模型选择的指标进行分析，揭示了房价波动是宏观政策调控的结果。为了保证房价稳健发展使房价能很好的通过市场机制调整，这里给出了合理的调控建议，有参考意义。 时间序列模型对短期经济预测的准确度较高，所以从计量经济学真正的应用到生产中之后，就一直在延续和发展中求完善[6]。

基金项目

校级大创项目资助(项目编号：202112026419)。

参考文献

[1] 武秀丽, 张锋. 时间序列分析法在房价预测中的应用——以广州市的数据为例[J]. 科学技术与工程, 2007, 7(21): 5631-5635.

[2] 尤梅芳, 黄敏, 程立. ARIMA 模型在房价预测中的应用——四川省商品住房价格指数未来走势的实证分析[J]. 中国物价, 2009(6): 40-42.

[3] 白晓东. 应用时间序列分析[M]. 北京: 清华大学出版社, 2017.

[4] 施仁杰, 卢科学. 时间序列分析引论[M]. 西安: 西安电子科技大学出版社, 1988: 28-31, 160-163.

[5] 杜金观, 项静怡, 戴捡华. 时间序列分析——建模与预报[M]. 安徽: 安徽教育出版社, 1991: 41-44, 126.

[6] 黄石松, 陈红梅. 房价之谜[M]. 北京: 科学文献出版社, 2009.

附录

附录 1：原始数据

	2012	2013	2014	2015	2016
1	12,510	12,941	14,752	14,717	15,553
2	12,363	12,097	14,906	14,932	15,557
3	12,379	13,304	15,012	14,685	16,138
4	12,488	13,433	15,176	14,843	16,948
5	12,491	13,386	15,223	14,889	17,405
6	12,426	13,370	15,263	15,064	17,713
7	12,490	13,603	15,335	15,104	17,795
8	12,517	13,778	15,268	15,137	18,415
9	12,638	13945	15,005	15,231	19,764
10	12645	14,059	14,960	15,425	22,659
11	12,654	14,207	14,921	15,477	22,922
12	12,793	14,519	14,961	15,553	23,220

	2017	2018	2019	2020	2021
1	23,290	21,899	22,308	20,785	19,696
2	23,468	21,751	22,056	20,804	19,687
3	24,561	21,784	21,993	21,262	19,600
4	25,048	21,586	22,381	20,774	19,146
5	25,283	21,461	22,395	20,368	18,954
6	24,941	21,734	22,263	20,165	18,917
7	24,802	22,819	22,208	20,123	
8	24,242	23,391	22,047	20,088	
9	23,852	23,283	21,713	20,114	
10	23,382	22,826	21,376	19,902	
11	22,706	22,591	21,057	19,800	
12	22,205	22,188	20,952	19,729	

附录 2：程序

```
a<-read.table("D:/Gemini/天津房价.csv",sep=",",header=T);a
x1<-ts(a$房价,start=c(2012,1),frequency=12)
plot(x1)
acf(x1)
pacf(x1)
x.fix<-diff(x1,1,2)
```

```
plot(x.fix)
acf(x.fix)
pacf(x.fix)
z1.fix<-arima(x1,order=c(1,2,0),method="ML")
for(i in 1:2)print(Box.test(z1.fix$residuals,lag=6*i))
z1.fix$aic
z2.fix<-arima(x1,order=c(1,2,1),method="ML")
for(i in 1:2)print(Box.test(z2.fix$residuals,lag=6*i))
z2.fix$aic
z3.fix<-arima(x1,order=c(2,2,0),method="ML")
for(i in 1:2)print(Box.test(z3.fix$residuals,lag=6*i))
z3.fix$aic
z4.fix<-arima(x1,order=c(2,2,1),method="ML")
for(i in 1:2)print(Box.test(z4.fix$residuals,lag=6*i))
z4.fix$aic
x1.fix<-arima(x1,order=c(1,2,1));x1.fix
library(forecast)
x.fore<-forecast(x1.fix,h=5)
x.fore
plot(x.fore)
Q1<-x1.fore$fitted-1.96*sqrt(x1.fix$sigma2)
Q2<-x1.fore$fitted+1.96*sqrt(x1.fix$sigma2)
B1<-ts(x1.fore$lower[,2])
B2<-ts(x1.fore$upper[,2])
J1<-min(x1,Q1,B1)
J2<-max(x1,Q2,B2)
plot(x1,type="o",pch=1,ylim=c(J1,J2))
lines(x1.fore$fitted,col=4,lwd=2)
lines(x1.fore$mean,col=4,lwd=2)
lines(Q1,col=6,lty=2,lwd=2)
lines(Q2,col=6,lty=2,lwd=2)
lines(B1,col=7,lty=2,lwd=2)
lines(B2,col=7,lty=2,lwd=2)
```