基于ARMA模型进行未来5年北京省的居民消费价格指数的预测

学 号： 202120113220

姓 名： 蒋婧雯

专业名称： 数据科学与大数据技术

指导教师：马健

# 摘要

**AＲMA( Auto Ｒegression Moving Average) ，即自回归移动平均模型，是目前最常用的平稳时间序列模型，可分为 AＲ( Auto Ｒegression ) 模型、MA( Moving Average) 模型和 AＲMA 模型**

**关键词：预测，平稳分析，ARMA模型**

# **Abstract**

ARMA (AutoRegressive Moving Average), which stands for AutoRegressive Moving Average model, is currently the most commonly used stationary time series model. It can be divided into AR (AutoRegressive) model, MA (Moving Average) model, and ARMA model.

**Keywords:Forecasting, Stationary Analysis, ARMA model**

# **目录**

[摘要 I](#_Toc22444)

[Abstract II](#_Toc17505)

[目录 III](#_Toc1189)

[前言 1](#_Toc1826)

[1 数据相关背景概述 2](#_Toc29316)

[1.1 通过官方新闻媒体渠道收集了2020-2024年北京省的居民消费价格指数（上年同月＝100） 2](#_Toc25023)

[2 相关理论方法 3](#_Toc8725)

[3. 数据统计 5](#_Toc4205)

[3. 1 数据统计 5](#_Toc27402)

[4 结果与分析 6](#_Toc3893)

[4.1 稳定性分析： 6](#_Toc22492)

[4.2自相关函数和偏自相关函数的计算 8](#_Toc10974)

[4.3 模型定阶 10](#_Toc1444)

[4.4 系数确定及模型预测结果分析 11](#_Toc10720)

[5 结论 15](#_Toc20403)

[参考文献 16](#_Toc22042)

# 前言

ARMA（AutoRegressive Moving Average）模型是用于分析和预测时间序列数据的一种常用统计方法。它结合了自回归（AR）和移动平均（MA）组件，以捕捉数据中的依赖关系。AR部分表示观测值与一定数量的滞后观测值之间的线性关系，而MA部分则基于滞后的误差项值捕捉观测值与随机误差项之间的线性关系。ARMA模型广泛应用于经济学、金融学和工程学等各个领域，因其有效建模和预测时间序列数据的能力。

# 1 数据相关背景概述

## 通过官方新闻媒体渠道收集了2020-2024年北京省的居民消费价格指数（上年同月＝100）

现在对这个数据进行分析，确定其是否平稳，同时利用ARMA模型进行预测未来5年消费价格指数的变化情况，计算他的纯随机性，判断是否具有周期性或者季节性，是否适用于ARMA模型进行预测。利用绝对误差和相对误差对数据进行整体对比，判断最优模型的预测效果。

# 2 相关理论方法

公式（2-1）：

（2-1）

公式（2-2）：

（2-2）

公式（2-3）

（2-3）

AＲ( p) 模型中，

t 时刻的时间序列数值 Xt 与其前 p 时刻的时间序列数据 Xt － 1，Xt － 2，…，Xt － p 有关［5］，见 公式（2-4）:

Xt = φ1Xt － 1 + φ2Xt － 2 + … + φpXp － 1 + Ct 公式（2-4）

92其中 φ1，φ2，…，φp 为系数，Ct 是常数项。

MA( q) 模型中，t 时刻的时间序列数值 Xt 与其前 p时刻的时间序列数据 Xt － 1，Xt － 2，…，Xt － p 及前 q 时刻的误差扰动数据

εt － 1，εt － 2，…，εt － q均相关［5］，见公式（2-5）:

Xt = Ct － θ1εt － 1 － θ2εt － 2 － … － θqεt － q 公式（2-5）

其中 θ1，θ2，…，θq 为系数，εt 是均值为零的误差扰动。

AＲMA( p，q) 模型中，t 时刻的时间序列数值 Xt 与其前 p 时刻的时间序列数据 Xt － 1，Xt － 2，…，Xt － p及前 q时刻的误差扰动数据 εt － 1，εt － 2，…，εt － q均相关［5］，见式( 2-6) :

Xt = φ1Xt － 1 + φ2Xt － 2 + … + φpXp － 1

+ Ct － θ1εt － 1 － θ2εt － 2 － … － θqεt － q 式（2-6）

由式( 3) 可知: 当 q = 0 时，AＲMA( p，q) 模型即 AＲ ( p) 模型; 当 p = 0 时，AＲMA( p，q) 模 型 即 MA ( q) 模型。

# 数据统计

## 3. 1 数据统计

通过官方新闻媒体渠道收集2020-2024年每月北京省的居民消费价格指数（上年同月＝100），按给出的对应的学号的表中的数据在官方下搜索，如图下表,表3.1

表3.1 2020-2024年北京居民消费价格指数

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Jan-20 | Feb-20 | Mar-20 | Apr-20 | May-20 | Jun-20 | Jul-20 |
| 104.47 | 103.55 | 103.25 | 102.38 | 101.86 | 101.44 | 100.73 |
| Aug-20 | Sep-20 | Oct-20 | Nov-20 | Dec-20 | Jan-21 | Feb-21 |
| 100.89 | 101 | 100.9 | 100.2 | 100.2 | 99.2 | 99.9 |
| Mar-21 | Apr-21 | May-21 | Jun-21 | Jul-21 | Aug-21 | Sep-21 |
| 100.6 | 101.1 | 101.2 | 100.9 | 101.5 | 101.4 | 101.2 |
| Oct-21 | Nov-21 | Dec-21 | Jan-22 | Feb-22 | Mar-22 | Apr-22 |
| 102 | 102.4 | 101.8 | 101.3 | 101.2 | 101.8 | 102 |
| May-22 | Jun-22 | Jul-22 | Aug-22 | Sep-22 | Oct-22 | Nov-22 |
| 102.2 | 102.5 | 102.1 | 101.9 | 102.1 | 101.3 | 101.2 |
| Dec-22 | Jan-23 | Feb-23 | Mar-23 | Apr-23 | May-23 | Jun-23 |
| 101.8 | 102 | 101.2 | 100.7 | 100.3 | 100.2 | 100.1 |
| Jul-23 | Aug-23 | Sep-23 | Oct-23 | Nov-23 | Dec-23 | Jan-24 |
| 100.1 | 100.3 | 100.4 | 100 | 99.7 | 99.5 | 99.6 |

# 4 结果与分析

## 4.1 稳定性分析：

原始数据的时序图如图4-1

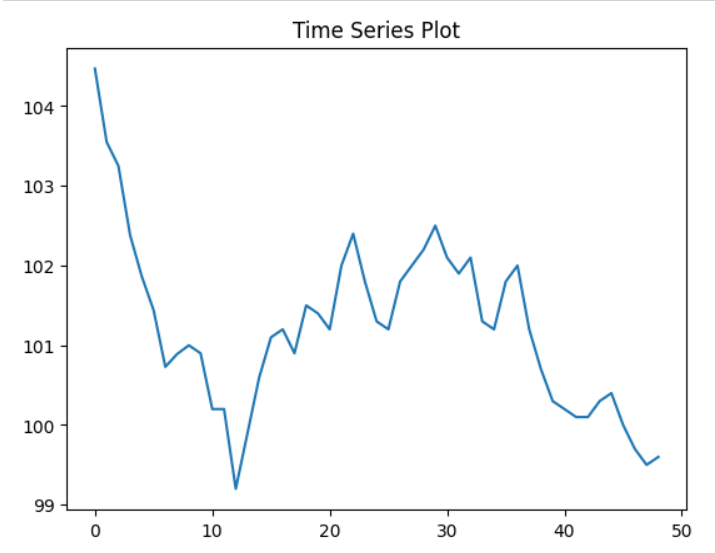


图4-1

对其进行平稳性分析：如表4-2

表4-2 平稳性结果

|  |  |
| --- | --- |
| ADF Statistic | P-value |
| -2.920274 | 0.043041 |

发现是平稳的无需进行差分。ADF统计值越小，意味着越有可能拒绝原假设，p值小于显著性水平（通常为0.05）时，我们拒绝原假设，因此这个时间序列是平稳的。

纯随机性结果如图4-22

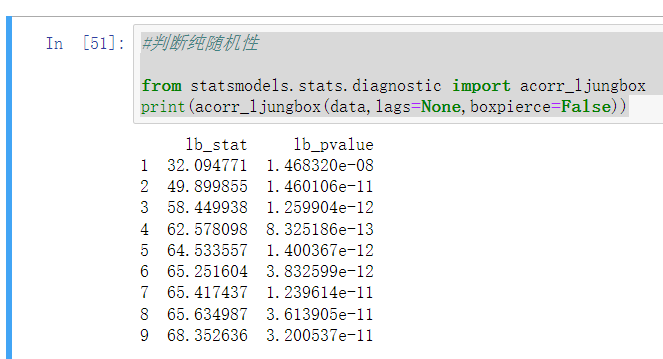


图4-22

判断纯随机性：通过结果知道，对应的p值都比较小，接近0了，因此时间序列数据中存在显著的自相关性的作用，因此用ARMA模型预测是比较合适的，预测的结果也比较恰当。

**4.2**自相关函数和偏自相关函数的计算

自相关函数( ACF) 反映时间序列 Xt，Xt － 1，Xt － 2，

…，Xt － k 间 的 关 系，自相关程度通过下式

偏自相关函数( PACF) 反映 Xt 在 Xt － 1，Xt － 2，…，

Xt － k + 1条件下，Xt 与 Xt － k间的关系，通过式( 5) 、式( 6)

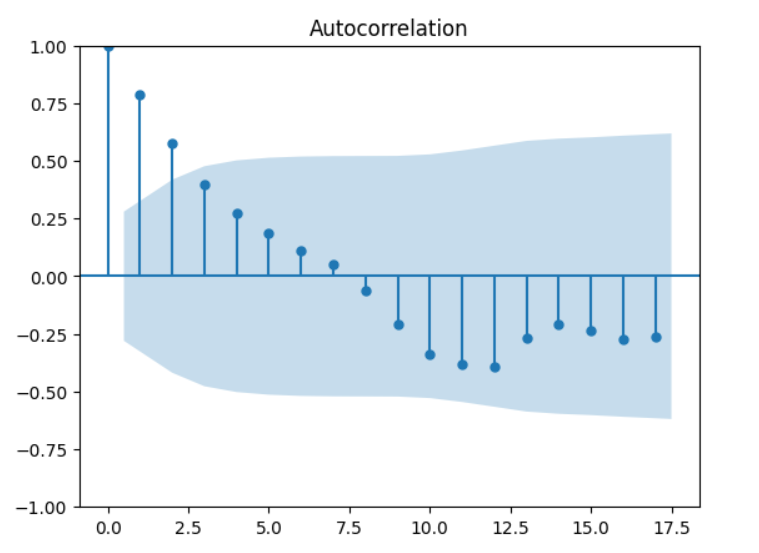
进行计算［5］。

=,k=2,3,... (5)

(6)

**自相关图和偏自相关图如图4-3**

可以看出自相关图拖尾，偏自相关图截尾，判断为AR(P)模型。且根据PACF图，初步判断p为2。



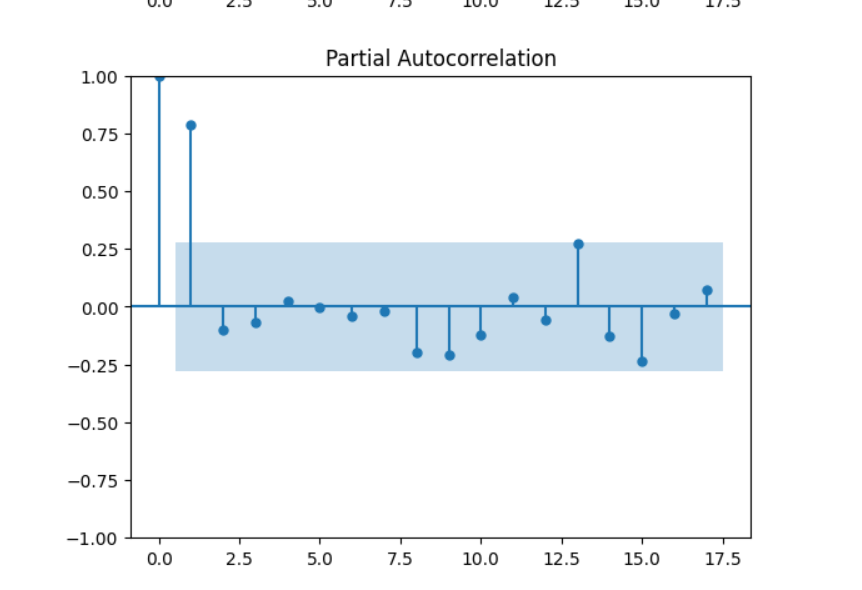


图4-3

**4.3** 模型定阶

使用 AIC 准则函数［6］来确定 AＲMA( p，q) 中的 p， q 值。函数具体形式为:

AIC(S)=2ln^ σ+ (7)

其中 S 为模型未知参数的总数，^ σ2 是标准差的估

计值，N 为样本总数，S = p + q。 p，q的确定方法为寻求 AIC( S) 的最小值点( p，q) 。Matlab 7 有专门计算 AIC 值的指令 aicbic。令 p,q ∈［0，2］，计算各 AIC 值，见下表：表4-7

表4-7 不同p,q下的AIC值

|  |  |  |
| --- | --- | --- |
| p | q | AIC |
| 0 | 0 | 149.026 |
| 0 | 1 | 109.202 |
| 0 | 2 | 93.816 |
| 1 | 0 | 73.174 |
| 1 | 1 | 71.758 |
| 1 | 2 | 73.714 |
| 2 | 0 | 71.706 |
| 2 | 1 | 73.684 |
| 2 | 2 | 72.834 |

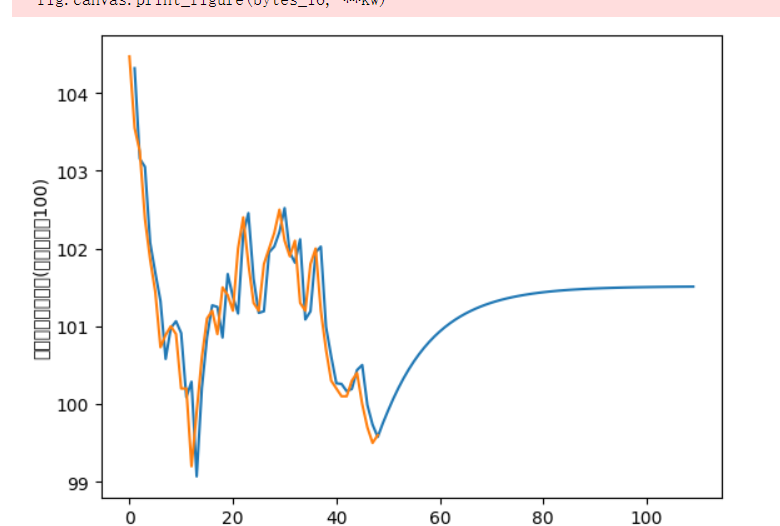
可以看出当p=2，q=0时，AIC最小，且利用求最小AIC的值的代码，发现p=2,q=0时最小，判断为最优模型。

**4.4** 系数确定及模型预测结果分析

预测函数为：

Xt=φ1Xt-1+φ2Xt-2+φ3－1 + Ct－θ1εt (8)

利用最优模型预测2020-2024年和未来5年每月的居民消费价格指标情况，其中，蓝色为预测值，橙色为实际值，结果如图4-21所示

 图4-21

**可以看出这个预测效果是比较好的，比较符合原数据。**

分析：通过结果发现，未来5年北京省的居民消费价格指数会慢慢的升高，然后在2027年到2029年趋于稳定。2020年到2021年这段时间居民指数骤降，可能是因为疫情的原因，然后缓慢增加，大概是疫情结束，从大概2023年到2024年居民指数骤降，到2024年底下降为最低，疫情结束后居民消费指数缓慢升高，但是最后还是在101到102之间，消费指数虽然在升高，但是还是不如2020年的消费指数水平，甚至跟疫情期间的消费指数水平差不多，可能反映出北京省的人们因为疫情的影响，消费思想转变了，花钱不那么大手大脚，也可能反映出疫情过后整体的就业形式不咋好。

**计算各省2020年的实际值与预测值，以及计算其绝对误差和相对误差：用来判断最优模型是否可以用来进行预测，结果如表4-28**

其中，相对误差的计算方法为“( 绝对误差/实际值) × 100% ”。

绝对误差的计算方法为：abs(预测值-实际值)

**表4-28 2020年1月到2024年1月相对误差和绝对误差**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **指标** | **居民消费价格指数(上年同月＝100)** | **预测值** | **绝对误差** | **相对误差 %** |
| Jan-20 | 104.47 | 101.5136062 | 2.956393843 | 0.028298974 |
| Feb-20 | 103.55 | 104.3173968 | 0.767396826 | 0.007410882 |
| Mar-20 | 103.25 | 103.154038 | 0.095961982 | 0.000929414 |
| Apr-20 | 102.38 | 103.0498245 | 0.669824496 | 0.006542533 |
| May-20 | 101.86 | 102.075752 | 0.215752016 | 0.002118123 |
| Jun-20 | 101.44 | 101.6895115 | 0.2495115 | 0.002459695 |
| Jul-20 | 100.73 | 101.3257142 | 0.595714213 | 0.00591397 |
| Aug-20 | 100.89 | 100.5804817 | 0.309518265 | 0.003067879 |
| Sep-20 | 101 | 100.9767176 | 0.023282389 | 0.000230519 |
| Oct-20 | 100.9 | 101.0652426 | 0.165242572 | 0.001637687 |
| Nov-20 | 100.2 | 100.9122873 | 0.712287306 | 0.007108656 |
| Dec-20 | 100.2 | 100.0884384 | 0.111561567 | 0.001113389 |
| Jan-21 | 99.2 | 100.2870437 | 1.087043706 | 0.010958102 |
| Feb-21 | 99.9 | 99.06958506 | 0.830414944 | 0.008312462 |
| Mar-21 | 100.6 | 100.2055279 | 0.39447207 | 0.003921194 |
| Apr-21 | 101.1 | 100.8591437 | 0.240856288 | 0.002382357 |
| May-21 | 101.2 | 101.2692678 | 0.069267764 | 0.000684464 |
| Jun-21 | 100.9 | 101.2491527 | 0.34915272 | 0.003460384 |
| Jul-21 | 101.5 | 100.8555429 | 0.644457057 | 0.006349331 |
| Aug-21 | 101.4 | 101.6711347 | 0.271134679 | 0.002673912 |
| Sep-21 | 101.2 | 101.3791557 | 0.179155722 | 0.001770313 |
| Oct-21 | 102 | 101.1640362 | 0.835963826 | 0.008195724 |
| Nov-21 | 102.4 | 102.1947475 | 0.205252542 | 0.002004419 |
| Dec-21 | 101.8 | 102.4547535 | 0.654753463 | 0.006431763 |
| Jan-22 | 101.3 | 101.6107895 | 0.310789545 | 0.003068011 |
| Feb-22 | 101.2 | 101.1722933 | 0.027706689 | 0.000273782 |
| Mar-22 | 101.8 | 101.1924084 | 0.607591644 | 0.005968484 |
| Apr-22 | 102 | 101.9512557 | 0.048744272 | 0.000477885 |
| May-22 | 102.2 | 102.0245144 | 0.175485633 | 0.001717081 |
| Jun-22 | 102.5 | 102.2112617 | 0.288738267 | 0.002816959 |
| Jul-22 | 102.1 | 102.519755 | 0.419754964 | 0.004111214 |
| Aug-22 | 101.9 | 101.947655 | 0.047654958 | 0.000467664 |
| Sep-22 | 102.1 | 101.817652 | 0.282348044 | 0.002765407 |
| Oct-22 | 101.3 | 102.117888 | 0.81788805 | 0.00807392 |
| Nov-22 | 101.2 | 101.0871768 | 0.112823234 | 0.001114854 |
| Dec-22 | 101.8 | 101.1924084 | 0.607591644 | 0.005968484 |
| Jan-23 | 102 | 101.9512557 | 0.048744272 | 0.000477885 |
| Feb-23 | 101.2 | 102.0245144 | 0.824514367 | 0.008147375 |
| Mar-23 | 100.7 | 100.9938031 | 0.293803082 | 0.002917608 |
| Apr-23 | 100.3 | 100.6120512 | 0.312051213 | 0.003111179 |
| May-23 | 100.2 | 100.2669287 | 0.066928662 | 0.000667951 |
| Jun-23 | 100.1 | 100.2586715 | 0.158671525 | 0.00158513 |
| Jul-23 | 100.1 | 100.1652978 | 0.065297841 | 0.000652326 |
| Aug-23 | 100.3 | 100.19367 | 0.106329977 | 0.001060119 |
| Sep-23 | 100.4 | 100.4371618 | 0.037161753 | 0.000370137 |
| Oct-23 | 100 | 100.5021633 | 0.502163255 | 0.005021633 |
| Nov-23 | 99.7 | 99.98680761 | 0.286807613 | 0.002876706 |
| Dec-23 | 99.5 | 99.73505875 | 0.235058745 | 0.002362399 |
| Jan-24 | 99.6 | 99.57668356 | 0.023316439 | 0.000234101 |
| **总共** | 4959.57 | 4960.957228 | 19.34233744 | 0.0003 |

发现北京省2020年1月到2024年1月实际值总和约为4959.57，预测值总和约为4960.96，发现相对误差为0.0003%，可以看出说明该模型适合用于进行预测。

# 5 结论

利用ARMA（2，0）模型可以看出：

未来5年北京省的居民消费价格指数会慢慢的升高，然后在2027年到2029年趋于稳定。2020年到2021年这段时间居民指数骤降，可能是因为疫情的原因，然后缓慢增加，大概是疫情结束，从大概2023年到2024年居民指数骤降，到2024年底下降为最低，疫情结束后居民消费指数缓慢升高，但是最后还是在101到102之间，消费指数虽然在升高，但是还是不如2020年的消费指数水平，甚至跟疫情期间的消费指数水平差不多，可能反映出北京省的人们因为疫情的影响，消费思想转变了，花钱不那么大手大脚，也可能反映出疫情过后整体的就业形式不咋好。

# 参考文献

[1].Shumway, R. H., & Stoffer, D. S.《时间序列分析及其应用：带R示例》[书籍]，Springer，2006年。

[2].Montgomery, D. C., Jennings, C. L., & Kulahci, M.《时间序列分析与预测导论》[书籍]，Wiley，2008年。

[3].Box, G. E. P., Jenkins, G. M., & Reinsel, G. C.《时间序列分析：预测与控制》[书籍]，Wiley，2015年。

[4].Tsay, R. S.《金融时间序列分析》[书籍]，Wiley，2010年。

[5].Cryer, J. D., & Chan, K. S.《时间序列分析：R语言应用》[书籍]，Springer，2008年。