**云南大学数学与统计学院**

**实验报告**

**实验课名称： 应用时间序列分析实验**

**指导教师： 周建军**

**专业（年级）： 2021级统计学**

**学生姓名： 枫叶 学号:**

**实验名称： 非平稳时间序列建模**

**实验时间： 2024.5.21**

**实验成绩：**

1. **实验目的和要求：**

掌握单位根检验、arima模型；掌握疏稀疏模型

1. **实验内容和原理**

（1）对对数石油价格数据（oil.price）建立一个合适的ARIMA时间序列模型。该数据包含在TSA程序包中，为198601-200601期间每月的原油价格。

Data(oil.price)

1. 第四章的习题5（P157）:1867-1938年英国绵羊数量。确定序列的平稳性，建立合适的时间序列模型并预测1939-1945年的绵羊数。

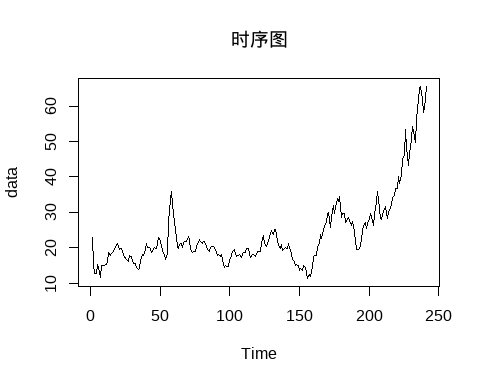
（3）基于中美两国的GDP数据，建模分析中国是否能超越美国，成为世界第一大经济体。如果能，请说明在哪一年能实现这一目标。

1. **实验步骤及方法（包含具体的程序）**

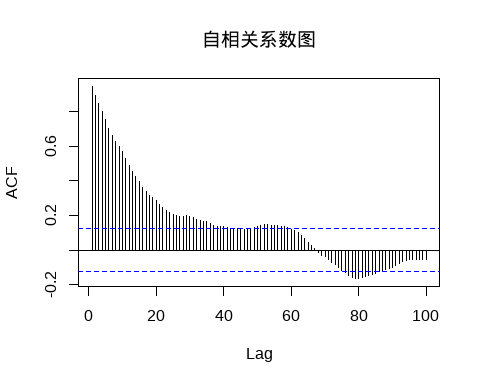
**library**(tidyverse)**library**(forecast)**library**(TSA)**library**(urca)**library**(modelsummary)**library**(readxl)**library**(patchwork)

**第一题**

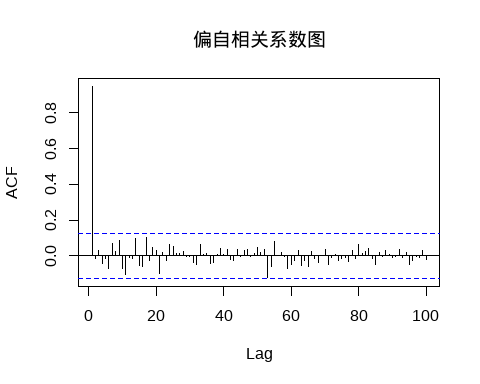
**data**("oil.price")data <- **ts**(oil.price,frequency = 1)**plot**(data,main="时序图")



**acf**(data,lag.max = 100,main="自相关系数图")

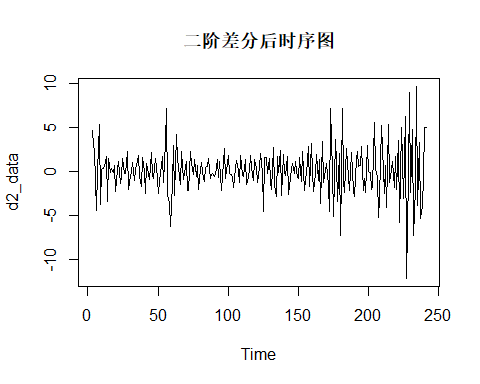


**acf**(data,type = "partial",lag.max = 100,main="偏自相关系数图")

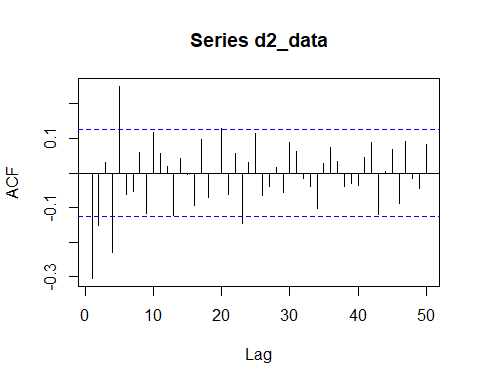


时序图显示出明显的趋势，ACF显示出长期相关性，不过不能确定趋势为线性还是曲线型，1999年科索沃战争与OPEC减产协议导致国际油价上涨，2001年起石油需求剧增、美元走弱、伊拉克战争等因素也促使国际油价大幅上涨，各个时期的数据是否可视作同一总体有待商榷，这里暂且对其进行统一建模

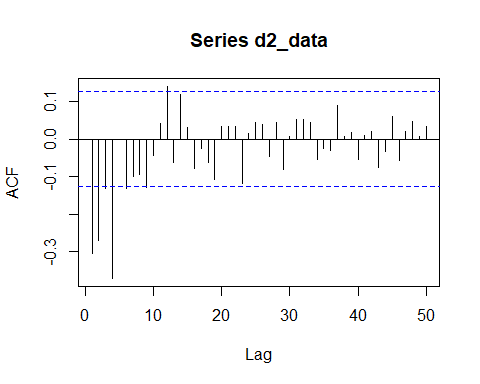
d2\_data <- **diff**(data,differences = 2)**plot**(d2\_data,main="二阶差分后时序图")



**acf**(d2\_data,lag.max = 50)



**acf**(d2\_data,type = "partial",lag.max =50)



经过尝试，二阶差分时在很大程度上消除了长期趋势，但个别年份波动较大，从时序图来看除了个别年份波动较大，整体来说围绕零波动，ACF图和PACF图没有显示出明显的拖尾或截尾特征，先做单位根检验

**ur.df**(d2\_data,type = "trend",lags = 10) **%>%** **summary**()

## ## ############################################### ## # Augmented Dickey-Fuller Test Unit Root Test # ## ############################################### ## ## Test regression trend ## ## ## Call:## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)## ## Residuals:## Min 1Q Median 3Q Max ## -6.6642 -1.0211 0.0383 1.0899 8.0362 ## ## Coefficients:## Estimate Std. Error t value Pr(>|t|) ## (Intercept) -0.129602 0.283173 -0.458 0.647646 ## z.lag.1 -8.002181 0.835396 -9.579 < 2e-16 \*\*\*## tt 0.001323 0.002013 0.657 0.511631 ## z.diff.lag1 6.258522 0.798515 7.838 2.10e-13 \*\*\*## z.diff.lag2 5.397827 0.743702 7.258 7.06e-12 \*\*\*## z.diff.lag3 4.565697 0.673278 6.781 1.13e-10 \*\*\*## z.diff.lag4 3.572257 0.591115 6.043 6.59e-09 \*\*\*## z.diff.lag5 2.847092 0.496928 5.729 3.38e-08 \*\*\*## z.diff.lag6 2.099752 0.404870 5.186 4.95e-07 \*\*\*## z.diff.lag7 1.431344 0.307392 4.656 5.63e-06 \*\*\*## z.diff.lag8 0.845743 0.218951 3.863 0.000148 \*\*\*## z.diff.lag9 0.359070 0.136665 2.627 0.009225 \*\* ## z.diff.lag10 0.092731 0.069610 1.332 0.184221 ## ---## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1## ## Residual standard error: 1.997 on 215 degrees of freedom## Multiple R-squared: 0.7974, Adjusted R-squared: 0.7861 ## F-statistic: 70.53 on 12 and 215 DF, p-value: < 2.2e-16## ## ## Value of test-statistic is: -9.5789 30.6222 45.9178 ## ## Critical values for test statistics: ## 1pct 5pct 10pct## tau3 -3.99 -3.43 -3.13## phi2 6.22 4.75 4.07## phi3 8.43 6.49 5.47

ADF检验结果显示可以认为不存在单位根，但仍存在趋势项，故再做二阶差分

d4\_data <- **diff**(data,differences = 2)**ur.df**(d4\_data,type = "trend",lags = 10) **%>%** **summary**()

## ## ############################################### ## # Augmented Dickey-Fuller Test Unit Root Test # ## ############################################### ## ## Test regression trend ## ## ## Call:## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)## ## Residuals:## Min 1Q Median 3Q Max ## -6.6642 -1.0211 0.0383 1.0899 8.0362 ## ## Coefficients:## Estimate Std. Error t value Pr(>|t|) ## (Intercept) -0.129602 0.283173 -0.458 0.647646 ## z.lag.1 -8.002181 0.835396 -9.579 < 2e-16 \*\*\*## tt 0.001323 0.002013 0.657 0.511631 ## z.diff.lag1 6.258522 0.798515 7.838 2.10e-13 \*\*\*## z.diff.lag2 5.397827 0.743702 7.258 7.06e-12 \*\*\*## z.diff.lag3 4.565697 0.673278 6.781 1.13e-10 \*\*\*## z.diff.lag4 3.572257 0.591115 6.043 6.59e-09 \*\*\*## z.diff.lag5 2.847092 0.496928 5.729 3.38e-08 \*\*\*## z.diff.lag6 2.099752 0.404870 5.186 4.95e-07 \*\*\*## z.diff.lag7 1.431344 0.307392 4.656 5.63e-06 \*\*\*## z.diff.lag8 0.845743 0.218951 3.863 0.000148 \*\*\*## z.diff.lag9 0.359070 0.136665 2.627 0.009225 \*\* ## z.diff.lag10 0.092731 0.069610 1.332 0.184221 ## ---## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1## ## Residual standard error: 1.997 on 215 degrees of freedom## Multiple R-squared: 0.7974, Adjusted R-squared: 0.7861 ## F-statistic: 70.53 on 12 and 215 DF, p-value: < 2.2e-16## ## ## Value of test-statistic is: -9.5789 30.6222 45.9178 ## ## Critical values for test statistics: ## 1pct 5pct 10pct## tau3 -3.99 -3.43 -3.13## phi2 6.22 4.75 4.07## phi3 8.43 6.49 5.47

ADF检验仍显示存在趋势项，怀疑确定性部分可能不是时间的多项式函数，为了不进一步扩大方差，只能根据二阶差分的自相关系数五阶截尾和偏自相关系数四阶截尾特征，分别尝试用ARI(4,2)，和IMA(2,5)进行拟合，同时也尝试拟合低阶的ARIMA模型

order\_list <- **list**(**c**(4,2,0),**c**(0,2,5),**c**(1,2,1),**c**(2,2,1),**c**(1,2,2),**c**(2,2,2))models <- **map**(order\_list,Arima,y=data)**modelsummary**(models,gof\_map = **c**("bic","rmse"),stars = T)

|  | IAR(4,2) | IMA(2,5) | ARIMA(1,2,1) | ARIMA(2,2,1) | ARIMA(1,2,2) | ARIMA(2,2,2) |
| --- | --- | --- | --- | --- | --- | --- |
| ar1 | -0.532\*\*\* |  | 0.194\*\* | 0.216\*\* | -0.528\*\*\* | -0.529\*\*\* |
|  | (0.061) |  | (0.067) | (0.067) | (0.112) | (0.111) |
| ar2 | -0.537\*\*\* |  |  | -0.161\* |  | 0.012 |
|  | (0.068) |  |  | (0.067) |  | (0.077) |
| ar3 | -0.387\*\*\* |  |  |  |  |  |
|  | (0.068) |  |  |  |  |  |
| ar4 | -0.426\*\*\* |  |  |  |  |  |
|  | (0.061) |  |  |  |  |  |
| ma1 |  | -0.722\*\*\* | -0.986\*\*\* | -0.982\*\*\* | -0.220\*\* | -0.214\* |
|  |  | (0.066) | (0.014) | (0.015) | (0.081) | (0.088) |
| ma2 |  | -0.358\*\*\* |  |  | -0.752\*\*\* | -0.759\*\*\* |
|  |  | (0.076) |  |  | (0.079) | (0.087) |
| ma3 |  | 0.002 |  |  |  |  |
|  |  | (0.090) |  |  |  |  |
| ma4 |  | -0.126+ |  |  |  |  |
|  |  | (0.069) |  |  |  |  |
| ma5 |  | 0.232\*\*\* |  |  |  |  |
|  |  | (0.060) |  |  |  |  |
| BIC | 1081.2 | 1052.5 | 1056.2 | 1056.0 | 1052.7 | 1058.2 |
| RMSE | 2.18 | 2.02 | 2.11 | 2.08 | 2.07 | 2.07 |
| * p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001 | | | | | | |

可以看到IAR(4,2)模型的系数全部显著，IMA(2,5)模型的不显著，ARIMA模型中仅ARIMA(2,2,2)的不显著

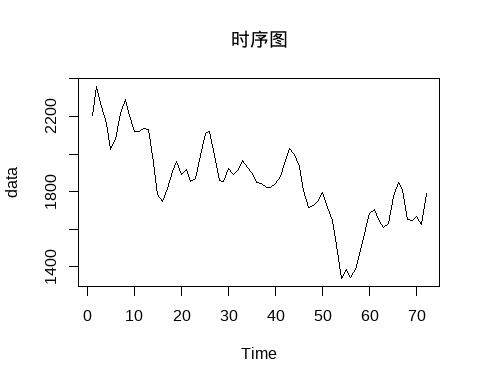
**map**(models,**~**.**$**residuals) **%>%** **map**(Box.test,lag=5)

## [[1]]## ## Box-Pierce test## ## data: .x[[i]]## X-squared = 10.712, df = 5, p-value = 0.0574## ## ## [[2]]## ## Box-Pierce test## ## data: .x[[i]]## X-squared = 3.4002, df = 5, p-value = 0.6385## ## ## [[3]]## ## Box-Pierce test## ## data: .x[[i]]## X-squared = 19.055, df = 5, p-value = 0.001877## ## ## [[4]]## ## Box-Pierce test## ## data: .x[[i]]## X-squared = 17.823, df = 5, p-value = 0.003176## ## ## [[5]]## ## Box-Pierce test## ## data: .x[[i]]## X-squared = 11.107, df = 5, p-value = 0.0493## ## ## [[6]]## ## Box-Pierce test## ## data: .x[[i]]## X-squared = 11.057, df = 5, p-value = 0.05026

对各模型的残差序列作纯随机性检验，可以看到只有IMA(2,5)可以认为残差序列是纯随机的，结合BIC来看，IMA(2,5)的BIC也是最小的，故选择IMA(2,5)来进行拟合

## 第二题

data <- read\_xlsx("D:/预删除文件夹/大三下/时间序列/时间序列第五章习题三绵羊数据.xlsx",col\_names = F) %>%  
 as.matrix() %>%  
 t() %>%  
 as.vector() %>%  
 ts()  
plot(data,main="时序图")



ur.df(data,type = "trend",lags = 3) %>% summary()

##   
## ###############################################   
## # Augmented Dickey-Fuller Test Unit Root Test #   
## ###############################################   
##   
## Test regression trend   
##   
##   
## Call:  
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -166.104 -34.910 5.548 33.817 188.323   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 299.84046 165.09750 1.816 0.0742 .   
## z.lag.1 -0.14716 0.07545 -1.950 0.0556 .   
## tt -0.87484 0.76930 -1.137 0.2598   
## z.diff.lag1 0.53931 0.12890 4.184 9.19e-05 \*\*\*  
## z.diff.lag2 -0.17673 0.13921 -1.270 0.2090   
## z.diff.lag3 -0.17560 0.12607 -1.393 0.1686   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 66.76 on 62 degrees of freedom  
## Multiple R-squared: 0.3907, Adjusted R-squared: 0.3416   
## F-statistic: 7.951 on 5 and 62 DF, p-value: 7.727e-06  
##   
##   
## Value of test-statistic is: -1.9505 1.7179 2.3425   
##   
## Critical values for test statistics:   
## 1pct 5pct 10pct  
## tau3 -4.04 -3.45 -3.15  
## phi2 6.50 4.88 4.16  
## phi3 8.73 6.49 5.47

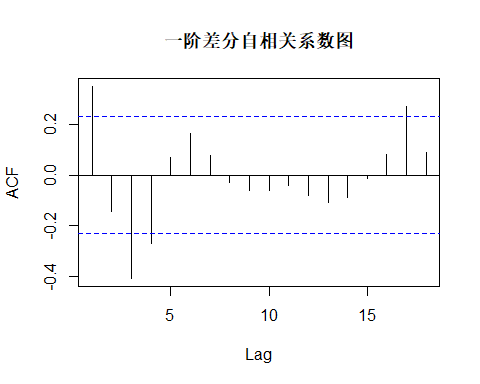
从时序图来看绵羊数存在稍微的下降趋势，首先检验序列平稳性，从ADF检验结果来看存在单位根，考虑一阶差分

d1\_data <- diff(data,differences = 1)  
ur.df(d1\_data,type = "trend",lags = 3) %>% summary()

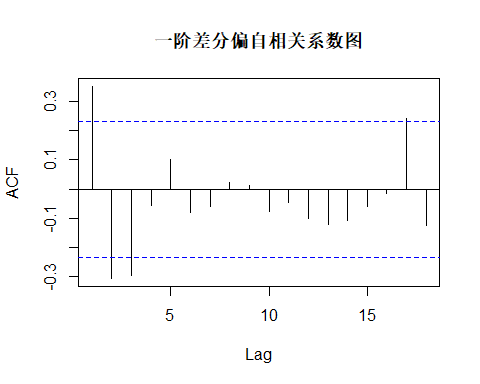
##   
## ###############################################   
## # Augmented Dickey-Fuller Test Unit Root Test #   
## ###############################################   
##   
## Test regression trend   
##   
##   
## Call:  
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -161.755 -35.530 6.604 43.744 172.771   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -15.82499 18.43477 -0.858 0.39402   
## z.lag.1 -1.06812 0.20972 -5.093 3.64e-06 \*\*\*  
## tt 0.29908 0.43617 0.686 0.49550   
## z.diff.lag1 0.54999 0.15988 3.440 0.00105 \*\*   
## z.diff.lag2 0.24519 0.14392 1.704 0.09354 .   
## z.diff.lag3 0.05553 0.12571 0.442 0.66026   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 68.62 on 61 degrees of freedom  
## Multiple R-squared: 0.4431, Adjusted R-squared: 0.3975   
## F-statistic: 9.707 on 5 and 61 DF, p-value: 7.419e-07  
##   
##   
## Value of test-statistic is: -5.0931 8.7474 12.9913   
##   
## Critical values for test statistics:   
## 1pct 5pct 10pct  
## tau3 -4.04 -3.45 -3.15  
## phi2 6.50 4.88 4.16  
## phi3 8.73 6.49 5.47

对一阶差分后的序列作ADF检验，可以看到此时可以认为不存在单位根，但趋势项趋势项显著，可能确定性趋势并非简单的线性模式，或者检验受到了异常值影响，下面考察一阶差分后的ADF图和PACF图

acf(d1\_data,main="一阶差分自相关系数图")



acf(d1\_data,type = "partial",main="一阶差分偏自相关系数图")



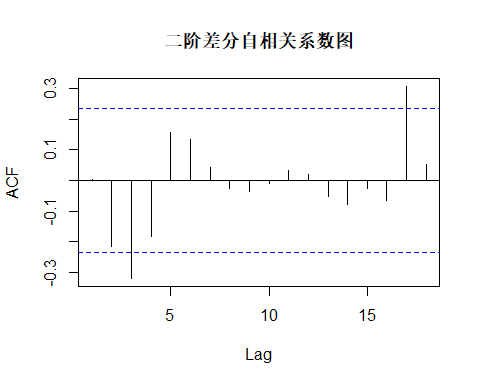
ur.df(d1\_data,lags = 5,type = "trend") %>% summary()

##   
## ###############################################   
## # Augmented Dickey-Fuller Test Unit Root Test #   
## ###############################################   
##   
## Test regression trend   
##   
##   
## Call:  
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -146.663 -38.765 9.884 36.429 163.458   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -26.11219 19.89505 -1.312 0.194613   
## z.lag.1 -1.05133 0.27964 -3.760 0.000403 \*\*\*  
## tt 0.51178 0.46167 1.109 0.272285   
## z.diff.lag1 0.55275 0.25415 2.175 0.033799 \*   
## z.diff.lag2 0.26954 0.22410 1.203 0.234044   
## z.diff.lag3 0.12860 0.17627 0.730 0.468667   
## z.diff.lag4 -0.03501 0.15317 -0.229 0.820038   
## z.diff.lag5 0.08701 0.13241 0.657 0.513775   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 68.53 on 57 degrees of freedom  
## Multiple R-squared: 0.4334, Adjusted R-squared: 0.3638   
## F-statistic: 6.228 on 7 and 57 DF, p-value: 1.946e-05  
##   
##   
## Value of test-statistic is: -3.7596 4.8012 7.1963   
##   
## Critical values for test statistics:   
## 1pct 5pct 10pct  
## tau3 -4.04 -3.45 -3.15  
## phi2 6.50 4.88 4.16  
## phi3 8.73 6.49 5.47

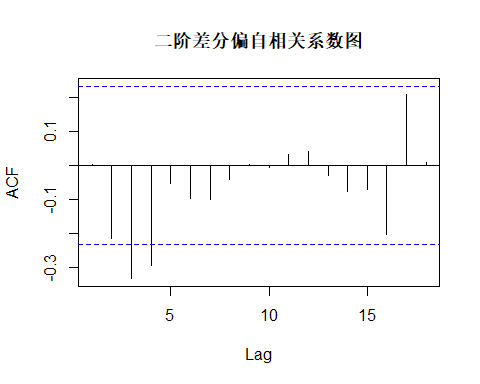
d2\_data <- diff(data,differences = 2)  
ur.df(d2\_data,lags = 5,type = "trend") %>% summary()

##   
## ###############################################   
## # Augmented Dickey-Fuller Test Unit Root Test #   
## ###############################################   
##   
## Test regression trend   
##   
##   
## Call:  
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -160.991 -57.912 6.386 50.907 174.291   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -9.5470 21.6473 -0.441 0.660893   
## z.lag.1 -3.0707 0.5659 -5.427 1.28e-06 \*\*\*  
## tt 0.2740 0.5181 0.529 0.599051   
## z.diff.lag1 1.7806 0.4939 3.605 0.000664 \*\*\*  
## z.diff.lag2 1.2706 0.3998 3.178 0.002416 \*\*   
## z.diff.lag3 0.8023 0.2992 2.682 0.009605 \*\*   
## z.diff.lag4 0.3434 0.2194 1.565 0.123198   
## z.diff.lag5 0.1725 0.1348 1.280 0.205981   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 76.12 on 56 degrees of freedom  
## Multiple R-squared: 0.6349, Adjusted R-squared: 0.5893   
## F-statistic: 13.91 on 7 and 56 DF, p-value: 2.731e-10  
##   
##   
## Value of test-statistic is: -5.4266 9.981 14.9562   
##   
## Critical values for test statistics:   
## 1pct 5pct 10pct  
## tau3 -4.04 -3.45 -3.15  
## phi2 6.50 4.88 4.16  
## phi3 8.73 6.49 5.47

acf(d2\_data,main="二阶差分自相关系数图")



acf(d2\_data,type = "partial",main="二阶差分偏自相关系数图")



发现一阶差分时样本存在长期自相关，且ADF检验结果显示存在单位根，经尝试，二阶差分后可以认为单位根不显著，故结合二阶差分下的ACF图和PACF图，分别尝试用ARI(3,2)和IMA(2,3)进行拟合

models <- map(list(c(3,2,0),c(0,2,3)),Arima,y=data)  
modelsummary(models,stars = T,gof\_map = c("bic","rmse"))

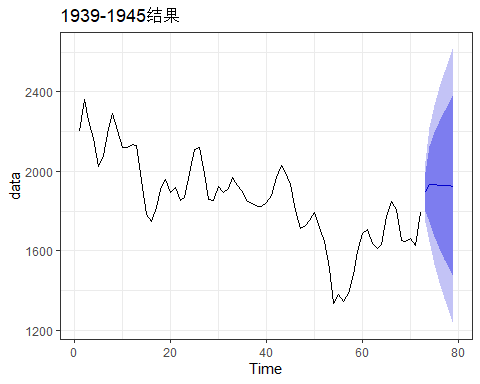
|  | ARI(3,2) | IMA(2,3) |
| --- | --- | --- |
| ar1 | -0.114 |  |
|  | (0.124) |  |
| ar2 | -0.239\* |  |
|  | (0.117) |  |
| ar3 | -0.402\*\*\* |  |
|  | (0.122) |  |
| ma1 |  | -0.403\*\* |
|  |  | (0.135) |
| ma2 |  | -0.393\* |
|  |  | (0.184) |
| ma3 |  | -0.204 |
|  |  | (0.195) |
| BIC | 834.4 | 824.1 |
| RMSE | 81.54 | 74.24 |
| * p < 0.1, \* p < 0.05,  **p < 0.01,** \* p < 0.001 | | |

map(models,~.$residuals) %>%  
 map(Box.test,lag=3)

## [[1]]  
##   
## Box-Pierce test  
##   
## data: .x[[i]]  
## X-squared = 1.2508, df = 3, p-value = 0.7409  
##   
##   
## [[2]]  
##   
## Box-Pierce test  
##   
## data: .x[[i]]  
## X-squared = 3.3775, df = 3, p-value = 0.337

可以看到两个模型系数均显著，对残差序列做纯随机性检验，可以看到均不能拒绝纯随机性假设，二者都在一定程度上提取了序列信息，从BIC来看，IMA(2,3)的BIC更小，故选择IMA(2,3)来进行拟合

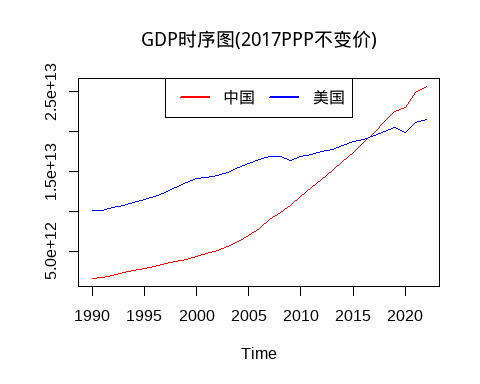
forecast(Arima(data,c(0,2,3)),7) %>%  
 autoplot() +  
 labs(title = "1939-1945结果") +  
 theme\_bw()



七年预测结果如上图所示，不过众所周知的是1939-1945爆发了第二次世界大战，英国绵羊数量想必会受到冲击

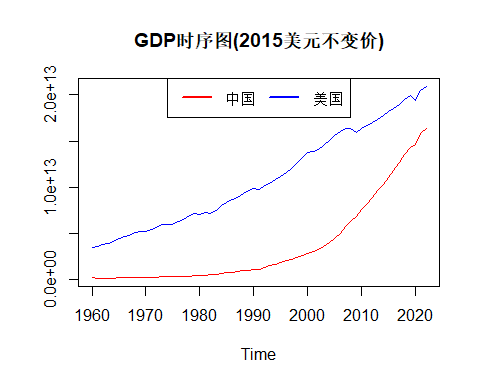
**第三题**

data\_PPP <- **read\_xls**("D:/预删除文件夹/大三下/时间序列/中美GDP2017PPP不变价.xls") **%>%** **ts**(start = 1990)**ts.plot**(data\_PPP,gpars=**list**(col=**c**("red","blue")),main="GDP时序图(2017PPP不变价)")**legend**("top",legend=**c**("中国","美国"),col=**c**("red","blue"),lty=1,lwd=2,ncol=2)



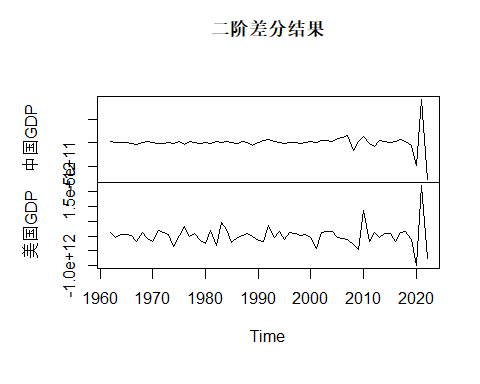
数据获取自世界银行，由于涉及多年的国际间比较，理论上来说应当选择购买力平价进行折算，如上图所示，按实际购买力来看，中国的GDP早在2017年就已超越美国，但由于题目要求估计中国超越美国的时间，故下面用基于2015年美元不变价折算的数据进行建模

data\_dollar <- **read\_xls**("D:/预删除文件夹/大三下/时间序列/中美GDP2015美元不变价.xls") **%>%** **ts**(start = 1960)**ts.plot**(data\_dollar,gpars=**list**(col=**c**("red","blue")),main="GDP时序图(2015美元不变价)")**legend**("top",legend=**c**("中国","美国"),col=**c**("red","blue"),lty=1,lwd=2,ncol=2)



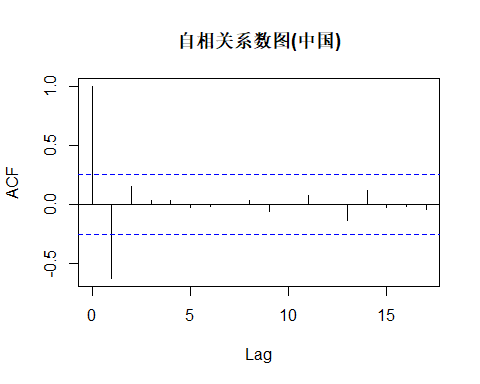
中美GDP时序图如上图所示，显然都不为平稳序列，下面尝试二阶差分

d2\_data <- **diff**(data\_dollar,differences = 2)**plot**(d2\_data,main="二阶差分结果")

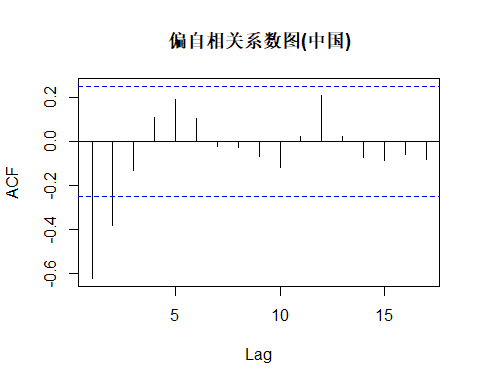


可以看到中美GDP经二阶差分后直观上来看均较为平稳，仅2020年受新冠疫情冲击有剧烈波动，下面分别尝试进行拟合

*#中国*d2\_data\_1 <- d2\_data[1**:**61,1]**acf**(d2\_data\_1,drop.lag.0 = F,main="自相关系数图(中国)")



**acf**(d2\_data\_1,type = "partial",main="偏自相关系数图(中国)")



从ACF图和PACF图来看，自相关系数一阶截尾，偏自相关系数呈现出一定的拖尾特征，也可以认为是二阶截尾，为保险起见先进行单位根检验

**ur.df**(d2\_data\_1,type = "trend") **%>%** **summary**()

## ## ############################################### ## # Augmented Dickey-Fuller Test Unit Root Test # ## ############################################### ## ## Test regression trend ## ## ## Call:## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)## ## Residuals:## Min 1Q Median 3Q Max ## -5.894e+11 -3.150e+10 -4.030e+09 3.075e+10 2.728e+11 ## ## Coefficients:## Estimate Std. Error t value Pr(>|t|) ## (Intercept) -6.400e+07 2.965e+10 -0.002 0.9983 ## z.lag.1 -2.642e+00 3.299e-01 -8.008 8.46e-11 \*\*\*## tt 1.128e+09 8.432e+08 1.338 0.1864 ## z.diff.lag 5.027e-01 2.325e-01 2.163 0.0349 \* ## ---## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1## ## Residual standard error: 1.09e+11 on 55 degrees of freedom## Multiple R-squared: 0.8735, Adjusted R-squared: 0.8666 ## F-statistic: 126.6 on 3 and 55 DF, p-value: < 2.2e-16## ## ## Value of test-statistic is: -8.0082 21.6519 32.4018 ## ## Critical values for test statistics: ## 1pct 5pct 10pct## tau3 -4.04 -3.45 -3.15## phi2 6.50 4.88 4.16## phi3 8.73 6.49 5.47

ADF检验结果显示不存在单位根，但存在趋势项，经尝试，多次差分反而使得趋势项更显著，可能确定性部分并非时间的多项式函数，下面尝试用IMA(2,1)进行拟合

model\_zh <- **Arima**(data\_dollar[1**:**61,1],**c**(0,2,1))**modelsummary**(model\_zh,stars = T,gof\_map = **c**("bic"))

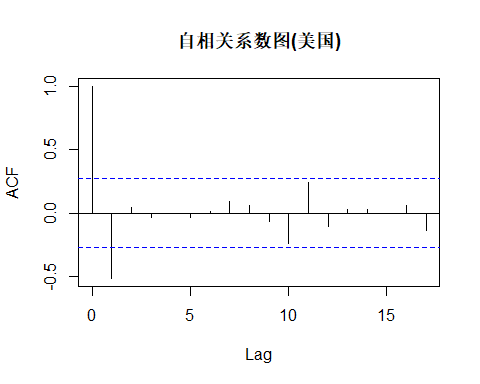
|  | IMA(2,1) |
| --- | --- |
| ma1 | 0.435\*\* |
|  | (0.167) |
| BIC | 3132.5 |
| * p < 0.1,  *p < 0.05,* *p < 0.01, p* < 0.001 | |

**Box.test**(model\_zh**$**residuals,lag = 2,type = "Ljung-Box")

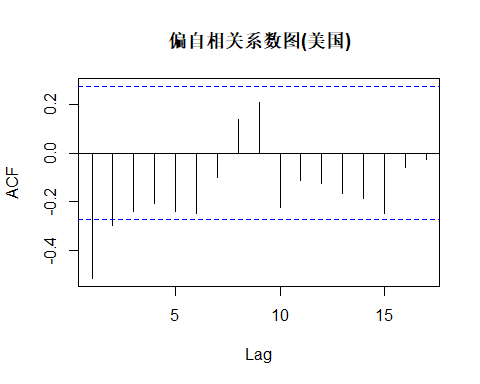
## ## Box-Ljung test## ## data: model\_zh$residuals## X-squared = 0.32212, df = 2, p-value = 0.8512

系数显著，对残差序列作纯随机性检验，结果显示可以认为残差是纯随机的，IMA(2,1)较好地提取了序列信息

*#美国*d2\_data\_2 <- d2\_data[1**:**61,2]d2\_data\_2 <- **diff**(d2\_data\_2,lag = 10)**acf**(d2\_data\_2,drop.lag.0 = F,main="自相关系数图(美国)")



**acf**(d2\_data\_2,type = "partial",main="偏自相关系数图(美国)")



**ur.df**(d2\_data\_2,type = "trend",lags = 5) **%>%** **summary**()

## ## ############################################### ## # Augmented Dickey-Fuller Test Unit Root Test # ## ############################################### ## ## Test regression trend ## ## ## Call:## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)## ## Residuals:## Min 1Q Median 3Q Max ## -1.096e+12 -8.238e+10 1.877e+10 1.387e+11 7.697e+11 ## ## Coefficients:## Estimate Std. Error t value Pr(>|t|) ## (Intercept) -3.213e+10 1.401e+11 -0.229 0.81981 ## z.lag.1 -5.071e+00 9.829e-01 -5.159 8.62e-06 \*\*\*## tt 2.070e+09 4.568e+09 0.453 0.65299 ## z.diff.lag1 3.106e+00 8.878e-01 3.499 0.00123 \*\* ## z.diff.lag2 2.311e+00 7.333e-01 3.151 0.00321 \*\* ## z.diff.lag3 1.540e+00 5.692e-01 2.706 0.01024 \* ## z.diff.lag4 9.167e-01 3.952e-01 2.320 0.02598 \* ## z.diff.lag5 3.812e-01 2.351e-01 1.621 0.11345 ## ---## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1## ## Residual standard error: 3.93e+11 on 37 degrees of freedom## Multiple R-squared: 0.835, Adjusted R-squared: 0.8037 ## F-statistic: 26.74 on 7 and 37 DF, p-value: 1.213e-12## ## ## Value of test-statistic is: -5.1589 8.8792 13.3188 ## ## Critical values for test statistics: ## 1pct 5pct 10pct## tau3 -4.04 -3.45 -3.15## phi2 6.50 4.88 4.16## phi3 8.73 6.49 5.47

美国的样本自相关系数和偏自相关系数分别在滞后11阶和10阶时显著不为零，出现了长期相关性，即使删除了受新冠疫情影响的2020年也依旧如此，考虑作10步差分，发现10步差分后长期自相关性消除，但没法解释其成因。此外ADF检验结果显示10步差分后不存在单位根，但趋势项显著不为零，目前已经遇到多次趋势项显著不为零的情况，但差分后反而会更显著，推测可能是存在非线性趋势，暂时不知道该如何处理此问题，下面分别尝试ARI(2,2),IMA(2,1)和ARIMA(1,1,1)

model\_us1 <- **Arima**(data\_dollar[1**:**61,2],**c**(2,2,0))model\_us2 <- **Arima**(data\_dollar[1**:**61,2],**c**(0,2,1))model\_us3 <- **Arima**(data\_dollar[1**:**61,2],**c**(1,2,1))**modelsummary**(**list**("ARI(2,2)"=model\_us1,"IMA(2,1)"=model\_us2,"ARIMA(1,1,1)"=model\_us3),stars = T,gof\_map = "bic")

|  | | ARI(2,2) | IMA(2,1) | | ARIMA(1,1,1) | |
| --- | --- | --- | --- | --- | --- | --- |
| ar1 | | -0.295+ |  | | 0.340+ | |
|  | | (0.152) |  | | (0.176) | |
| ar2 | | -0.247+ |  | |  | |
|  | | (0.149) |  | |  | |
| ma1 |  | | | -0.907\*\*\* | | -0.977\*\*\* |
|  |  | | | (0.074) | | (0.116) |
| BIC | 3270.0 | | | 3259.6 | | 3259.6 |
| * p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001 | | | | | | |

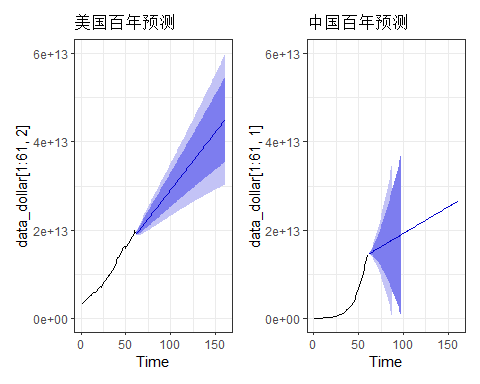
从回归结果来看，系数都是显著的，其中BIC相对较小的是IMA(2,1)和ARIMA(1,1,1)，进一步对残差序列做纯随机性检验

**map**(**list**(model\_us1,model\_us2,model\_us3),**~**.**$**residuals) **%>%** **map**(Box.test,lag=3,type="Ljung-Box")

## [[1]]## ## Box-Ljung test## ## data: .x[[i]]## X-squared = 3.6876, df = 3, p-value = 0.2972## ## ## [[2]]## ## Box-Ljung test## ## data: .x[[i]]## X-squared = 3.9129, df = 3, p-value = 0.271## ## ## [[3]]## ## Box-Ljung test## ## data: .x[[i]]## X-squared = 1.6126, df = 3, p-value = 0.6565

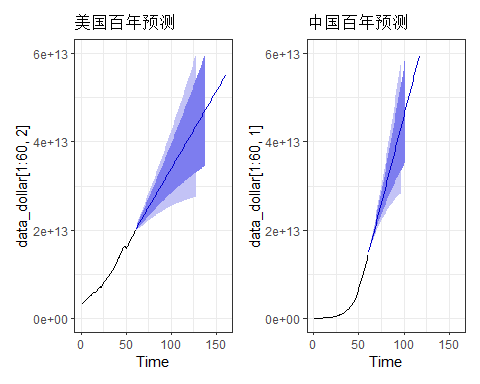
三个模型的残差序列都可以视作随机序列，其中ARIMA(1,1,1)的p值最大，结合BIC，不妨选择ARIMA(1,1,1)进行拟合

p1 <- **forecast**(model\_us3,100) **%>%** **autoplot**() **+** **ylim**(0,6e+13) **+** **labs**(title = "美国百年预测") **+** **theme\_bw**()p2 <- **forecast**(model\_zh,100) **%>%** **autoplot**() **+** **ylim**(0,6e+13) **+** **labs**(title = "中国百年预测") **+** **theme\_bw**()p1**|**p2



如果用IMA(2,1)拟合中国，ARIMA(1,2,1)拟合美国，那么中国将永远也赶不上美国，但是考虑到IMA模型受最后一期数据影响极大，而如果删去受新冠疫情影响的2020年，拟合结果如下

model\_zh <- **Arima**(data\_dollar[1**:**60,1],**c**(0,2,1))model\_us <- **Arima**(data\_dollar[1**:**60,2],**c**(1,2,1))p1 <- **forecast**(model\_us,100) **%>%** **autoplot**() **+** **ylim**(0,6e+13) **+** **labs**(title = "美国百年预测") **+** **theme\_bw**()p2 <- **forecast**(model\_zh,100) **%>%** **autoplot**() **+** **ylim**(0,6e+13) **+** **labs**(title = "中国百年预测") **+** **theme\_bw**()p1**|**p2



win <- 2019**+which.max**(**forecast**(model\_us,100)**$**mean **<** **forecast**(model\_zh,100)**$**mean)

可以看到中国将在2032年超越美国

1. **实验结果分析说明**

已附于图下