407_cohort1_group5_hw2

Zechen Liu, Jessica Ghai,Yuetong Li,Ashutosh Ikade

2024-01-23

Question 1

Fama and French (2015) propose a five-factor model for expected stock returns. One of the factor is based on cross-sectional sorts on firm profitability. In particular, the factor portfolio is long firms with high profitability (high earnings divided by book equity; high ROE) and short firms with low profitability (low earnings divided by book equity; low ROE). This factor is called RMW-Robust Minus Weak.

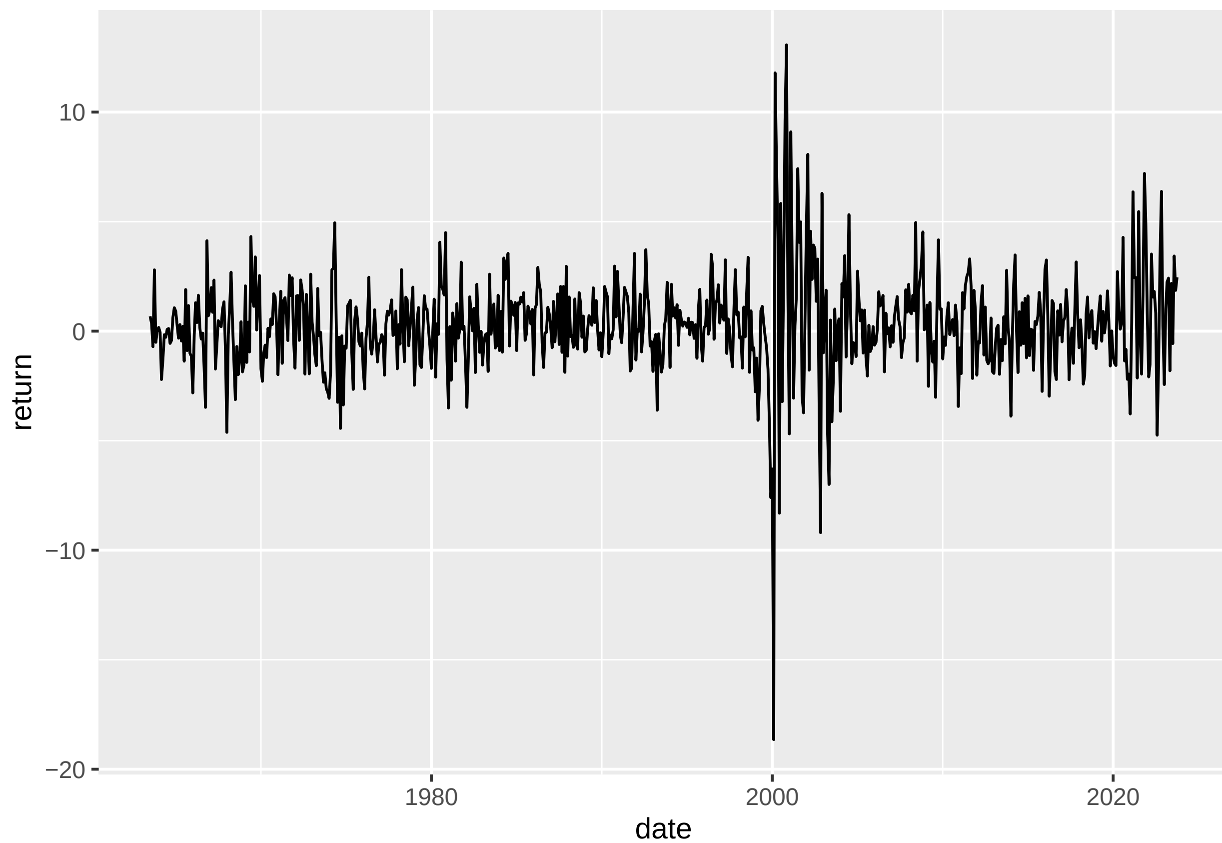
Part(1)

Denote the time series of value-weighted monthly factor returns for the RMW factor from 1963.07-2023.10 as "rmw." Plot the time.series, give the annualized mean and standard deviation of this return series.

```
data <- read.csv('F-F_Research_Data_5_Factors_2x3.csv')
data <- data[1:724 ,]
rmw <- as.numeric(data$RMW)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
df<-data.frame(date=as.Date(paste0(data$X,'01'),format="%Y%m%d"),return=rmw)

p <- ggplot(df,aes(x=date, y=return)) +
  geom_line()
print(p)
```




```r
annualized_mean = mean(rmw)*12
annualized_standard_deviation = sd(data$RMW)*sqrt(12)
cat("The annualized mean is:", annualized_mean, "\n")
```

```
## The annualized mean is: 3.523591
```

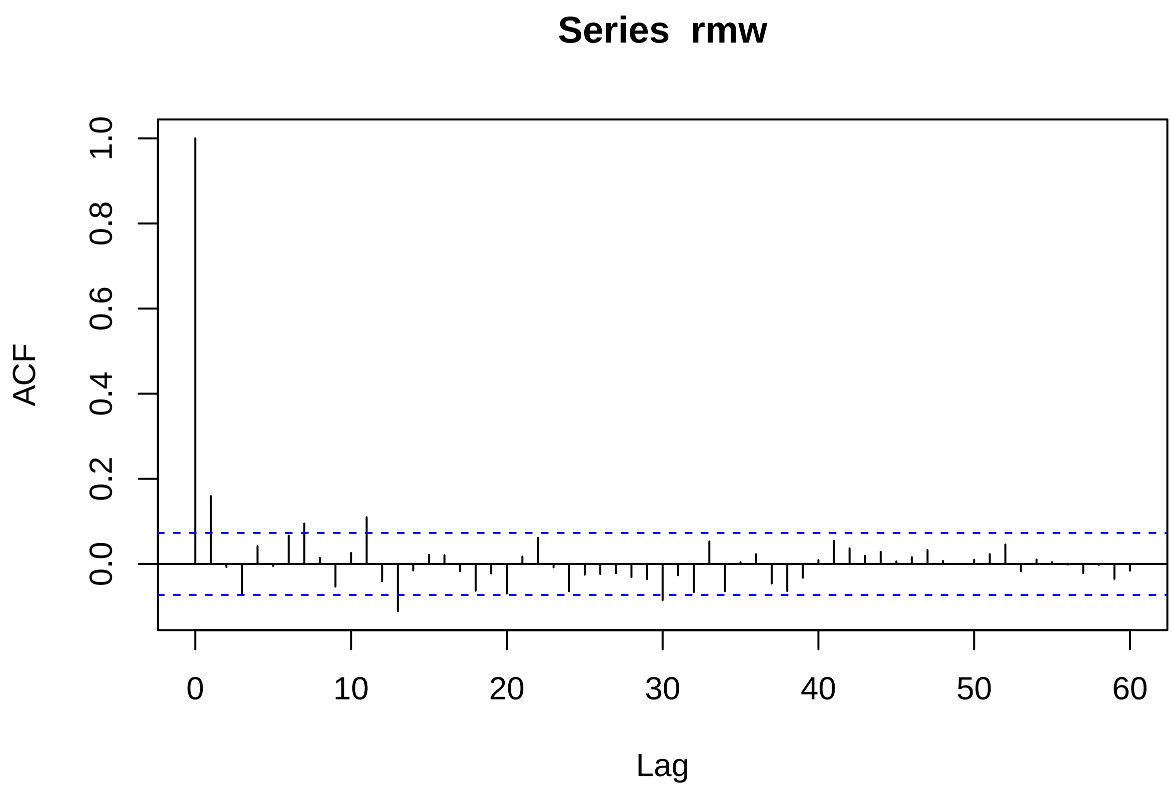
```r
cat("The annualized standard deviation is:", annualized_standard_deviation, "\n")
```

```
## The annualized standard deviation is: 7.688638
```

Part(2)

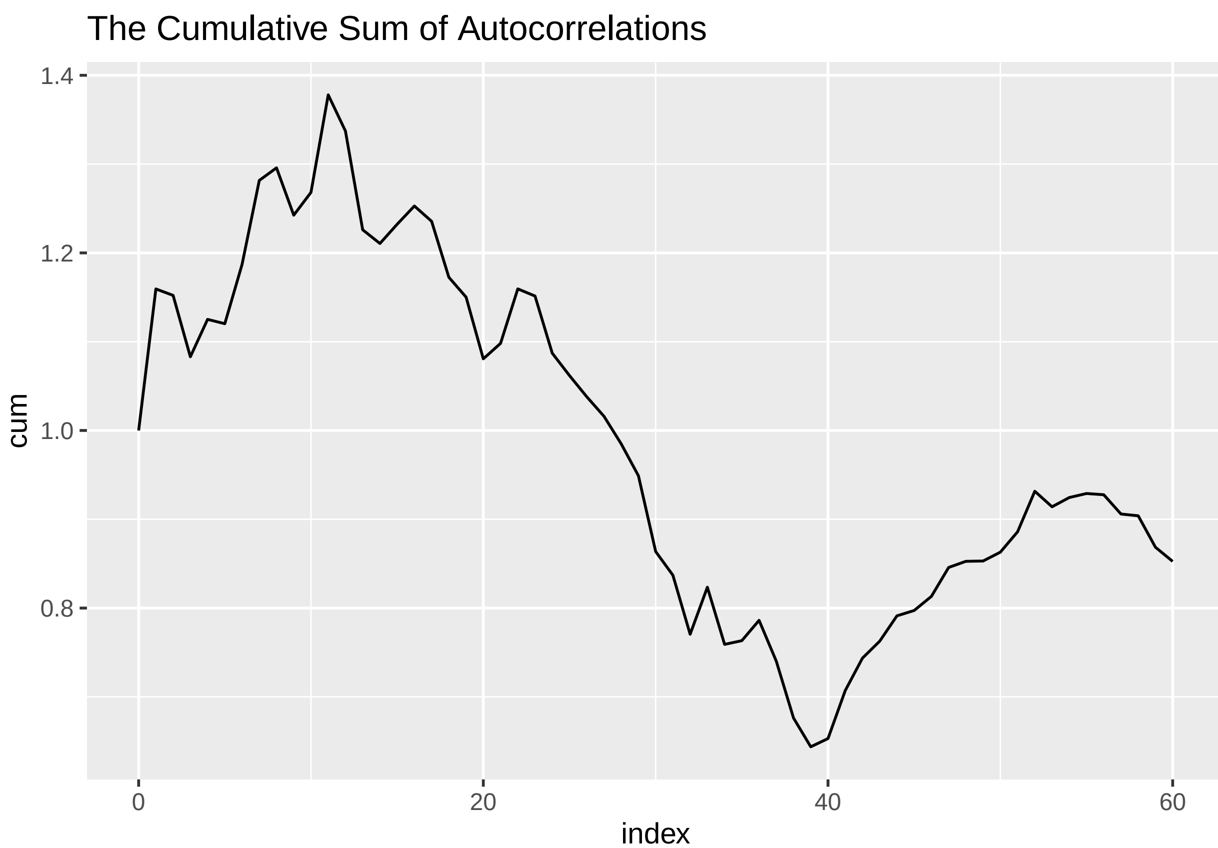
Plot the 1st to 60th order autocorrelations of rmw. Also plot the cumulative sum of these autocorrelations (that is, the 5th observation is the sum of the first 5 autocorrelations, the 11th observation is the sum of the first 11 autocorrelations, etc.). Describe these plots. In particular, do the plots hint at predictabilty of the factor returns? What are the salient patterns, if any?

```r
auto_cor <- acf(rmw,lag.max = 60,plot=TRUE)$acf
```



I think that the plot hints at predictability of the factor returns. A possible pattern that I see is that this data seems cyclical. After positive autocorrelation, there is a trend of more negative autocorrelation, which implies a cyclical pattern. We have that from the 1-60 autocorrelation plot, the first one lag is significant, showing short term momentous, but the rest lag did not show a meaningful pattern

```
df2<- data.frame(index=0:60,auto_correlation=auto_cor,cum=cumsum(auto_cor))

p3 <- ggplot(df2,aes(x=index, y=cum)) +
  geom_line()+ggtitle('The Cumulative Sum of Autocorrelations')
print(p3)
```



In this plot, I see that the as the lag increases, the cumulative sum of autocorrelations laregely decreases up until lag = 40. After lag= 40, the cumulative sum of autocorrelations begin to increase, but are less than 0. When looking at the culmulative sum graph, the lag between 30-48 cumulative sum is below 0, which shows a reversal effect

Part(3)

Perform a Ljung-Box test that the first 6 autocorrelations jointly are zero (see Lecture 3). Write out the form of the test and report the p-value. What do you conclude from this test?

```
## Warning: package 'tidyverse' was built under R version 4.2.3

## Warning: package 'tibble' was built under R version 4.2.3

## Warning: package 'tidyr' was built under R version 4.2.3

## Warning: package 'readr' was built under R version 4.2.3

## Warning: package 'purrr' was built under R version 4.2.3

## Warning: package 'dplyr' was built under R version 4.2.3

## Warning: package 'stringr' was built under R version 4.2.3

## Warning: package 'forcats' was built under R version 4.2.3

## Warning: package 'lubridate' was built under R version 4.2.3

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v lubridate 1.9.3     v tibble    3.2.1
```

```
## v purrr     1.0.2     v tidyr     1.3.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
ljung_box = Box.test(auto_cor[2:61], lag = 6, type = "Ljung-Box")
cat("The Ljung-Box Test Statistic is:", ljung_box$statistic, "\n")
```

```
## The Ljung-Box Test Statistic is: 10.20348
```

```
box_test<-Box.test(rmw,lag=6,type="Ljung-Box")
p_value<-box_test$p.value
cat("The p_value is:", p_value, "\n")
```

```
## The p_value is: 0.0001749616
```

What can be concluded from the test is that since the p value is less than 0.05, there is evidence that the first 6 autocorrelatoins are not jointly equal to 0. We have that we can reject the null hypothesis.

Part(4)

```
df$RMW_Lag = lag(rmw,1)
model = lm(return ~ RMW_Lag, data = df)
summary(model)
```

```
##
## Call:
## lm(formula = return ~ RMW_Lag, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.8928  -1.1264  -0.0111   1.0606  14.5200
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.24671    0.08230   2.998  0.00281 **
## RMW_Lag      0.15961    0.03679   4.339 1.64e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.194 on 721 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.02544,    Adjusted R-squared:  0.02409
## F-statistic: 18.82 on 1 and 721 DF,  p-value: 1.639e-05
```

The story here is that we need two explanatory variables, showing 2 effects one is short term momentum (from first one lag), one is long term reversal. We have that the first predictive variable is the one period lagged RMW. We use the one period lagged RMW because it is significant and it captures the short term momentum.

Part(5)

```
## Warning: package 'sandwich' was built under R version 4.2.3
```

```
ols_standard_errors = sqrt(diag(vcov(model)))
robust_standard_errors = sqrt(diag(vcovHC(model, type = "HC1")))
```

```r
cat("The regular OLS standard errors are:\n", ols_standard_errors, "\n\n")
```

```
## The regular OLS standard errors are:
##  0.08229668 0.03678863
```

```r
cat("The robust (White) standard errors are:\n", robust_standard_errors, "\n\n")
```

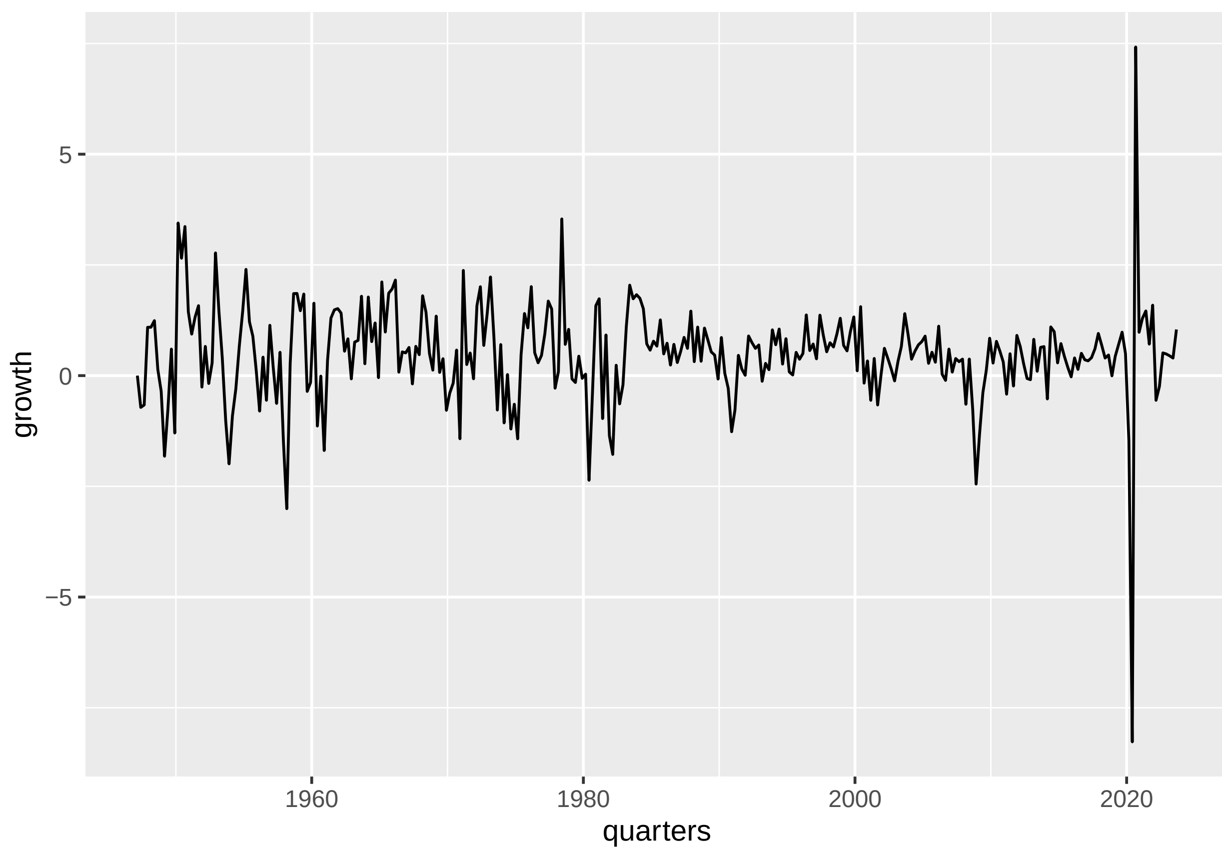
```
## The robust (White) standard errors are:
##  0.09379821 0.1026363
```

Problem 2

Part(1)

Plot the GDP growth time series and describe it.

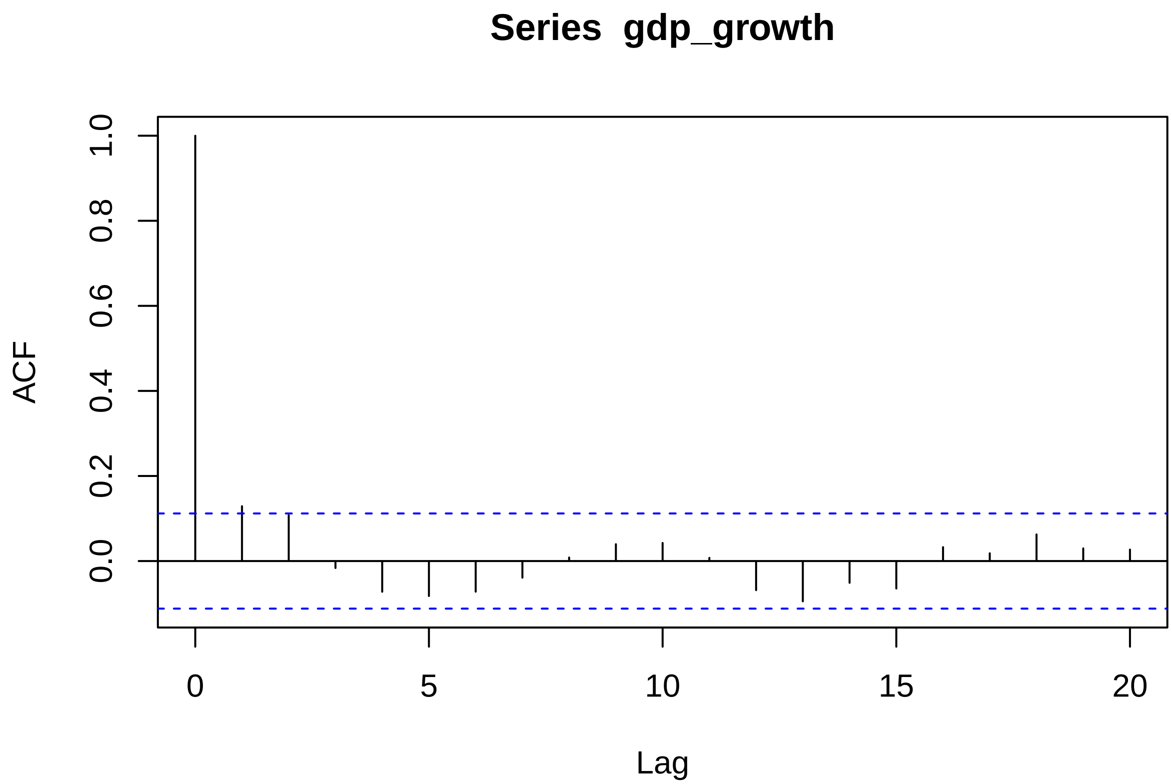
```r
data2 <- read.csv('Table.csv')[13,]
colnames(data2) <- substr(colnames(data2), 2, nchar(colnames(data2)))
gdps <- as.numeric(data2[3:309])
date <- as.Date(paste0(colnames(data2)[3:309],'01'),format="%Y%m%d")

gdps_adjusted <- gdps[-1]
gdp_growth <- c(0,log(gdps_adjusted/gdps[-length(gdps)])*100)
df3 <- data.frame(Date=date,growth=gdp_growth)
p4 <- ggplot(df3,aes(x=Date, y=growth)) +
  geom_line()+xlab("quarters")
print(p4)
```



Based on the time series plot provided, GDP growth appears to be volatile, with visible fluctuations indicating periods of economic expansion and contraction. The distribution suggests variability in growth rates, with potential outliers(covid) indicating the presence of heavy tails. The series also displays what could be interpreted as cyclical behavior, with patterns of rise and fall that might correspond to economic cycles.

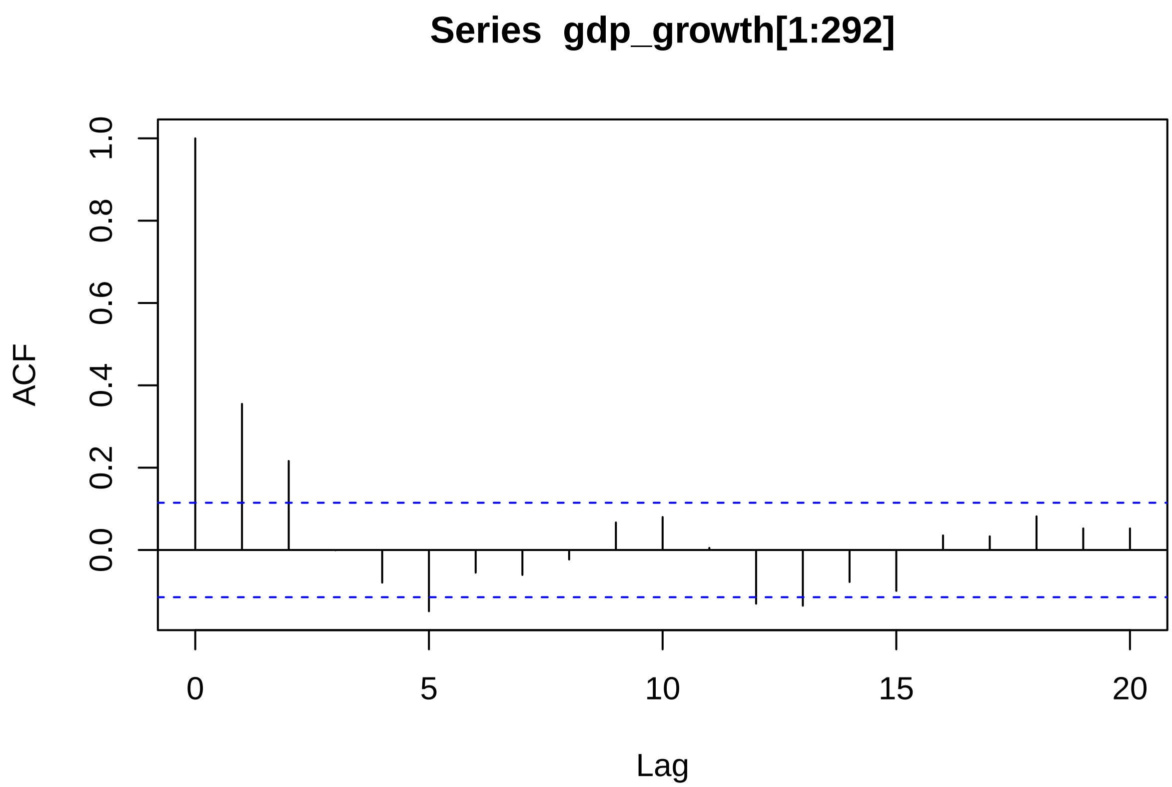
Part(2)

```
auto_cor_gdp <- acf(gdp_growth,lag.max = 20,plot=TRUE)
```



Part(3)

```
acf(gdp_growth[1:292],lag.max = 20,plot=TRUE)
```



The autocorrelation function (ACF) of GDP growth with all years included shows a significant initial lag but quickly drops, indicating short-term predictability in GDP growth. When years affected by COVID are removed, the ACF suggests slightly more predictability with a more gradual decline in autocorrelation values, indicating that the COVID years introduced additional volatility to GDP growth. This implies that the pandemic disrupted the usual GDP growth patterns, making the economic environment less predictable during that time.