**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： 实验六**

**实验时间： 2024.6.24**

**实验成绩：**

1. **实验目的和要求：**

本节主要针对多元时间序列介绍伪相关问题，以及如何建立多元时间序列的回归模型-AIRMAX模型。主要介绍单位根检验、协整分析和误差修正模型。

1. **实验内容和原理**

作业：1、以TSA中的bulebird公司的一款薯片数据data(bluebird)为例，检验是否存在互相关性，如果存在建立合适的ARIMAX模型。

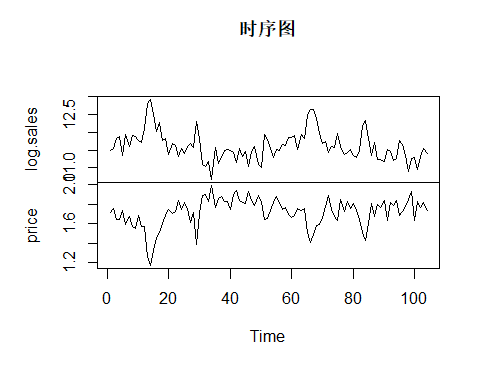
作业：2、基于我国1950-2008年进出口总额数据，考虑如下问题：

1. 使用单位根检验方法分别对进口和出口数据进行平稳性检验
2. 检验进口和出口序列间是否存在协整关系
3. 如果存在协整关系，则建立合适的模型拟合它们之间的关系
4. 如果存在协整关系，建立误差修正模型。
5. **实验步骤及方法（包含具体的程序）**

library(TSA)  
library(urca)  
library(dplyr)  
library(purrr)  
library(modelsummary)  
library(forecast)

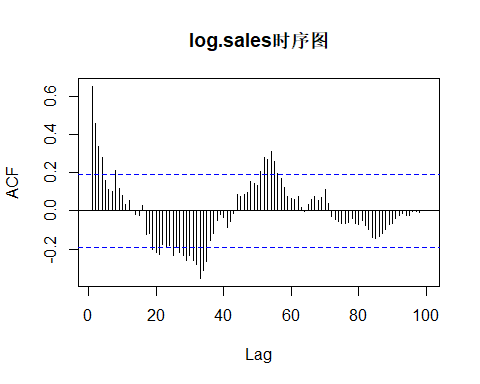
## 第一题

data("bluebird")  
plot(bluebird,main="时序图")

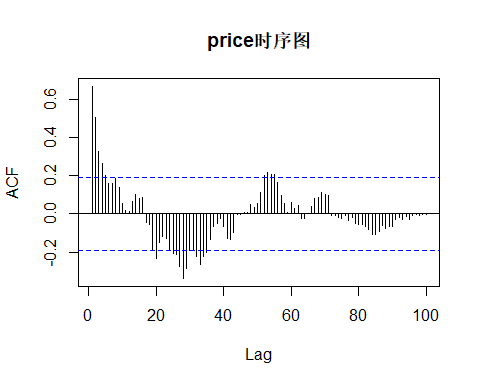


从时序图可以看出两个序列很可能有负相关关系，但为了避免伪相关问题，下面先考察二者的自相关结构

data1 <- bluebird[,1];data2 <- bluebird[,2]  
acf(data1,lag.max = 100,main="log.sales时序图")



acf(data2,lag.max = 100,main="price时序图")



可以看到二者都具有长期自相关性，虽然并非所有时点都有非零自相关系数，但也可不避免地增加了t统计量的方差，下面尝试作预白化处理来得到真实的相关关系，首先考察二者的平稳性

ur.df(data1,type = "drift",lags = 4) %>% summary()

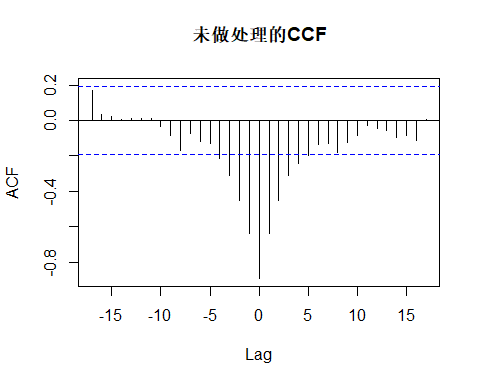
##   
## ###############################################   
## # Augmented Dickey-Fuller Test Unit Root Test #   
## ###############################################   
##   
## Test regression drift   
##   
##   
## Call:  
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.70594 -0.22842 -0.01764 0.15660 0.88945   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.706036 1.177205 3.148 0.00221 \*\*  
## z.lag.1 -0.318881 0.101191 -3.151 0.00219 \*\*  
## z.diff.lag1 -0.054328 0.120673 -0.450 0.65361   
## z.diff.lag2 -0.009012 0.115517 -0.078 0.93798   
## z.diff.lag3 -0.023406 0.110501 -0.212 0.83271   
## z.diff.lag4 0.107450 0.103117 1.042 0.30010   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3177 on 93 degrees of freedom  
## Multiple R-squared: 0.1854, Adjusted R-squared: 0.1416   
## F-statistic: 4.233 on 5 and 93 DF, p-value: 0.00164  
##   
##   
## Value of test-statistic is: -3.1513 4.9678   
##   
## Critical values for test statistics:   
## 1pct 5pct 10pct  
## tau2 -3.46 -2.88 -2.57  
## phi1 6.52 4.63 3.81

ur.df(data2,type = "drift",lags = 4) %>% summary()

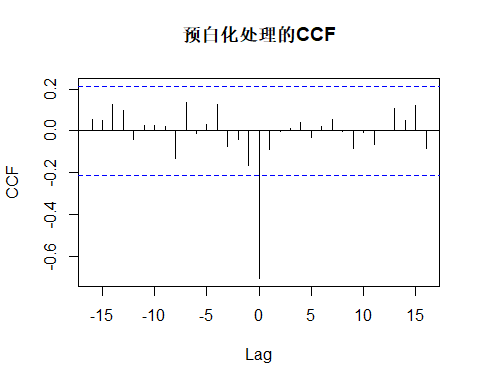
##   
## ###############################################   
## # Augmented Dickey-Fuller Test Unit Root Test #   
## ###############################################   
##   
## Test regression drift   
##   
##   
## Call:  
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.33589 -0.05001 0.01333 0.06890 0.23733   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.494640 0.169625 2.916 0.00444 \*\*  
## z.lag.1 -0.288083 0.098718 -2.918 0.00441 \*\*  
## z.diff.lag1 -0.092011 0.119265 -0.771 0.44238   
## z.diff.lag2 0.050497 0.113499 0.445 0.65741   
## z.diff.lag3 -0.084858 0.110237 -0.770 0.44338   
## z.diff.lag4 -0.006229 0.104442 -0.060 0.95257   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1124 on 93 degrees of freedom  
## Multiple R-squared: 0.1829, Adjusted R-squared: 0.139   
## F-statistic: 4.163 on 5 and 93 DF, p-value: 0.001858  
##   
##   
## Value of test-statistic is: -2.9182 4.2602   
##   
## Critical values for test statistics:   
## 1pct 5pct 10pct  
## tau2 -3.46 -2.88 -2.57  
## phi1 6.52 4.63 3.81

从时序图来看，两个序列都没有明显的趋势项，且围绕某一值波动，可以选择带截距项的模型进行ADF检验，检验结果如上所示，可以看到在0.1的显著性水平下拒绝单位根的存在，并且不拒绝非零截距项，故可以认为二者都是平稳序列

ccf(data2,data1,main="未做处理的CCF")



prewhiten(data2,data1,main="预白化处理的CCF")



预白化处理后的互相关系数仅有同期负相关性，其余滞后阶数都为零，这与时序图特征一致，而直接计算的CCF在各自滞后五阶之内都显著非零，若如此则时序图不应该呈现这么对称的分布，而且未来的销量或价格影响过去的价格或销量也是不太合理的。

下面进一步拟合ARIMAX模型，由于价格与销量是同期相关，不能判定二者的因果关系，从理论上来说，价格与销量也是互相决定的，从实用性的角度考虑，下面用销量作为被解释变量，价格作为解释变量

model1 <- auto.arima(data1)  
model2 <- lm(data1~data2)  
model3 <- Arima(data1,order = c(2,0,1),xreg = data2)  
modelsummary(list("ARMA"=model1,  
 "线性回归"=model2,  
 "ARMAX模型"=model3),stars = T,gof\_map = c("bic","rmse"),)

|  | ARMA | 线性回归 ARMA | |
| --- | --- | --- | --- |
| ar1 | -0.311\*\*\* |  | 0.807 |
|  | (0.078) |  | (0.153) |
| ar2 | 0.669\*\*\* |  | 0.105 |
|  | (0.072) |  | (0.116) |
| ma1 | 0.959\*\*\* |  | -0.713 |
|  | (0.051) |  | (0.123) |
| intercept | 11.615\*\*\* | 15.891\*\*\* | 15.874 |
|  | (0.088) | (0.217) | (0.238) |
| price |  | -2.489\*\*\* | -2.487 |
|  |  | (0.126) | (0.136) |
| BIC | 69.5 | -40.6 | -46.3 |
| RMSE | 0.30 | 0.19 | 0.17 |
| * p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001 | | | |

#modelsummary无法提取ARMAX的显著性，手动计算ARMAX的p值  
get\_estimates(model3) %>%  
 mutate(t=estimate/std.error,  
 p=2\*pt(abs(t),103,lower.tail = F)) %>%  
 select(c("term","p"))

## # A tibble: 5 × 2  
## term p  
## <chr> <dbl>  
## 1 ar1 7.28e- 7  
## 2 ar2 3.69e- 1  
## 3 ma1 8.13e- 8  
## 4 intercept 1.62e-86  
## 5 xreg 4.20e-34

分别使用单变量时间序列模型、线性回归模型和ARIMAX模型，结果如上表所示，，除了ARMAX模型中的二阶自回归系数在0.05显著性水平下不显著之外，三个模型的所有系数都是显著的。此外还可以注意到ARMAX模型的截距项以及价格的系数与线性回归中的值极为接近，可以说ARMAX同时兼顾了序列的时序信息和相关变量的作用，并且可以看到ARIMAX模型的BIC和RMSE都是最小的，从拟合的角度来说是最佳模型。

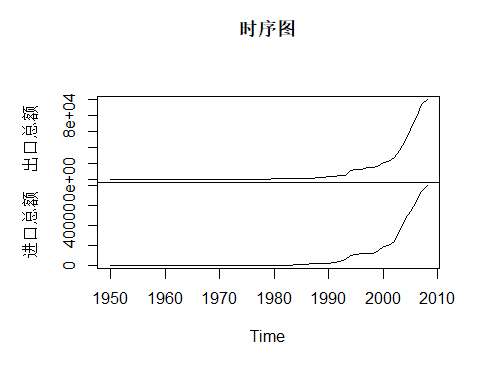
map(list(model1,model2,model3),~.$residuals) %>%  
 map(Box.test,lag=10)

## [[1]]  
##   
## Box-Pierce test  
##   
## data: .x[[i]]  
## X-squared = 8.7308, df = 10, p-value = 0.5578  
##   
##   
## [[2]]  
##   
## Box-Pierce test  
##   
## data: .x[[i]]  
## X-squared = 57.657, df = 10, p-value = 1.003e-08  
##   
##   
## [[3]]  
##   
## Box-Pierce test  
##   
## data: .x[[i]]  
## X-squared = 15.451, df = 10, p-value = 0.1165

在0.01的显著性水平上ARMA模型和ARMAX模型的残差都通过了纯随机性检验，可以认为它们较好地提取了序列信息，而线性回归模型拒绝纯随机性假设，并没有充分提取序列信息。

## 第二题

data <- read.table("D:/预删除文件夹/大三下/时间序列/进出口数据.txt",sep = "\t",header = T) %>%  
 select(-1) %>%  
 ts(start = 1950)  
export <- data[,1]  
import <- data[,2]  
plot(data,main="时序图")



ur.df(export,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   
## -6836.6 -540.2 -60.2 368.9 4163.0   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -631.87205 552.45848 -1.144 0.257869   
## z.lag.1 -0.02562 0.04899 -0.523 0.603094   
## tt 36.67893 19.33813 1.897 0.063321 .   
## z.diff.lag 0.92524 0.23422 3.950 0.000232 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1819 on 53 degrees of freedom  
## Multiple R-squared: 0.7981, Adjusted R-squared: 0.7867   
## F-statistic: 69.85 on 3 and 53 DF, p-value: < 2.2e-16  
##   
##   
## Value of test-statistic is: -0.5231 1.6026 1.7989   
##   
## 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

ur.df(import,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   
## -3456.3 -319.3 -66.7 242.1 5341.9   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -434.45562 427.46663 -1.016 0.314080   
## z.lag.1 0.03149 0.03146 1.001 0.321505   
## tt 25.54872 15.15690 1.686 0.097749 .   
## z.diff.lag 0.64024 0.16910 3.786 0.000392 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1397 on 53 degrees of freedom  
## Multiple R-squared: 0.7876, Adjusted R-squared: 0.7756   
## F-statistic: 65.52 on 3 and 53 DF, p-value: < 2.2e-16  
##   
##   
## Value of test-statistic is: 1.0007 2.2279 2.6274   
##   
## 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，显然两个序列都是不平稳的，ADF检验也得到了同样地结论，在0.1的显著性水平下不能拒绝单位根存在，可以认为序列不平稳

ur.df(diff(export,differences = 2)) %>% summary()

##   
## ###############################################   
## # Augmented Dickey-Fuller Test Unit Root Test #   
## ###############################################   
##   
## Test regression none   
##   
##   
## Call:  
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8648.7 -12.9 4.2 89.9 4500.1   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## z.lag.1 -1.2371 0.2562 -4.828 1.21e-05 \*\*\*  
## z.diff.lag 0.1153 0.1774 0.650 0.519   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1912 on 53 degrees of freedom  
## Multiple R-squared: 0.4326, Adjusted R-squared: 0.4112   
## F-statistic: 20.2 on 2 and 53 DF, p-value: 3.005e-07  
##   
##   
## Value of test-statistic is: -4.8282   
##   
## Critical values for test statistics:   
## 1pct 5pct 10pct  
## tau1 -2.6 -1.95 -1.61

ur.df(diff(import,differences = 2)) %>% summary()

##   
## ###############################################   
## # Augmented Dickey-Fuller Test Unit Root Test #   
## ###############################################   
##   
## Test regression none   
##   
##   
## Call:  
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3067.2 -13.7 10.8 127.2 4286.9   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## z.lag.1 -1.5023 0.1928 -7.792 2.41e-10 \*\*\*  
## z.diff.lag 0.4221 0.1341 3.148 0.0027 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1375 on 53 degrees of freedom  
## Multiple R-squared: 0.5808, Adjusted R-squared: 0.565   
## F-statistic: 36.71 on 2 and 53 DF, p-value: 9.879e-11  
##   
##   
## Value of test-statistic is: -7.7921   
##   
## Critical values for test statistics:   
## 1pct 5pct 10pct  
## tau1 -2.6 -1.95 -1.61

考虑到二者的曲线上升趋势，直接作二阶差分处理，ADF检验结果显示在0.01的显著性水平下拒绝单位根的存在，那么可以认为两个序列都是二阶单整的，下面对二者进行线性回归并检验残差的平稳性

model <- lm(进口总额~出口总额,data)  
summary(model)

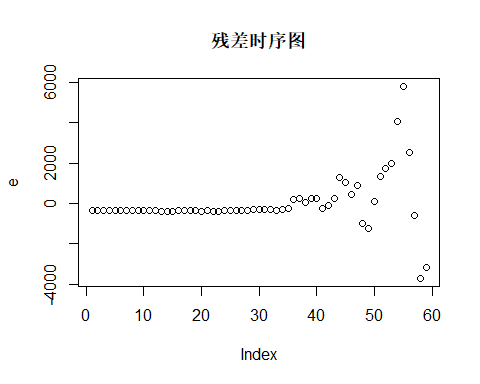
##   
## Call:  
## lm(formula = 进口总额 ~ 出口总额, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3713.9 -368.2 -355.0 167.5 5802.2   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.732e+02 1.899e+02 1.965 0.0543 .   
## 出口总额 8.199e-01 7.758e-03 105.684 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1327 on 57 degrees of freedom  
## Multiple R-squared: 0.9949, Adjusted R-squared: 0.9948   
## F-statistic: 1.117e+04 on 1 and 57 DF, p-value: < 2.2e-16

e <- model$residuals  
ur.df(e) %>% summary()

##   
## ###############################################   
## # Augmented Dickey-Fuller Test Unit Root Test #   
## ###############################################   
##   
## Test regression none   
##   
##   
## Call:  
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1918.0 -162.2 -151.2 136.3 2799.7   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## z.lag.1 -0.43298 0.08311 -5.210 2.92e-06 \*\*\*  
## z.diff.lag 0.67101 0.11516 5.827 3.06e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 698.8 on 55 degrees of freedom  
## Multiple R-squared: 0.4325, Adjusted R-squared: 0.4118   
## F-statistic: 20.95 on 2 and 55 DF, p-value: 1.718e-07  
##   
##   
## Value of test-statistic is: -5.2096   
##   
## Critical values for test statistics:   
## 1pct 5pct 10pct  
## tau1 -2.6 -1.95 -1.61

在0.01的显著性水平下拒绝单位根存在，残差序列是平稳的，故进口总额和出口总额存在协整关系，下面考察残差序列的时序特征

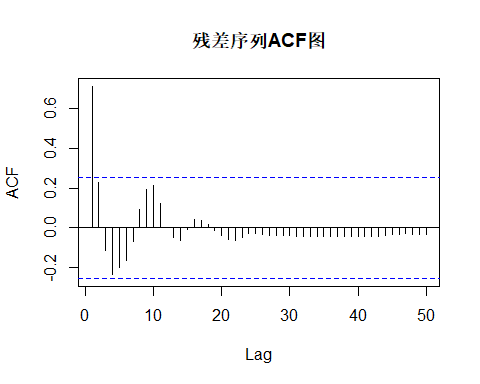
plot(e,main="残差时序图")



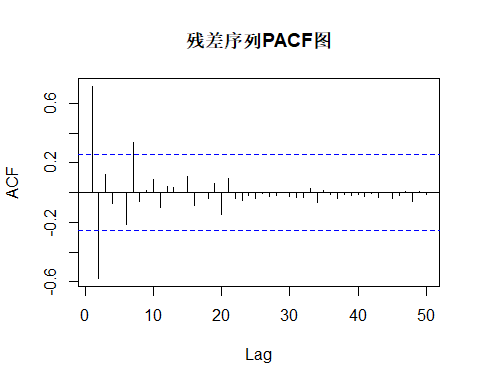
Box.test(e,lag = 10)

##   
## Box-Pierce test  
##   
## data: e  
## X-squared = 46.81, df = 10, p-value = 1.022e-06

acf(e,lag.max = 50,main="残差序列ACF图")



acf(e,lag.max = 50,type = "partial",main="残差序列PACF图")



首先，基于Q统计量对残差序列作纯随机性检验，p值极小，可以认为残差序列不是纯随机的，而从ACF图和PACF图来看，自相关系数一阶截尾，偏自相关系数没有明显的拖尾特征，可以尝试MA(1)、AR(2)、AR(6)模型。需要指出的是，从残差序列的时序图来看，其存在明显的异方差特征，为简单起见，这里不考虑对异方差进行拟合。

model1 <- Arima(e,c(0,0,1),include.mean = F)  
model2 <- Arima(e,c(2,0,0),include.mean = F)  
model3 <- Arima(e,c(6,0,0),include.mean = F)  
modelsummary(list(model1,model2,model3),stars = T,gof\_map = c("rmse","bic"))

|  | MA (1) | AR(2) | AR(6) |
| --- | --- | --- | --- |
| ma1 | 0.896\*\*\* |  |  |
|  | (0.079) |  |  |
| ar1 |  | 1.217\*\*\* | 1.225\*\*\* |
|  |  | (0.100) | (0.136) |
| ar2 |  | -0.646\*\*\* | -0.667\*\* |
|  |  | (0.106) | (0.233) |
| ar3 |  |  | -0.013 |
|  |  |  | (0.263) |
| ar4 |  |  | 0.168 |
|  |  |  | (0.286) |
| ar5 |  |  | -0.197 |
|  |  |  | (0.318) |
| ar6 |  |  | -0.004 |
|  |  |  | (0.215) |
| RMSE | 827.49 | 675.55 | 668.43 |
| BIC | 970.0 | 950.4 | 965.6 |
| * p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001 | | | |

Box.test(model2$residuals)

##   
## Box-Pierce test  
##   
## data: model2$residuals  
## X-squared = 0.17042, df = 1, p-value = 0.6797

从拟合的角度来说，AR(2)模型的系数均显著，RMSE和BIC值相对而言都较小，且通过了纯随机性检验，可以认为该模型较好地提取了残差序列的信息，那么最终的模型即

下面进一步拟合误差修正模型

ecm\_data <- data.frame(  
 y=diff(import),  
 x=diff(export),  
 ecm=e[1:58]  
)  
lm(y~x+ecm,ecm\_data) %>% summary()

##   
## Call:  
## lm(formula = y ~ x + ecm, data = ecm\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1921.80 -135.86 -125.79 -41.18 2956.85   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 118.16896 123.65699 0.956 0.343   
## x 0.72475 0.03511 20.641 <2e-16 \*\*\*  
## ecm -0.03017 0.10915 -0.276 0.783   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 834 on 55 degrees of freedom  
## Multiple R-squared: 0.9218, Adjusted R-squared: 0.9189   
## F-statistic: 324 on 2 and 55 DF, p-value: < 2.2e-16

即误差修正模型为

但可以注意到，误差修正系数并不显著

1. **实验结果分析说明**

已附于图下