|  |  |
| --- | --- |
| 成绩 |  |

**云南大学2024学年春季学期数学与统计学院2021级**

**《应用时间序列分析实验》**

**期末论文**

**题目**

**我国外商直接投资额的趋势分析及预测**

学 院：　数学与统计学院

专 业：　　　统计学

姓 名：　　　枫叶

学 号：

教 师：　　　周建军

**摘 要：**

本文旨在对我国外商直接投资（FDI）额的历史趋势及未来发展进行分析与预测。本文基于国家统计局公布的月度FDI数据，采用GDP平减指数将其调整为1999年不变价数据，在此基础上使用ARIMA模型及干预分析方法，重点考察了2008年全球金融危机对我国FDI流入的影响。通过对原始序列进行差分处理及模型拟合，识别出适合的ARIMA模型，并结合Chen C. 和 Liu Lon-Mu提出的异常点识别方法，进一步评估了金融危机及其他重大事件对FDI的冲击。最终，本文构建了包含异常点的ARIMAX模型，对我国FDI进行了短期预测。实证结果表明，我国FDI总体保持稳定增长，外部冲击对其长期趋势影响有限。

**关键词：**外商直接投资（FDI）;时间序列分析;ARIMA模型;干预分析;预测

# 引言

## 背景和意义

外商直接投资（FDI）在全球经济一体化进程中扮演着至关重要的角色，对中国经济的发展起到了巨大的推动作用。研究我国FDI的历史趋势、波动及其影响因素，不仅有助于全面了解其发展规律，总结经验，识别潜在问题，还能科学预测未来的FDI发展趋势，为政策制定者提供可靠的数据支持和决策参考。

## 国内外研究现状

对FDI的预测研究，马静和赵果庆(2009)基于logistic模型对中国FDI进行了混沌吸引子检测及非线性动力学预测;郑岩岩等(2016)运用灰色马尔科夫预测模型和时间序列模型分别对我国FDI进行预测并对两种方法进行比较;索寒蕾(2016)运用了乘积季节的ARIMA模型进行预测，并对所建立的模型进行了修正;祁馨禾(2019)比较了三种方法，认为相对非线性BP神经网络模型和ARIMA模型，Holt－Winters模型对FDI值的预测较好。本文采用经典的时间序列分析预测法ARIMA模型对FDI进行预测。

## 本文主要研究内容

本文主要围绕我国外商直接投资（FDI）的趋势分析及预测展开。首先，收集了国家统计局公布的月度FDI数据，并使用GDP平减指数将数据调整为1999年不变价，进行数据的预处理。接着，通过时序图展示FDI的历史趋势，发现2002年至2007年期间FDI呈现稳定增长趋势，而2008年金融危机对FDI产生了显著冲击，但其影响未持续太久，自2010年起，FDI流入保持相对平稳，未出现较大波动。

在模型构建与识别部分，本文依据时序图及自相关函数（ACF）和偏自相关函数（PACF）图，对原始数据进行一阶差分和12步差分，以确保序列平稳。随后进行模型识别和定阶，选择最优的ARIMA模型并进行拟合。干预分析部分，本文采用ARIMA模型对2008年金融危机后的FDI进行预测，评估干预形式，并识别出2008年12月和2017年11月的显著异常点。基于Chen C.和Liu Lon-Mu的异常点识别方法，进一步分析这些异常点对FDI的影响，并构建包含异常点的ARIMAX模型。

在模型检验与预测部分，对模型进行拟合和检验，确保其准确性和稳定性，并使用构建的ARIMAX模型，对2019年至2023年的FDI进行预测，提供未来几年的发展趋势。在实证结果与讨论部分，分析实证结果，发现我国FDI总体保持稳定增长，外部冲击对其长期趋势影响有限，并讨论了模型预测的可靠性及政策建议，为进一步优化投资环境和吸引外资提供参考。

# 研究方法

## ARIMA模型

SARIMA模型是季节性自回归移动平均模型（Seasonal Autoregressive Integrated Moving Average），它是ARIMA模型的一种扩展。ARIMA模型用于处理时间序列数据，特别是在数据具有明显的自相关和季节性变化时。SARIMA模型则考虑了时间序列数据中的季节性因素，通过增加季节性差分项来更好地拟合数据。

SARIMA模型有四个主要部分：季节性部分（Seasonal）：指数据中的重复周期性变化，如每年、每季度或每月的变化。自回归部分（Autoregressive，AR）：指模型中使用过去时间点的数据来预测当前时间点的数据，用AR表示。差分部分（Integrated，I）：指对原始数据进行差分操作，使数据变得平稳（stationary）。移动平均部分（Moving Average，MA）：指模型中使用过去预测误差的数据来预测当前时间点的数据，用MA表示。

SARIMA模型的表示形式为SARIMA(p, d, q)(P, D, Q)s，其中p、d、q分别表示非季节性自回归阶数、差分阶数和移动平均阶数。P、D、Q分别表示季节性自回归阶数、季节性差分阶数和季节性移动平均阶数。s表示季节周期的长度，如s=12表示每年的季节周期为12个月。

通过调整这些参数，SARIMA模型可以灵活地适应不同时间序列的季节性和趋势变化，是处理季节性时间序列数据的重要工具之一。

## 干预分析

干预分析（Intervention Analysis）是一种用于评估特定事件或干预对时间序列数据影响的方法。干预分析通过引入干预变量，评估事件发生前后时间序列的变化情况，通常应用于研究经济政策、自然灾害、金融危机等重大事件对经济指标的影响。干预分析可以帮助理解事件的短期冲击和长期影响。

在干预分析中，干预变量可以是瞬时冲击（如某一时点的突然变化）或持续影响（如政策实施后的长期效果）。模型通常以ARIMA模型为基础，加入干预项进行修正。干预模型可以表示为：

通过估计和检验干预模型，可以量化干预事件的影响，识别显著的异常点，并评估其对时间序列的长期影响。这种方法在经济学、金融学、公共政策等领域广泛应用，为决策者提供了重要的实证依据。

# 实证分析

## 数据来源

本文以实际利用外商直接投资额衡量外商投资情况，数据来源于国家统计局网站。原始数据为月度现价数据，本文使用GDP平减指数将其折算为1999年不变价数据，由于国家统计局仅公布季度GDP指数及现价GDP，为方便起见，本文假定同一季度中GDP平均分配，并假定基期各季度GDP平减指数均为1。

## 描述性分析

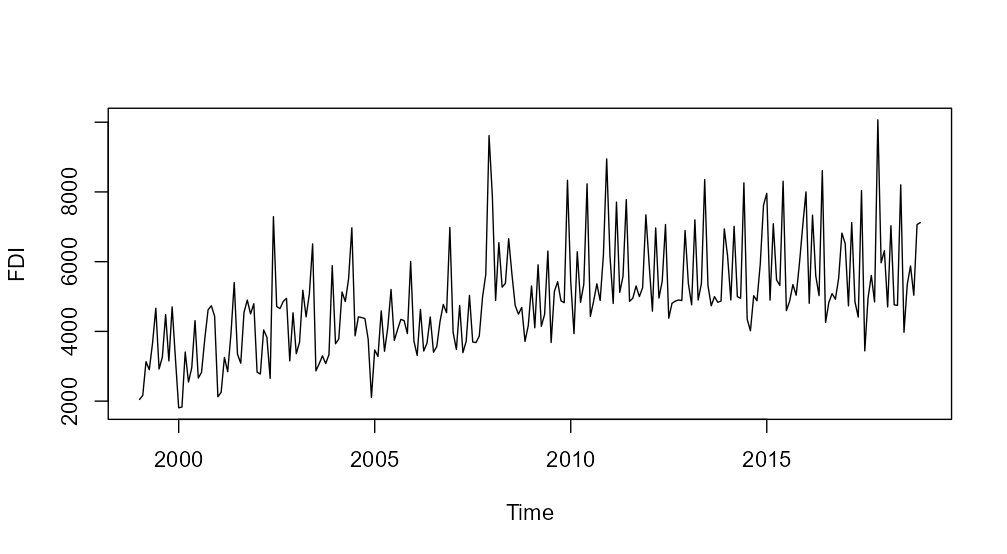


图 1 FDI时序图

图 1展示了我国FDI的历史趋势，其中纵坐标表示外商投资额，单位为百万美元。从时序图中可以看出，2002年至2007年期间，我国的FDI流入呈现稳定增长的趋势。而2008年的全球金融危机对我国FDI的流入造成了一定冲击，使得FDI大幅下降，但该冲击并未持续太久。 自2010年起，我国FDI流入处于较为平稳的状态，并未出现较大波动。总体来看，我国FDI保持了稳中向好的态势，偶发性的外生事件冲击也并没有影响FDI的长期趋势。

## 识别ARIMA模型

考虑到2008年金融危机的冲击，首先对2008年1月之前的序列进行分析。从时序图来看，序列具有长期趋势且存在季节性，结合原始数据的ACF图和PACF图（图 2和图 3）来看，可以对序列进行一阶差分和12步差分。对差分后序列做ADF检验，检验统计量为-5.26，小于1%显著性水平的临界值-2.58，故可在1%显著性水平下认为差分后的序列不存在单位根，序列平稳。同时，基于Q统计量的纯随机性检验结果也显著地拒绝了纯随机性的假设，可以认为序列不是白噪声，故下面进一步考察差分后序列的ACF图和PACF图以识别ARIMA模型。

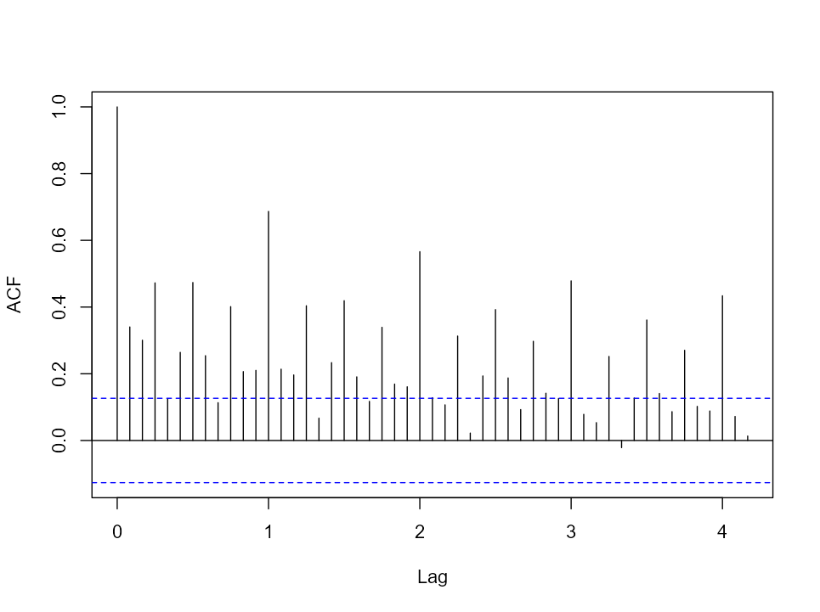


图 2 金融危机前FDI自相关系数图

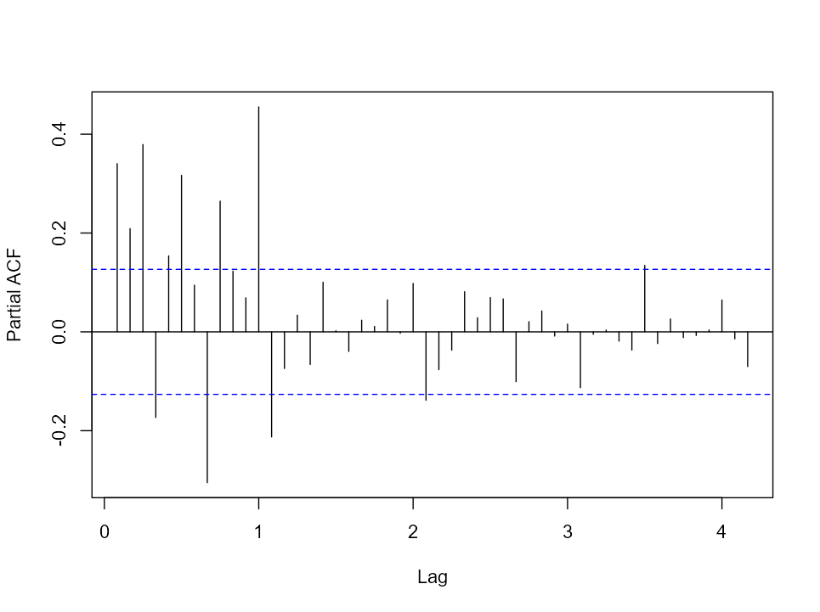


图 3 金融危机前FDI偏自相关系数图

差分后序列的ACF图和PACF（图 4和图 5），可以认为在周期长度上ACF一阶截尾，PACF拖尾，因此P=0，Q=1。而在一个周期内，ACF没有明显拖尾性，PACF可以视作一阶截尾，若认为ACF拖尾，则p=1，q=0，若认为ACF截尾，则可尝试p=0，q=3或p=0，q=11。

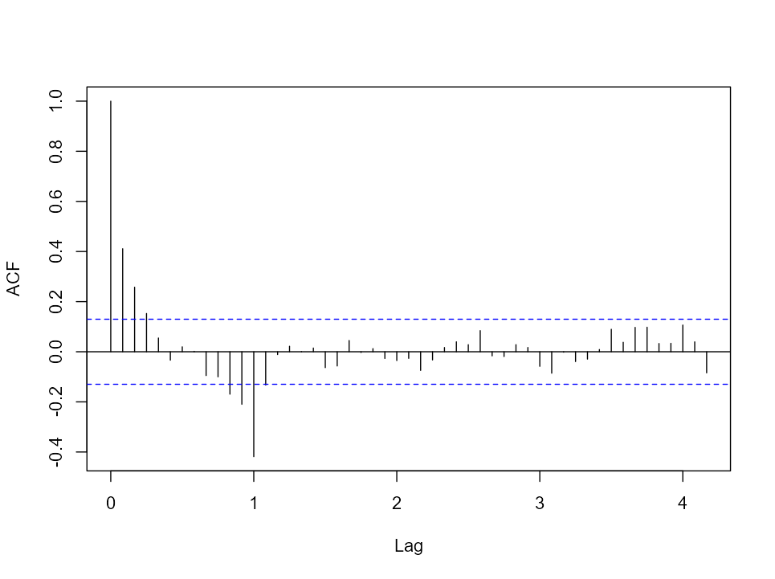


图 4 差分后序列自相关系数图

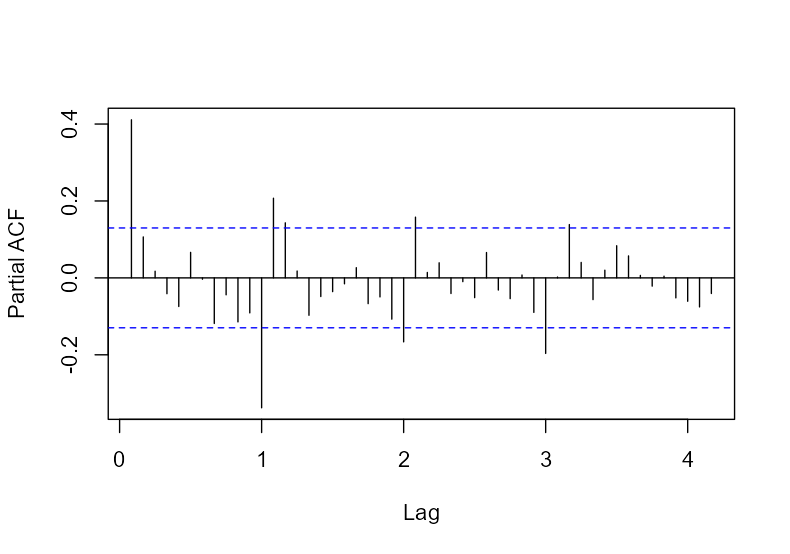


图 5 差分后序列偏自相关系数图

## 模型拟合与检验

上述三种模型的拟合结果如表 1所示，可以看到三个模型的拟合效果相差不多，从RMSE的大小来看，模型2的效果最佳，但其参数较多，故BIC相对于其他模型更大。对三个模型均做纯随机性检验，其p值均在0.9以上，在0.05的显著性水平下不能拒绝纯随机性假设，这说明三个模型均较好地提取了序列信息。考虑到预测问题，本文选择RMSE和BIC均居于中间的模型1作为进一步分析的基础。

表 1 三种模型拟合结果

|  | Arima(1,0,0)(0,1,1) | Arima(0,0,3)(0,1,1) | Arima(0,0,11)(0,1,1) |
| --- | --- | --- | --- |
| ar1 | 0.570\*\*\* |  |  |
|  | (0.098) |  |  |
| sma1 | -0.498\*\*\* | -0.490\*\*\* | -0.462\*\*\* |
|  | (0.127) | (0.122) | (0.127) |
| ma1 |  | 0.496\*\*\* | 0.473\*\*\* |
|  |  | (0.101) | (0.102) |
| ma2 |  | 0.391\*\*\* | 0.290\*\* |
|  |  | (0.105) | (0.094) |
| ma3 |  | 0.180+ |  |
|  |  | (0.106) |  |
| RMSE | 855.24 | 849.41 | 863.54 |
| BIC | 1613.9 | 1621.6 | 1619.7 |
| + p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001 | | | |

## 干预分析

基于ARIMA(1,0,0)(0,1,1)模型对2008年1月后的FDI进行预测，并计算预测值与实际值的差值，由此考察干预形式。

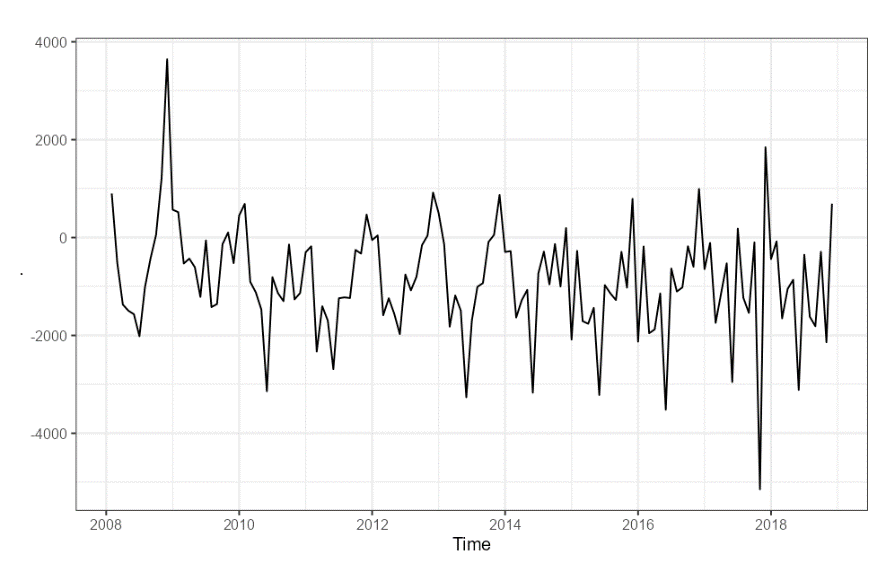


图 6 事件发生后无干预模型预测值与实际值差值

从图 6从可以发现，所建立的模型总体上而言低估了我国FDI的规模，而2018年12月远远高估了实际的FDI水平，2017年11月则远远低估了实际的FDI水平，前者可能为金融危机造成的影响，后者的机制尚不明确。据此可以认为，金融危机所造成的影响是脉冲干预形式。

除了依靠经验确定的2008年金融危机冲击，本文进一步基于Chen, C. and Liu, Lon-Mu (1993)的研究对异常点进行识别，识别结果如表 2所示，其中AO表示加性异常，或称为脉冲干预，即冲击仅在当期起作用，其中2008年12月的异常点和2017年11月的异常点在上文中已经指出。下面将所有异常点均纳入ARIMAX模型，回归系数已列于表 2中。

可以看到金融危机使得我国FDI当年下降了约3769百万美元。

表 2 异常点识别

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 异常点 | 异常类型 | 时间 | 系数 | t统计量 |
| 1 | AO | 2002:05 | -2530 | -6.799 |
| 2 | AO | 2004:12 | -3764 | -10.120 |
| 3 | AO | 2008:01 | 1744 | 4.558 |
| 4 | AO | 2008:12 | -3769 | -9.768 |
| 5 | AO | 2017:11 | 4284 | 11.465 |

考虑到所有异常点均为加性异常，且基于1999-2008的数据所确定的模型总体上而言低估了实际值，故有必要基于完整的数据建立模型，拟合结果如表 3所示。从中可以看到，和相对于之前的模型均有所减小，并且在样本增多的情况下RMSE不升反降，不过BIC的增幅较大。

表 3 完整数据模型拟合结果

|  | 完整数据模型 |
| --- | --- |
| ar1 | 0.456\*\*\* |
|  | (0.059) |
| sma1 | -0.555\*\*\* |
|  | (0.055) |
| RMSE | 834.69 |
| BIC | 3747.2 |
| + p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001 | |

## 预测

对2019-2023的预测结果如xxx所示，为美观起见没有给出置信区间。

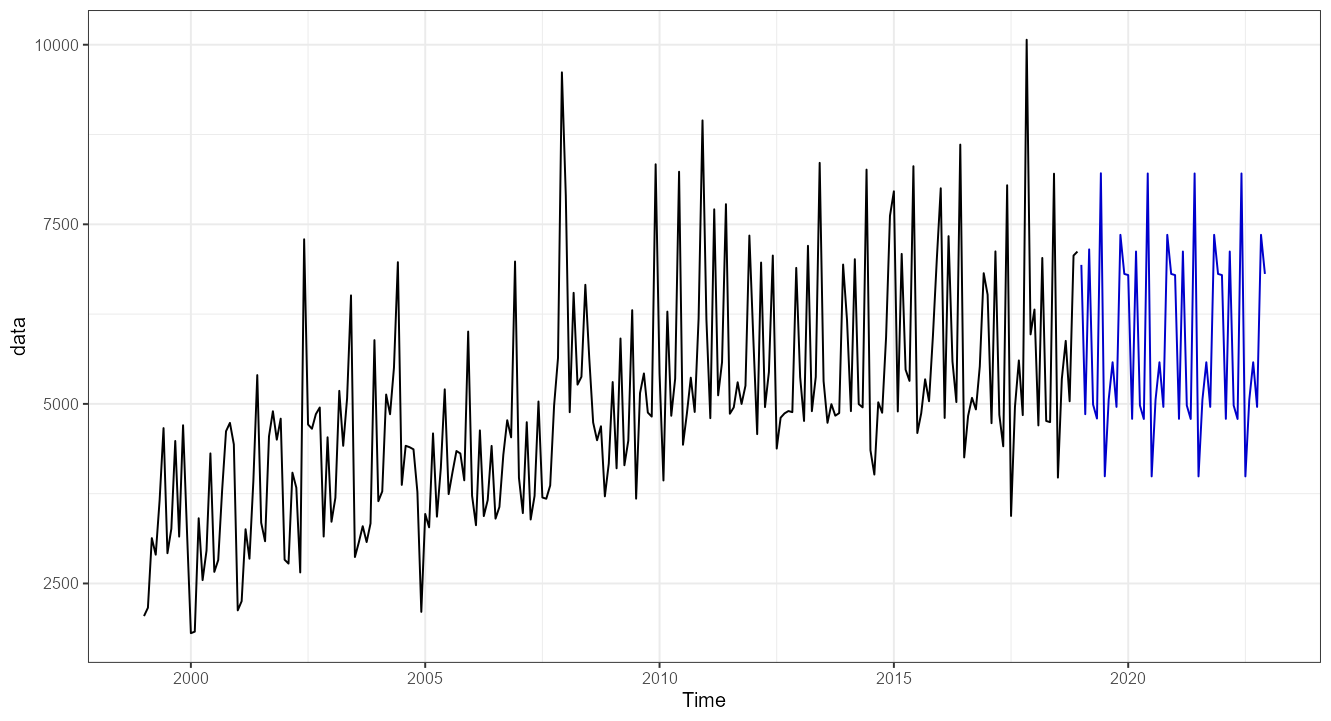


图 7 2019-2023预测结果

# 参考文献

[1] Chen, C. and Liu, Lon-Mu (1993). ‘Joint Estimation of Model Parameters and Outlier Effects in Time Series’. Journal of the American Statistical Association, 88(421), pp. 284-297.

# 附录

附录1数据见压缩包中文件“月度FDI.csv”和“季度GDP.csv”

附录2 R程序代码

|  |
| --- |
| library(tidyverse) library(forecast) library(modelsummary) library(tsoutliers) library(urca) library(flextable) #预处理 deflator <- read.csv("D:/预删除文件夹/大三下/时间序列/季度GDP.csv") %>%  select(-1) %>%  mutate(GDP指数=GDP指数/100,季度=rep(1:4,20)) %>%  group\_by(季度) %>%  mutate(基期指数=cumprod(GDP指数),  平减指数=GDP/(基期指数\*19361.9)) %>%  .$平减指数 %>%  rep(each=3) data <- read.csv("D:/预删除文件夹/大三下/时间序列/月度FDI.csv") %>%  select("FDI") %>%  mutate(FDI=FDI/deflator) %>%  ts(start = c(1999,01),frequency = 12) #差分 plot(data) acf(data,lag.max = 50,main=NA) acf(data,lag.max = 50,type = "partial",main=NA) data1 <- window(data,end=c(2008,01)) d1 <- diff(data) acf(d1,lag.max = 50) acf(d1,lag.max = 50,type = "partial") d2 <- diff(data,lag = 12) plot(d2) ur.df(d2,lags = 10) %>% summary() Box.test(d2,lag = 10) #定阶 acf(d2,lag.max = 50,main=NA) acf(d2,lag.max = 50,type = "partial",main=NA) model1 <- Arima(data1,c(1,0,0),c(0,1,1)) model2 <- Arima(data1,c(0,0,3),c(0,1,1)) model3 <- Arima(data1,c(0,0,2),c(0,1,1)) modelsummary(list("ARIMA(1,0,0)(0,1,1)"=model1,  "ARIMA(0,0,3)(0,1,1)"=model2,  "ARIMA(0,0,11)(0,1,1)"=model3),  stars = T,gof\_map = c("rmse","bic"),output = "flextable") %>%  save\_as\_docx(path = "临时.docx") map(list(model1,model2,model3),~.$residuals) %>%  map(Box.test,lag=12) tso(data) #干预 e <- model2 %>%  forecast(131) %>%  .$mean %>%  `-`(data[110:240]) e %>%  autoplot() +  theme\_bw() auto.arima(e) %>% modelsummary(stars = T,gof\_map = c("rmse","bic")) tso(data) out\_point <- data.frame(x1=c(rep(0,40),41,rep(0,199)),  x2=c(rep(0,59),60,rep(0,180)),  x3=c(rep(0,68),69,rep(0,171)),  x4=c(rep(0,79),80,rep(0,160)),  x5=c(rep(0,226),227,rep(0,13))) %>%  as.matrix() model4 <- Arima(data,c(1,0,0),c(0,1,1),xreg = out\_point) #完整建模 all\_d <- diff(data) %>%  diff(lag = 12) acf(all\_d,lag.max = 50) pacf(all\_d,lag.max = 50) model5 <- Arima(data,c(1,0,0),c(0,1,1)) modelsummary(model5,stars = T,gof\_map = c("rmse","bic"),output = "flextable") %>%  save\_as\_docx(path = "临时.docx") #预测 model5 %>%  forecast(48) %>%  autoplot(PI=F) +  labs(title = NULL) +  theme\_bw() |