**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： 2024.6.11**

**实验成绩：**

1. **实验目的和要求：**

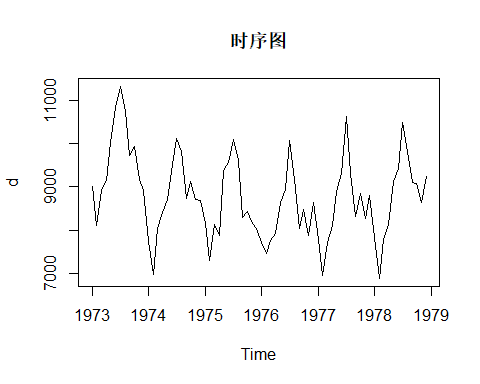
掌握掌握乘法季节模型；掌握ARCH、GARCH模型

1. **实验内容和原理**
2. 已知美国1973年1月-1978年12月的事故死亡数据（习题6.6.2），请对该数据建立合适的时间序列模型。
3. 已知1969年1月-1994年8月澳大利亚储备银行2年期有价证券月度利数据（按行排列）（见习题7.7.2），考察该数据是否存在条件异方差性，并对该序列建立合适的时间序列模型。
4. **实验步骤及方法（包含具体的程序）**

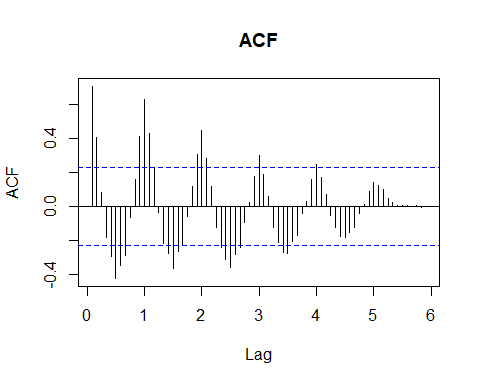
library(tidyverse)  
library(forecast)  
library(modelsummary)  
library(TSA)  
library(rugarch)

## 第一题

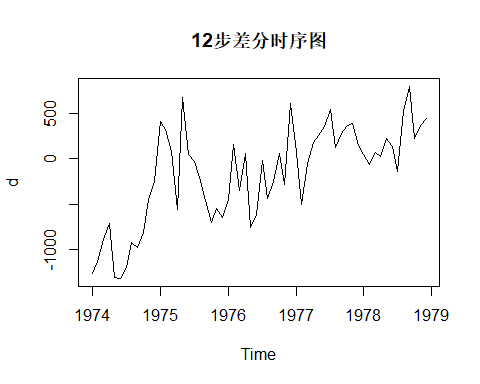
data <- read.table("D:/预删除文件夹/大三下/时间序列/月度死亡数据.txt",sep = "\t",header = T) %>%  
 pivot\_longer(everything(),  
 names\_to = c(".value", "id"),  
 names\_pattern = "(m|d)([1-3])") %>%  
 arrange(m) %>%  
 select(3) %>%  
 ts(start = 1973,frequency = 12)  
plot(data,main="时序图")



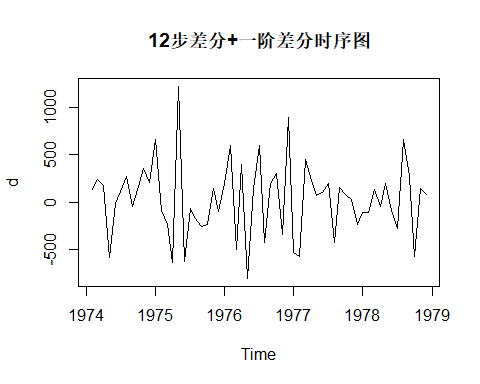
acf(data,main="ACF",lag.max = 80)



s12\_data <- diff(data,lag = 12)  
plot(s12\_data,main="12步差分时序图")



s12\_d1\_data <- diff(s12\_data)  
plot(s12\_d1\_data,main="12步差分+一阶差分时序图")



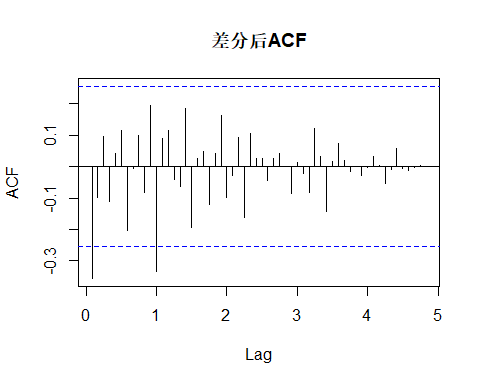
Box.test(s12\_d1\_data,lag = 12,type = "Ljung")

##   
## Box-Ljung test  
##   
## data: s12\_d1\_data  
## X-squared = 26.34, df = 12, p-value = 0.009605

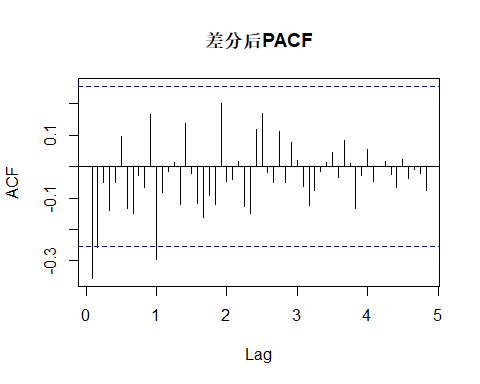
首先考察数据的时序图和ACF图，注意到其存在明显的季节性特征，考虑作12步差分，但发现12步差分后的时序图存在一定的线性上升趋势，故进一步作一阶差分，其时序图没有表现出明显的不平稳特征。稳健起见，对差分后的序列基于LB统计量作纯随机性检验，结果显示在0.05显著性水平下拒绝纯随机性假设。

下面进一步考察差分后序列的ACF和PACF

acf(s12\_d1\_data,main="差分后ACF",lag.max = 80)



acf(s12\_d1\_data,type = "partial",main="差分后PACF",lag.max = 80)



ACF和PACF均在一个周期长度上显著不为零，而在一个周期以内ACF在一阶滞后处显著不为零，PACF在一阶和二阶滞后处显著不为零，由于并不能确定明显的拖尾或截尾特征，下面对低阶SARIMA模型均作尝试，首先尝试P=0,Q=1

model1 <- Arima(data,c(0,1,1),c(0,1,1))  
model2 <- Arima(data,c(1,1,0),c(0,1,1))  
model3 <- Arima(data,c(1,1,1),c(0,1,1))  
model4 <- Arima(data,c(0,1,2),c(0,1,1))  
model5 <- Arima(data,c(2,1,0),c(0,1,1))  
model6 <- Arima(data,c(2,1,2),c(0,1,1))  
model7 <- Arima(data,c(1,1,2),c(0,1,1))  
model8 <- Arima(data,c(2,1,1),c(0,1,1))  
modelsummary(list("(0,1,1)[0,1,1]"=model1,  
 "(1,1,0)[0,1,1]"=model2,  
 "(1,1,1)[0,1,1]"=model3,  
 "(0,1,2)[0,1,1]"=model4,  
 "(2,1,0)[0,1,1]"=model5,  
 "(2,1,2)[0,1,1]"=model6,  
 "(1,1,2)[0,1,1]"=model7,  
 "(2,1,1)[1,1,1]"=model8),stars = T,gof\_map = c("bic","rmse"))

|  | (0,1,1)  [0,1,1] | (1,1,0)  [0,1,1] | (1,1,1)  [0,1,1] | (0,1,2)  [0,1,1] | (2,1,0)  [0,1,1] | (2,1,2)  [0,1,1] | (1,1,2)  [0,1,1] | (2,1,1)  [1,1,1] |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ma1 | -0.426\*\*\* |  | -0.497+ | -0.409\*\* |  | 0.570+ | 0.506\*\* | 0.488 |
|  | (0.123) |  | (0.284) | (0.134) |  | (0.330) | (0.193) | (0.561) |
| sma1 | -0.558\*\* | -0.596\*\* | -0.551\*\* | -0.549\*\* | -0.554\*\* | -0.548\*\* | -0.539\*\* | -0.583\*\* |
|  | (0.179) | (0.182) | (0.179) | (0.179) | (0.178) | (0.187) | (0.180) | (0.190) |
| ar1 |  | -0.334\*\* | 0.084 |  | -0.404\*\* | -0.941\*\* | -0.868\*\*\* | -0.866 |
|  |  | (0.122) | (0.320) |  | (0.127) | (0.328) | (0.127) | (0.532) |
| ma2 |  |  |  | -0.043 |  | -0.383 | -0.452\*\*\* |  |
|  |  |  |  | (0.149) |  | (0.316) | (0.132) |  |
| ar2 |  |  |  |  | -0.206 | -0.079 |  | -0.347\* |
|  |  |  |  |  | (0.127) | (0.324) |  | (0.166) |
| BIC | 863.3 | 865.5 | 867.3 | 867.3 | 867.1 | 874.0 | 870.0 | 870.8 |
| RMSE | 285.55 | 289.33 | 285.73 | 285.77 | 285.00 | 279.54 | 280.05 | 282.70 |
| * p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001 | | | | | | | | |

综合考虑BIC和RMSE以及系数显著性，并且考虑到低阶模型中自回归阶数和移动平均阶数的影响并不是很大，为了方便起见，选择系数均显著且BIC和RMSE均小的SARIMA(0,1,1)[0,1,1]与SARIMA(1,1,2)[0,1,1]两个模型，下面对其残差序列进行纯随机性检验

map(list(model1,model7),~.$residuals) %>%  
 map(Box.test,lag=12,type="Ljung")

## [[1]]  
##   
## Box-Ljung test  
##   
## data: .x[[i]]  
## X-squared = 11.11, df = 12, p-value = 0.5195  
##   
##   
## [[2]]  
##   
## Box-Ljung test  
##   
## data: .x[[i]]  
## X-squared = 9.3403, df = 12, p-value = 0.6736

可以看到三个模型的残差均通过了纯随机性检验，可以认为它们都较好地提取了序列信息

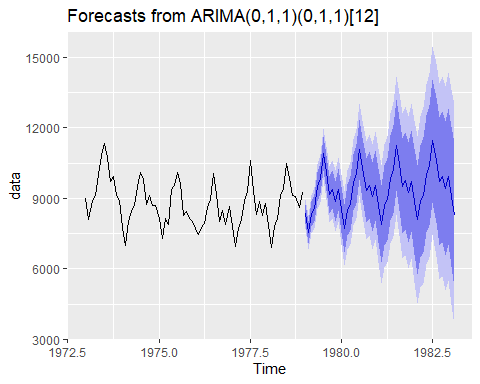
下面尝试P=1,Q=0的各低阶SARIMA模型

model9 <- Arima(data,c(0,1,1),c(1,1,0))  
model10 <- Arima(data,c(1,1,0),c(1,1,0))  
model11 <- Arima(data,c(1,1,1),c(1,1,0))  
model12 <- Arima(data,c(0,1,2),c(1,1,0))  
model13 <- Arima(data,c(2,1,0),c(1,1,0))  
model14 <- Arima(data,c(2,1,2),c(1,1,0))  
model15 <- Arima(data,c(1,1,2),c(1,1,0))  
model16 <- Arima(data,c(2,1,1),c(1,1,0))  
modelsummary(list("(0,1,1)[1,1,0]"=model9,  
 "(1,1,0)[1,1,0]"=model10,  
 "(1,1,1)[1,1,0]"=model11,  
 "(0,1,2)[1,1,0]"=model12,  
 "(2,1,0)[1,1,0]"=model13,  
 "(2,1,2)[1,1,0]"=model14,  
 "(1,1,2)[1,1,0]"=model15,  
 "(2,1,1)[1,1,0]"=model16),stars = T,gof\_map = c("bic","rmse"))

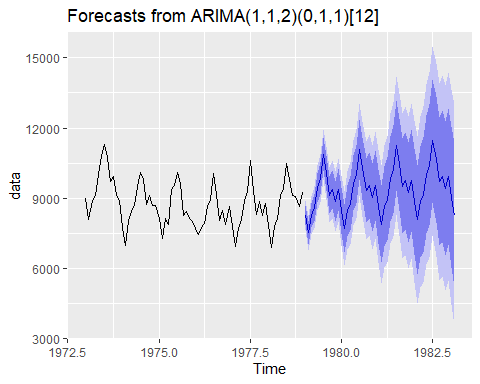
|  | (0,1,1)  [1,1,0] | (1,1,0)  [1,1,0] | (1,1,1)  [1,1,0] | (0,1,2)  [1,1,0] | (2,1,0)  [1,1,0] | (2,1,2)  [1,1,0] | (1,1,2)  [1,1,0] | (2,1,1)  [1,1,0] |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ma1 | -0.465\*\*\* |  | -0.584\* | -0.432\*\* |  | 0.482 | 0.474\*\* | 0.334 |
|  | (0.124) |  | (0.276) | (0.135) |  | (0.301) | (0.169) | (0.466) |
| sar1 | -0.346\*\* | -0.350\*\* | -0.346\*\* | -0.346\*\* | -0.355\*\* | -0.338\*\* | -0.338\*\* | -0.373\*\* |
|  | (0.127) | (0.125) | (0.127) | (0.127) | (0.126) | (0.130) | (0.130) | (0.128) |
| ar1 |  | -0.336\*\* | 0.145 |  | -0.422\*\*\* | -0.878\*\* | -0.868\*\*\* | -0.732+ |
|  |  | (0.121) | (0.320) |  | (0.125) | (0.313) | (0.112) | (0.438) |
| ma2 |  |  |  | -0.079 |  | -0.484+ | -0.492\*\*\* |  |
|  |  |  |  | (0.156) |  | (0.292) | (0.129) |  |
| ar2 |  |  |  |  | -0.246\* | -0.010 |  | -0.346\* |
|  |  |  |  |  | (0.124) | (0.310) |  | (0.163) |
| BIC | 866.5 | 869.7 | 870.4 | 870.3 | 870.0 | 876.9 | 872.8 | 873.9 |
| RMSE | 300.68 | 309.25 | 300.08 | 299.98 | 299.07 | 293.83 | 293.84 | 298.03 |
| * p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001 | | | | | | | | |

对P=1,Q=0的这一批模型，虽然相当一部分模型的系数均显著，但其BIC和RMSE普遍高于同类型的P=0,Q=1情况下的模型，从拟合的角度来说，这一批模型的效果不如之前的模型。 综上所述，SARIMA(0,1,1)[0,1,1]，SARIMA(1,1,2)[0,1,1]两个模型是较好的选择，下面分别用其作预测尝试

forecast(model1,50) %>% autoplot()



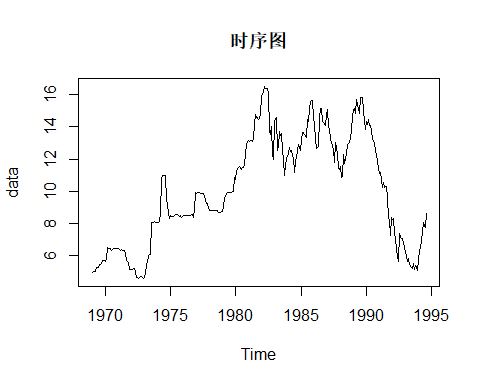
forecast(model7,50) %>% autoplot()



从预测结果来说，两个模型的差异不大，由于二者的拟合水平也没有太大差异，那么模型更简单的SARIMA(0,1,1)[0,1,1]可能是一个不错的选择

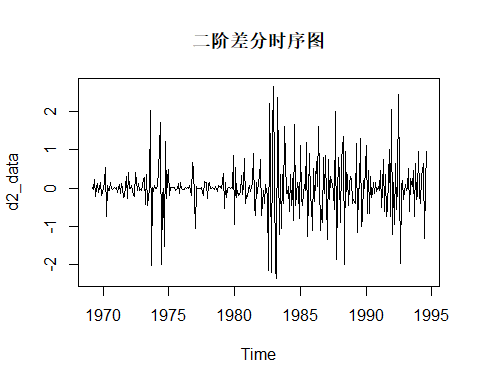
## 第二题

data <- read.table("D:/预删除文件夹/大三下/时间序列/月度利率数据.txt") %>%  
 t() %>%  
 as.vector() %>%  
 ts(start = 1969,frequency = 12)  
plot(data,main="时序图")



从时序图来看，序列呈现出先上升后下降的趋势，考虑作二阶差分

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

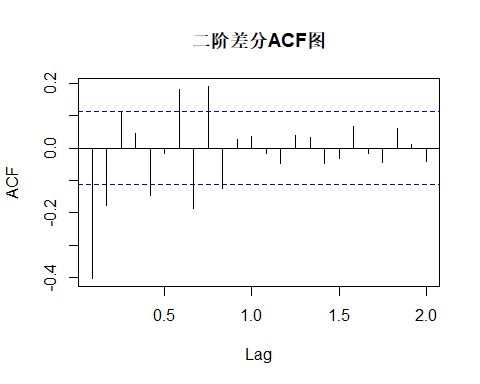


Box.test(d2\_data,lag = 12)

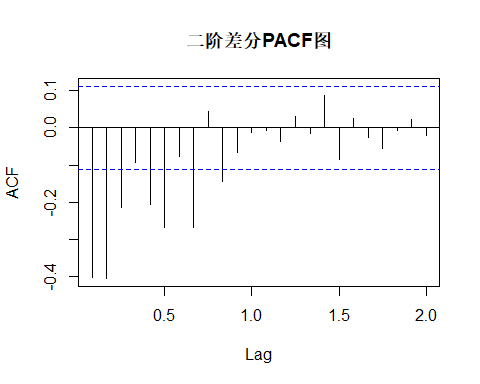
##   
## Box-Pierce test  
##   
## data: d2\_data  
## X-squared = 107.49, df = 12, p-value < 2.2e-16

二阶差分后的时序图没有明显趋势，但可以注意到从1983年开始波动幅度有明显增加，基于Q统计量进行纯随机性检验，结果显示序列不是纯随机的，下面进一步考察ACF图和PACF图

acf(d2\_data,main="二阶差分ACF图")



acf(d2\_data,type = "partial",main="二阶差分PACF图")



ACF图和PACF图没有明显的拖尾或截尾特征，可以尝试ARI(10,2)或IMA(2,10)

model1 <- Arima(data,c(10,2,0))  
model2 <- Arima(data,c(0,2,10))  
model3 <- Arima(data,c(4,2,5))  
modelsummary(list("ARIMA(10,2,0)"=model1,  
 "ARIMA(0,2,10)"=model2,  
 "ARIMA(4,2,5)"=model3),stars = T,statistic = NULL,gof\_map = c("bic","rmse"))

|  | ARIMA(10,2,0) | ARIMA(0,2,10) | ARIMA(4,2,5) |
| --- | --- | --- | --- |
| ar1 | -0.780\*\*\* |  | -0.784\*\*\* |
| ar2 | -0.849\*\*\* |  | -1.148\*\*\* |
| ar3 | -0.678\*\*\* |  | -0.780\*\*\* |
| ar4 | -0.619\*\*\* |  | -0.791\*\*\* |
| ar5 | -0.662\*\*\* |  |  |
| ar6 | -0.583\*\*\* |  |  |
| ar7 | -0.354\*\*\* |  |  |
| ar8 | -0.364\*\*\* |  |  |
| ar9 | -0.083 |  |  |
| ar10 | -0.149\*\* |  |  |
| ma1 |  | -0.840\*\*\* | -0.074\* |
| ma2 |  | -0.258\*\*\* | 0.325\*\*\* |
| ma3 |  | 0.145\* | -0.312\*\*\* |
| ma4 |  | 0.004 | 0.086\*\* |
| ma5 |  | -0.178\* | -0.994\*\*\* |
| ma6 |  | 0.101 |  |
| ma7 |  | 0.253\*\*\* |  |
| ma8 |  | -0.273\*\*\* |  |
| ma9 |  | 0.218\*\* |  |
| ma10 |  | -0.173\*\* |  |
| BIC | 518.7 | 502.3 | 503.8 |
| RMSE | 0.51 | 0.49 | 0.49 |
| * p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001 | | | |

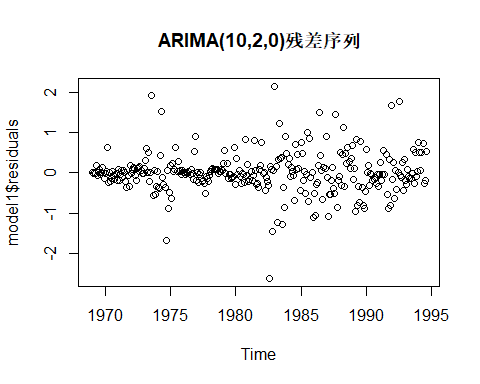
外经尝试，除了AR和MA模型外，低阶ARIMA模型或高阶ARIMA模型都出现了大量系数不显著的情况，但ARIMA(4,2,5)却出现了所有显著的情况，三个模型的回归结果如上表所示

map(list(model1,model2,model3),~.$residuals) %>%  
 map(Box.test,lag=12)

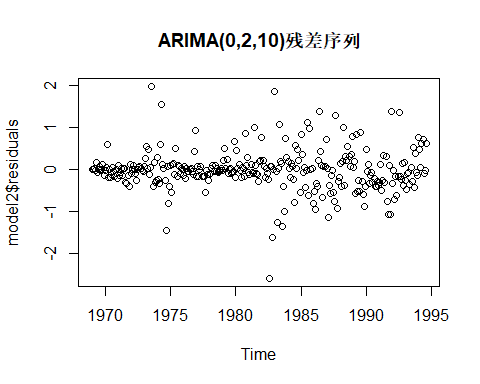
## [[1]]  
##   
## Box-Pierce test  
##   
## data: .x[[i]]  
## X-squared = 2.8146, df = 12, p-value = 0.9967  
##   
##   
## [[2]]  
##   
## Box-Pierce test  
##   
## data: .x[[i]]  
## X-squared = 1.0777, df = 12, p-value = 1  
##   
##   
## [[3]]  
##   
## Box-Pierce test  
##   
## data: .x[[i]]  
## X-squared = 11.121, df = 12, p-value = 0.5186

三个模型的残差序列都通过了纯随机性检验，下面进一步检验其是否具有异方差问题

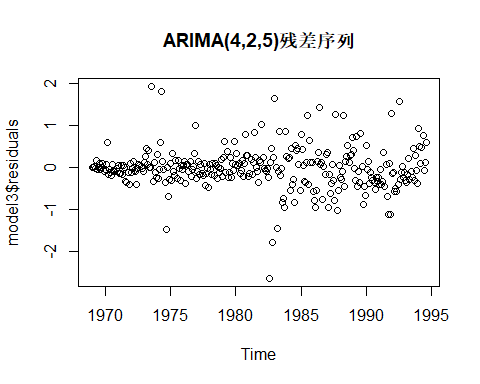
plot(model1$residuals,type = "p",main="ARIMA(10,2,0)残差序列")



plot(model2$residuals,type = "p",main="ARIMA(0,2,10)残差序列")

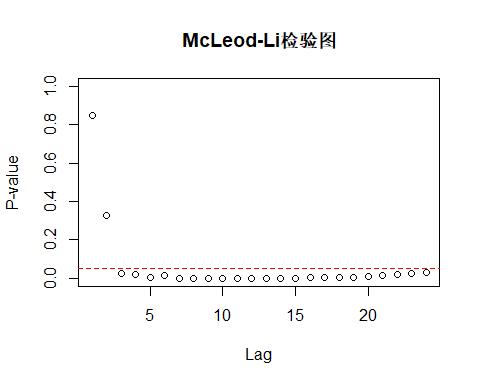


plot(model3$residuals,type = "p",main="ARIMA(4,2,5)残差序列")

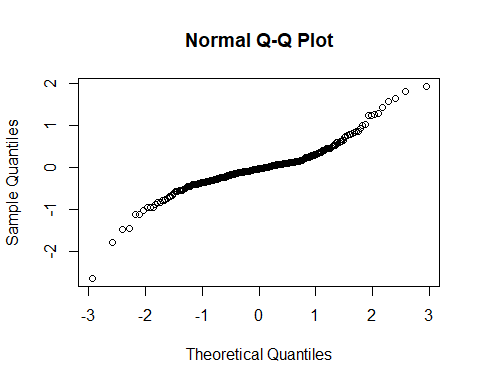


从散点图来看，三个模型的残差都表现出了一定的递增型异方差特征，这与最初基于时序图的判断是一致的，为了简便起见，下面选用ARIMA(4,2,5)进行后续分析

McLeod.Li.test(y=model3$residuals,main="McLeod-Li检验图")



qqnorm(model3$residuals)

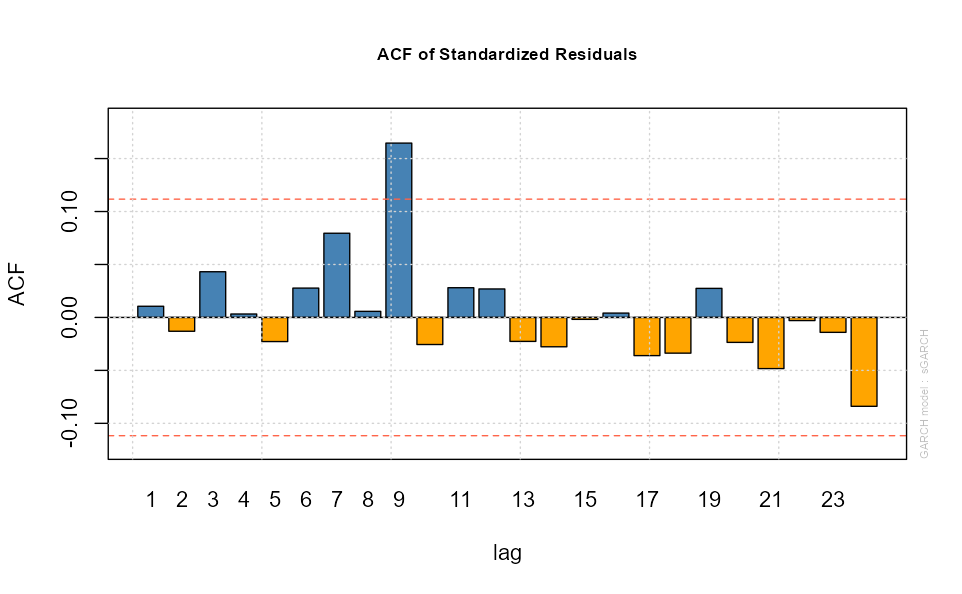


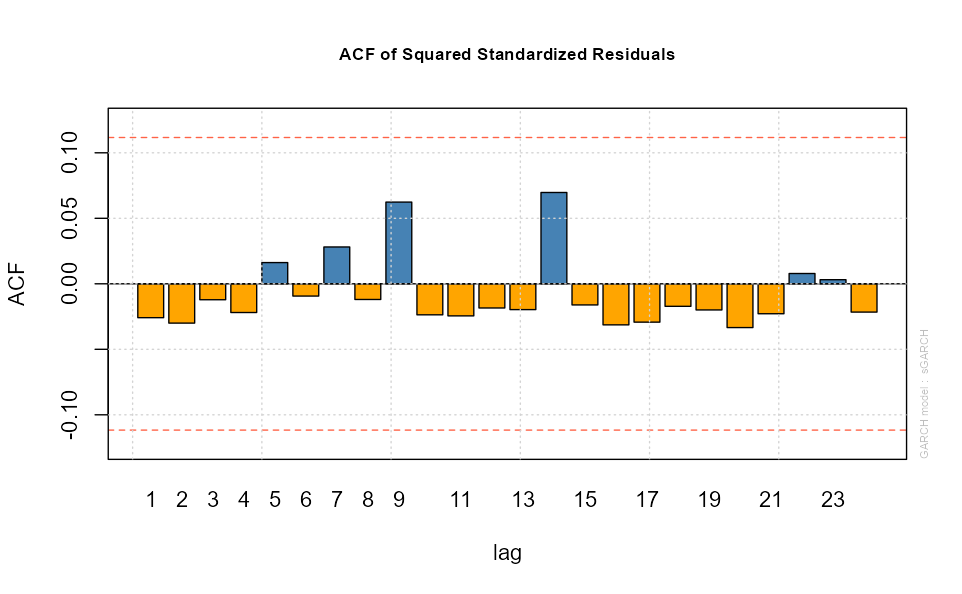
进一步的检验表明残差序列的平方自延迟三阶之后存在显著的自相关性，说明存在条件异方差，从正态QQ图可知残差序列也不服从正态分布，且为非对称厚尾分布，下面尝试拟合GARCH(1,1)+ARIMA(4,d,5)模型，其中残差分布假定为SGED

spec <- ugarchspec(variance.model = list(model = "sGARCH", garchOrder = c(1, 1)),mean.model = list(armaOrder = c(4, 5),include.mean = F,arfima = T),distribution.model = "sged")  
model <- ugarchfit(spec,data,solver = "hybrid")  
show(model)

##   
## \*---------------------------------\*  
## \* GARCH Model Fit \*  
## \*---------------------------------\*  
##   
## Conditional Variance Dynamics   
## -----------------------------------  
## GARCH Model : sGARCH(1,1)  
## Mean Model : ARFIMA(4,d,5)  
## Distribution : sged   
##   
## Optimal Parameters  
## ------------------------------------  
## Estimate Std. Error t value Pr(>|t|)  
## ar1 0.815296 0.001313 621.0492 0  
## ar2 0.166147 0.001336 124.3448 0  
## ar3 0.052090 0.001012 51.4650 0  
## ar4 -0.033480 0.000466 -71.8521 0  
## ma1 0.315183 0.004813 65.4848 0  
## ma2 0.080304 0.001967 40.8331 0  
## ma3 0.012783 0.000680 18.7908 0  
## ma4 -0.003108 0.000478 -6.5044 0  
## ma5 -0.012330 0.001248 -9.8777 0  
## arfima 0.033612 0.000737 45.6305 0  
## omega 0.006042 0.000371 16.2671 0  
## alpha1 0.234226 0.007492 31.2614 0  
## beta1 0.756704 0.013608 55.6088 0  
## skew 1.080639 0.015242 70.8980 0  
## shape 0.581707 0.017498 33.2444 0  
##   
## Robust Standard Errors:  
## Estimate Std. Error t value Pr(>|t|)  
## ar1 0.815296 0.001987 410.2552 0  
## ar2 0.166147 0.002466 67.3875 0  
## ar3 0.052090 0.002557 20.3709 0  
## ar4 -0.033480 0.001122 -29.8476 0  
## ma1 0.315183 0.003359 93.8285 0  
## ma2 0.080304 0.002340 34.3241 0  
## ma3 0.012783 0.000539 23.7349 0  
## ma4 -0.003108 0.000400 -7.7626 0  
## ma5 -0.012330 0.000959 -12.8597 0  
## arfima 0.033612 0.000758 44.3700 0  
## omega 0.006042 0.000344 17.5666 0  
## alpha1 0.234226 0.007666 30.5538 0  
## beta1 0.756704 0.012479 60.6389 0  
## skew 1.080639 0.013040 82.8736 0  
## shape 0.581707 0.016948 34.3224 0  
##   
## LogLikelihood : -127.4271   
##   
## Information Criteria  
## ------------------------------------  
##   
## Akaike 0.92485  
## Bayes 1.10651  
## Shibata 0.92040  
## Hannan-Quinn 0.99749  
##   
## Weighted Ljung-Box Test on Standardized Residuals  
## ------------------------------------  
## statistic p-value  
## Lag[1] 0.4264 0.5138  
## Lag[2\*(p+q)+(p+q)-1][26] 10.4081 1.0000  
## Lag[4\*(p+q)+(p+q)-1][44] 16.8143 0.9578  
## d.o.f=9  
## H0 : No serial correlation  
##   
## Weighted Ljung-Box Test on Standardized Squared Residuals  
## ------------------------------------  
## statistic p-value  
## Lag[1] 0.2412 0.6233  
## Lag[2\*(p+q)+(p+q)-1][5] 0.7743 0.9083  
## Lag[4\*(p+q)+(p+q)-1][9] 1.3231 0.9688  
## d.o.f=2  
##   
## Weighted ARCH LM Tests  
## ------------------------------------  
## Statistic Shape Scale P-Value  
## ARCH Lag[3] 0.1057 0.500 2.000 0.7450  
## ARCH Lag[5] 0.3459 1.440 1.667 0.9278  
## ARCH Lag[7] 0.5914 2.315 1.543 0.9694  
##   
## Nyblom stability test  
## ------------------------------------  
## Joint Statistic: 13.3284  
## Individual Statistics:   
## ar1 0.6914  
## ar2 0.7142  
## ar3 0.9017  
## ar4 0.8772  
## ma1 0.5090  
## ma2 0.2005  
## ma3 0.1798  
## ma4 1.3880  
## ma5 0.6047  
## arfima 0.2071  
## omega 0.1206  
## alpha1 0.3301  
## beta1 0.9010  
## skew 1.4064  
## shape 2.5209  
##   
## Asymptotic Critical Values (10% 5% 1%)  
## Joint Statistic: 3.26 3.54 4.07  
## Individual Statistic: 0.35 0.47 0.75  
##   
## Sign Bias Test  
## ------------------------------------  
## t-value prob sig  
## Sign Bias 0.3377 0.7358   
## Negative Sign Bias 0.6757 0.4998   
## Positive Sign Bias 0.6381 0.5239   
## Joint Effect 1.0069 0.7996   
##   
##   
## Adjusted Pearson Goodness-of-Fit Test:  
## ------------------------------------  
## group statistic p-value(g-1)  
## 1 20 21.61 0.30407  
## 2 30 41.68 0.06004  
## 3 40 55.64 0.04087  
## 4 50 56.29 0.22095  
##   
##   
## Elapsed time : 5.443396

结果如上所示，可以看到系数均显著，即使考虑稳健标准误也是显著的，对残差序列及残差平方序列的加权LB检验结果显示残差序列是纯随机的，下面的ACF图和PACF图显示残差序列几乎没有自相关性，但是由于尚不清楚ugarchfit对差分的处理，无法给出模型的形式





1. **实验结果分析说明**

已附于图下