Advanced Econometrics: Time Series
Assignment 1

Wilson Tai
21204700

27 March 2023

1 Section 1

1.1 Question a

In this part, we are required to plot the data:

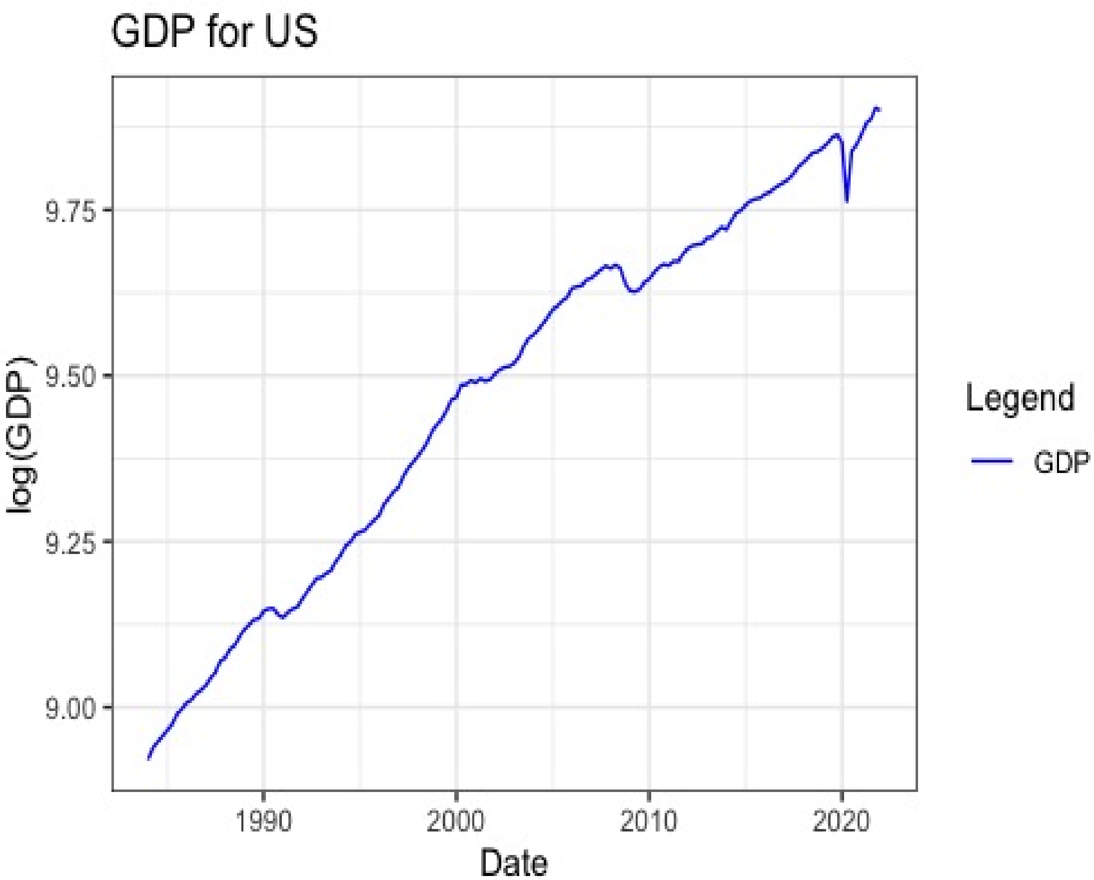


Figure 1: US GDP

We are also asked about its series: stationary, trend stationary, or stochastic. Loosely speaking, a synonym for stationary is "settle". For trend stationary, this must imply there is a trend settlement, in which applies to the graph above. Rather than deterministic (a constant linear trend), it seems quite stochastic: there is an upward trend but has some deviations possibly due to some unexplained random variable. However, through graphics is not a robust conclusion: I test this next.

1.2 Question b

The formal test for non-stationarity is called a Dickey Fuller test, whereby the null states that a unit root (or non-stationarity) is present, whilst the alternative says it is not. Test this, I find:

Augmented Dickey Fuller

Test Statistic	Critical Values	Levels
-1.23	-3.13	10%
	-3.43	5%
	-3.99	1%

ADF is peculiar: its critical values always change. Thus, if our test statistic is above our critical values, we fail to reject null of presence of unit root; else we reject presence of unit root. From the above, as it is above our critical values, we fail to reject presence of unit root. Note in my R command, I introduced "trend" because I think there is a trend.

We don't like non-stationary data because it implies our series is not independent and identically distributed: a bias is present.
I reimplement the whole process thus far but with a first difference, or growth, which is in logarithmic form.

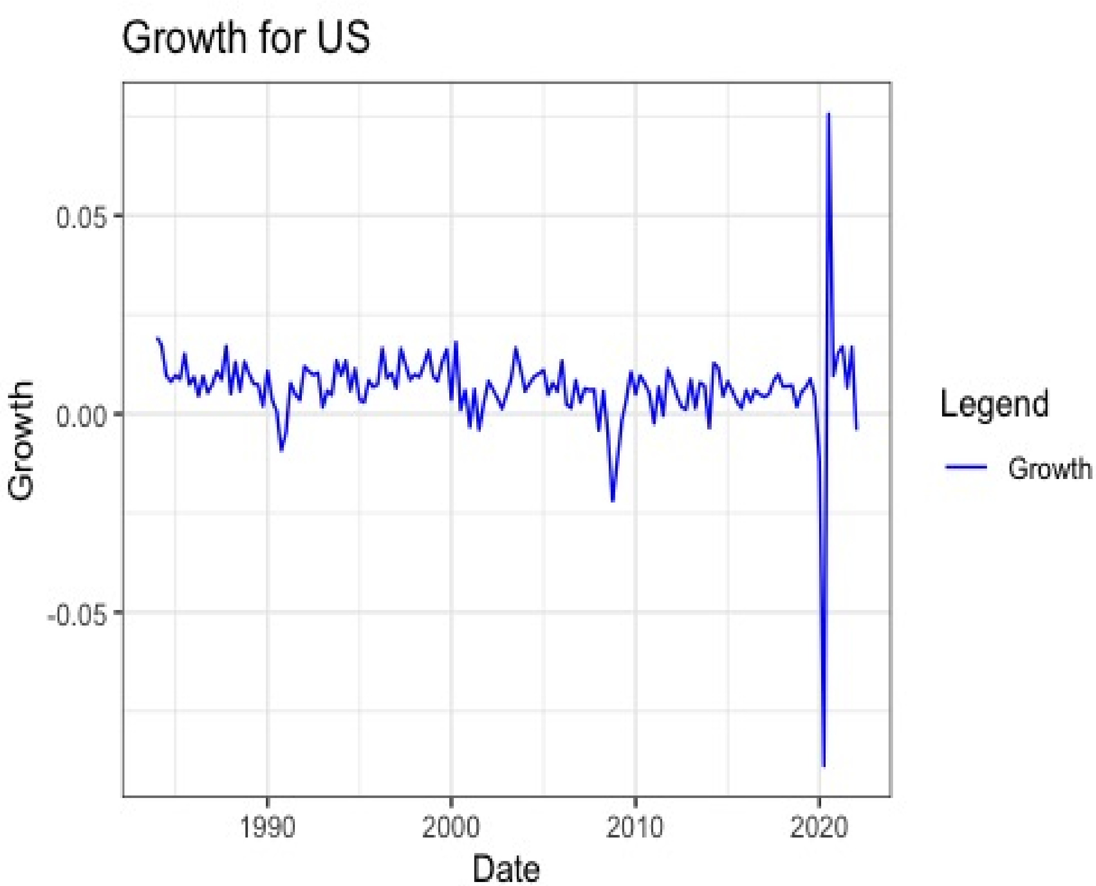


Figure 2: US Growth (in logarithmic form)

Augmented Dickey Fuller

Test Statistic	Critical Values	Levels
-6.12	-1.62	10%
	-1.95	5%
	-2.58	1%

From the above, we can reject null that a unit root is present: our test statistic is below the critical values.

1.3 Question c

Here my autocorrelation function is:

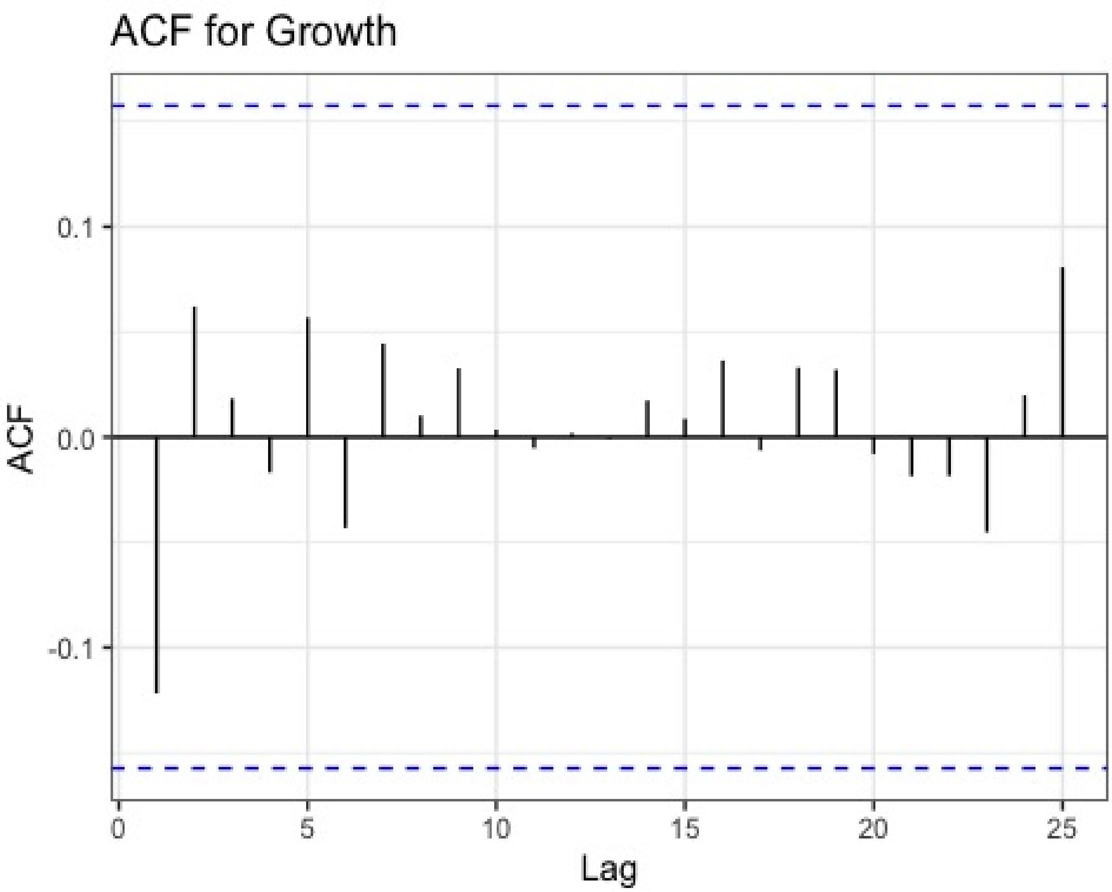


Figure 3: ACF Growth

We see that there is an oscillation between positive and negative values rather than a gradual fall: this must imply we use a moving average. The problem is that these spikes are graphically statistically insignificant. However, ACF is just a benchmark. I estimate an MA(1) and MA(2) to test this out.

But the problem with MA(q) models is that it can only predict q periods ahead: for t > q implies forecasts reverts to the mean instantaneously. To mitigate this, I also model an AR(p), as the forecasts for this model can depend upon forecasts.

1.4 Question d

The proposed models are as follows:

	Dependent variable: Growth			
	AR(1)	AR(2)	MA(1)	MA(2)
y_{t-1}	0.158**	0.117		
	(0.080)	(0.078)		
y_{t-2}		0.278***		
		(0.078)		
ε_{t-1}			0.106	0.068
			(0.067)	(0.082)
ε_{t-2}				0.214***
				(0.072)
Observations	154	154	154	154
Log Likelihood	451.734	457.784	451.070	455.214
σ^2	0.0002	0.0002	0.0002	0.0002
Akaike Inf. Crit.	−899.469	−909.569	−898.141	−904.429
Note:				*p<0.1; **p<0.05; ***p<0.01

1.5 Question e

An AR(p) and MA(q) are expressed as respectively:

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + ... + \rho_p y_{t-p} + \varepsilon_t$$

$$y_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

An AR depends on past values of itself, and MA relies on the past changes in the mean, hence the average is moving.

Before I examine my "best" model, I want to discuss the coefficients. All the summation of ρ coefficients of our models in absolute value is less than one: this implies our model is stationary and thus mean reverting. If they were greater than one, we will obtain a model with trend, whereby our model is dependent rather than independent. One tends to choose the one with the lowest Information Criteria, or here AIC. The lowest is an AR(2), but I also consider an MA(2) as it is the second lowest. I am only using an MA(2) as another model to double-check. Of course, as explained, an MA(q) can only predict q periods ahead, and will revert to the mean when t>q.

I ensure these models are stationarity: I implement this visually, and through inferences called Box Ljung test.

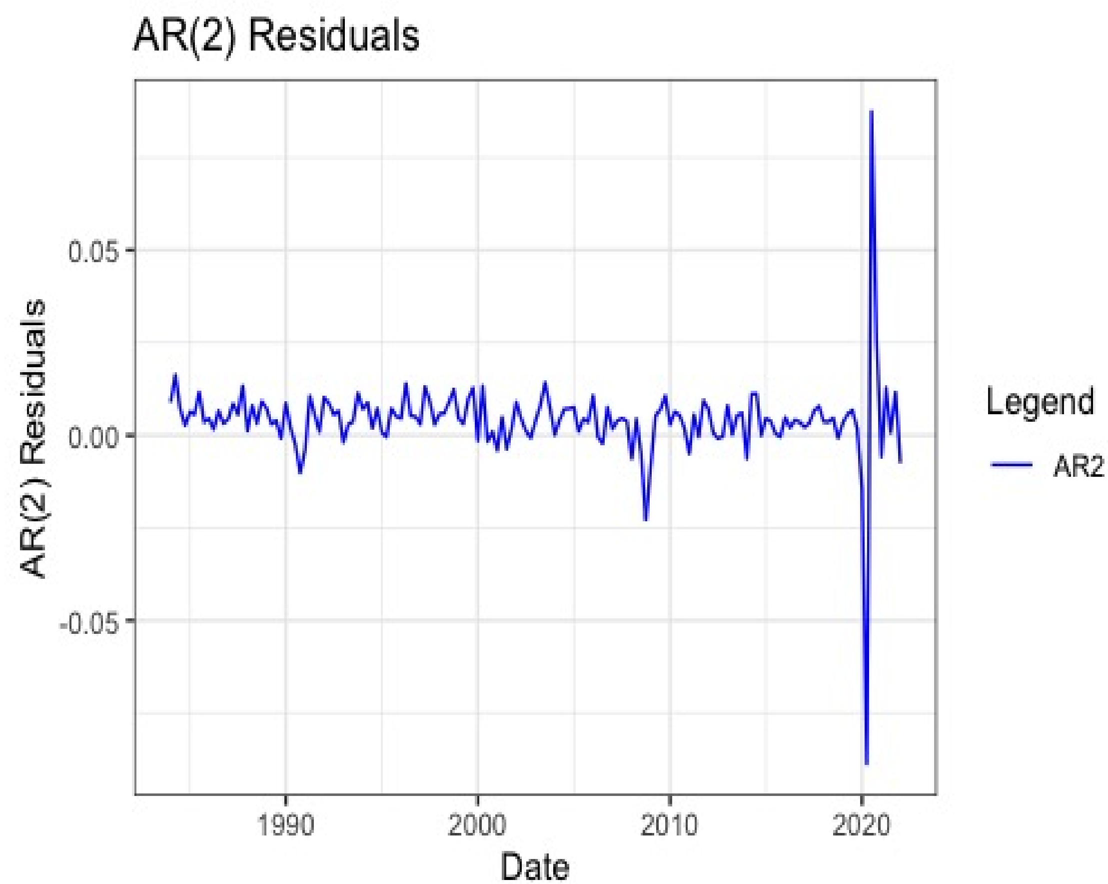


Figure 4: AR(2) residuals

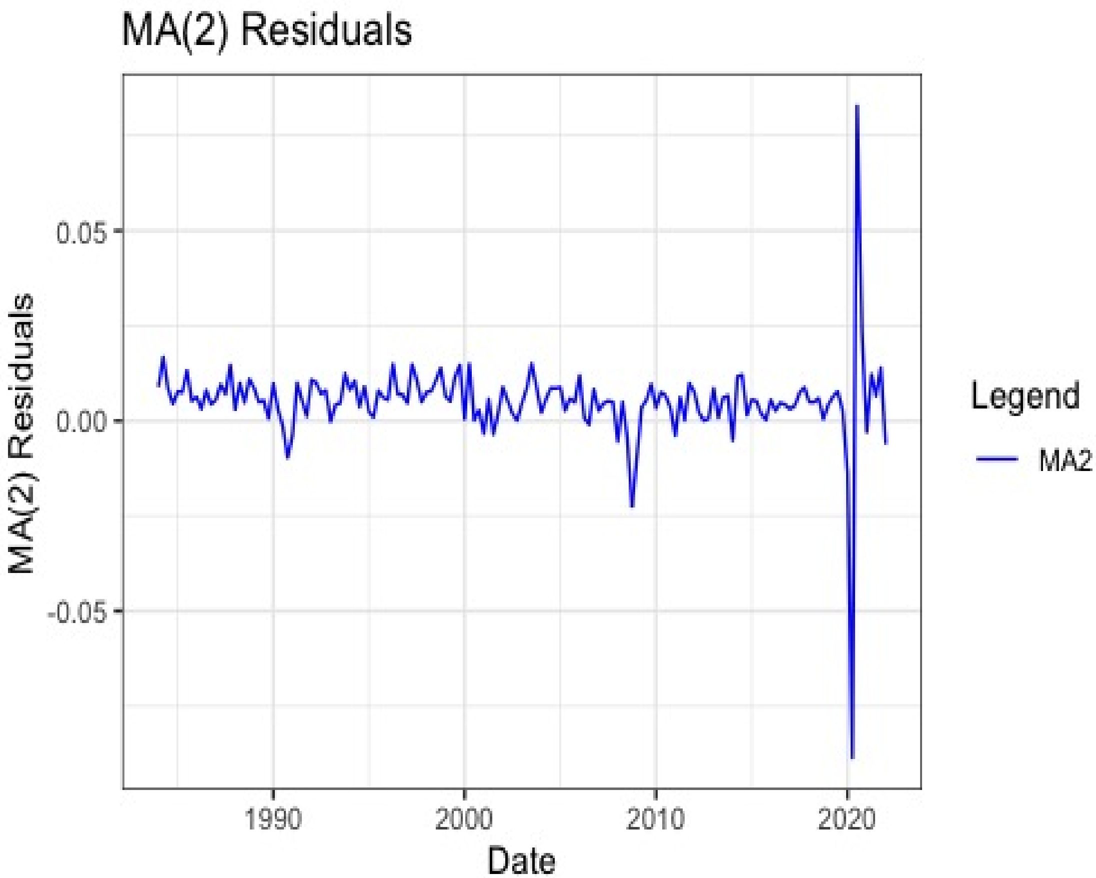


Figure 5: MA(2) residuals

In which both residuals look like white noise: covariances are equal to zero and also series is mean reverting. Note there is an extreme spike both upwards and downwards at early 2020. This is because of the Covid-19 pandemic which was a shock that the model could not explain.

The formal tests results are:

Box-Ljung Test

Model	χ^2	p-value
AR(2)	11.64	0.31
MA(2)	8.77	0.55

Where the null states that our series and therefore our model is white noise: we fail to reject the null for both our models that it is white noise.

1.6 Question f and g

Let's first have a look at how they fit:

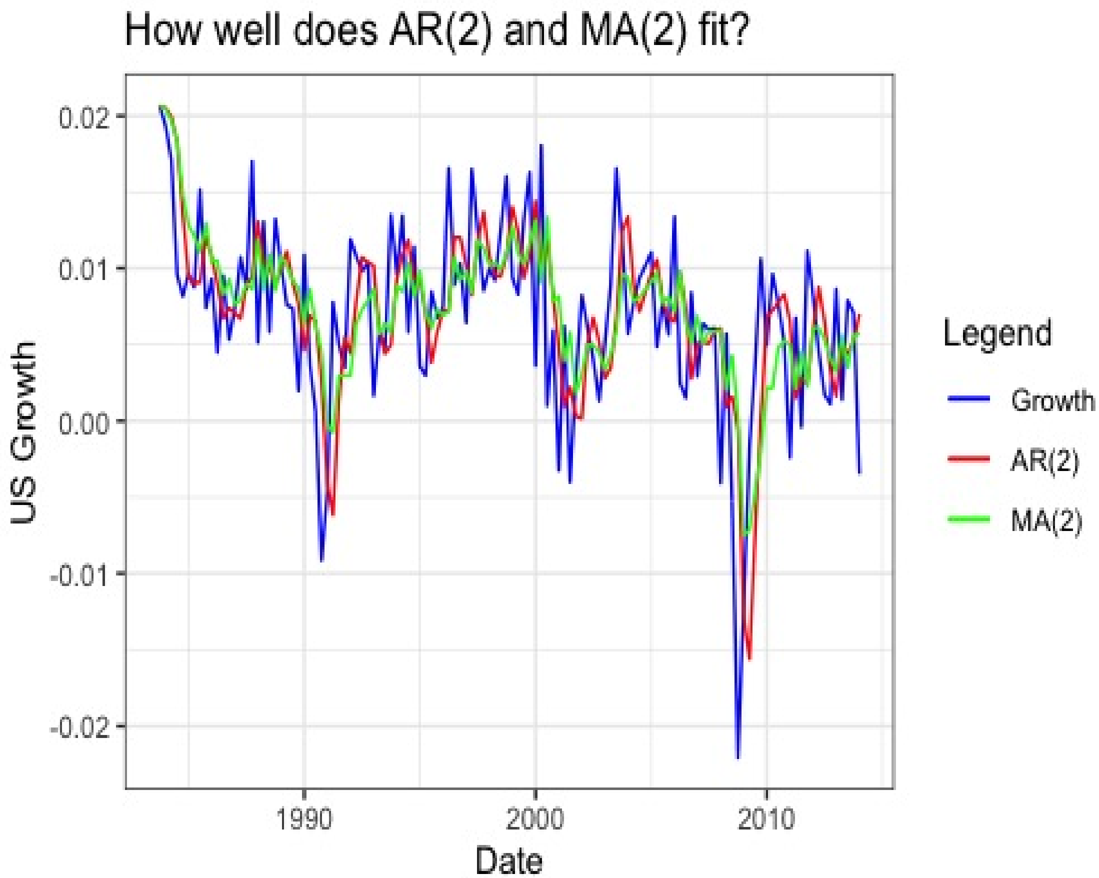


Figure 6: Fitted AR(2) and MA(2)

We can see from the above that both MA and AR captures the dynamics of the model quite well. AR may perform slightly better on capturing the more volatile components in the series i.e. 2008.

I now plot the forecasts:

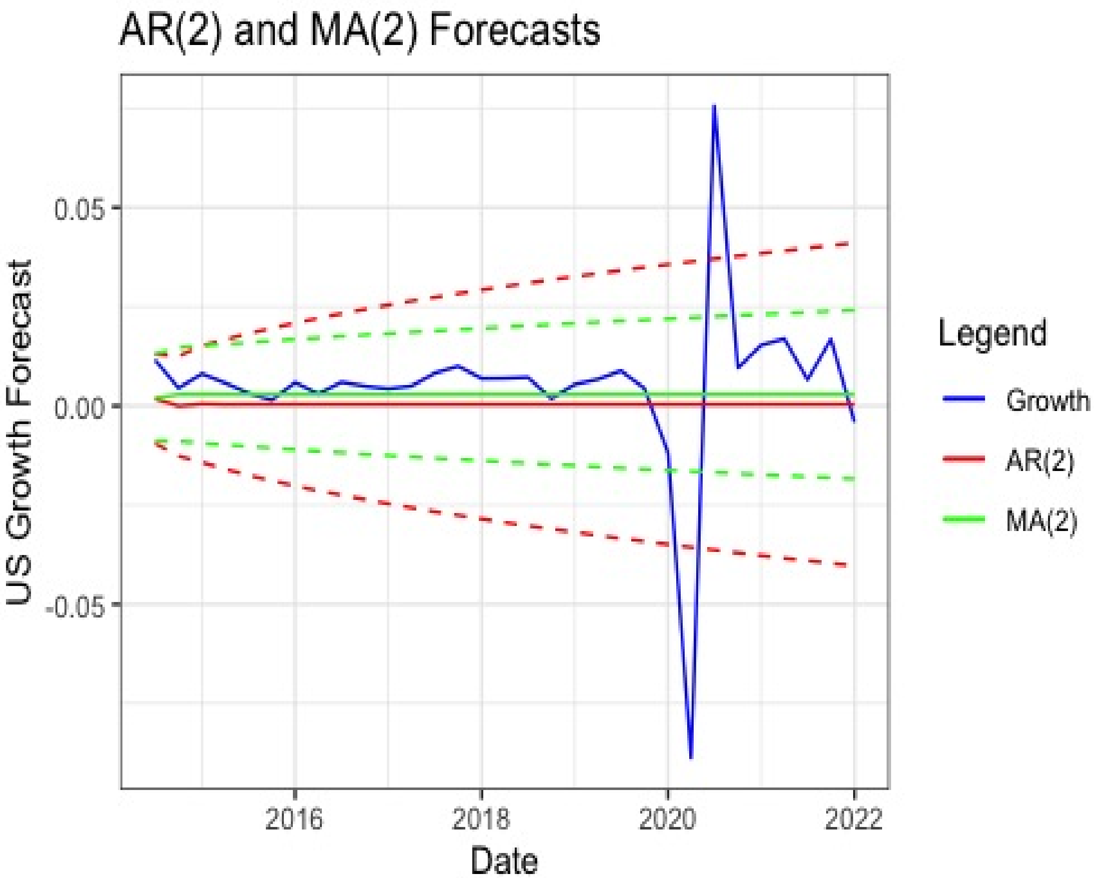


Figure 7: Forecasts AR(2) and MA(2)

On average, the forecasts are correct and justifies our model is stationary as it is mean reverting. However, if we look ever so closely, the MA(2) forecast achieves a stable forecast more quickly than the AR(2): the line gets flatter quicker. As aforementioned, this is because MA(2) can only predict q periods ahead, and will reverts to the mean instantaneously afterwards. The standard error bands, or confidence intervals also increase as time increases: this is expected because there is more uncertainty in the distant future than near future.

2 Section 2

2.1 Question a

In a 2012 paper by Messerli, he found a strong correlation between chocolate consumption and Nobel laureates as chocolate stimulates cognitive abilities. In essence: consumption of chocolate increases its value which makes people smarter and therefore higher output.

Note I said correlation as correlation does not imply correlation. If we plot this out, we see that both follow the same pattern. But of course, it also may be spurious regression: a fake relationship. We can test this all out throughout the rest of this question.

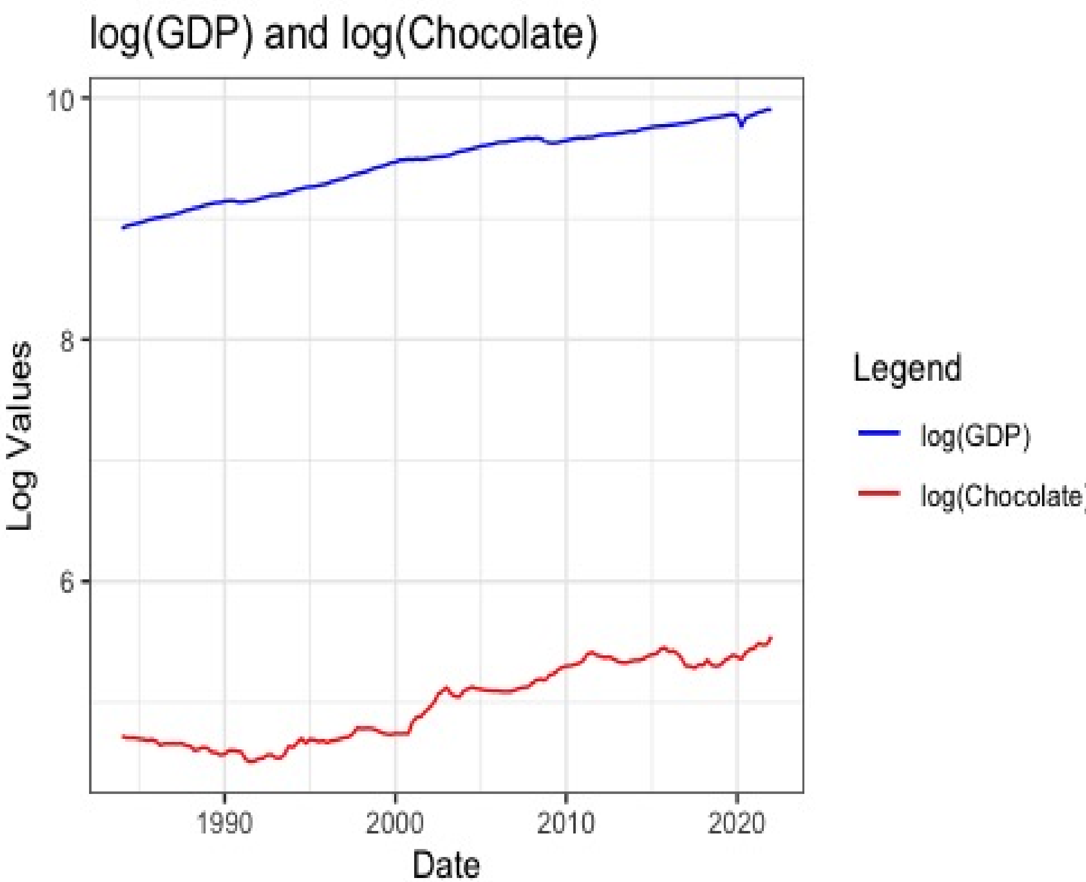


Figure 8: Chocolate and GDP

2.2 Question b

This question asks about the stationarity of my new variable Chocolate. Let's repeat same process: plot, Dicky-Fuller, and if unit root is present, first difference.

Plot of the Chocolate variable:

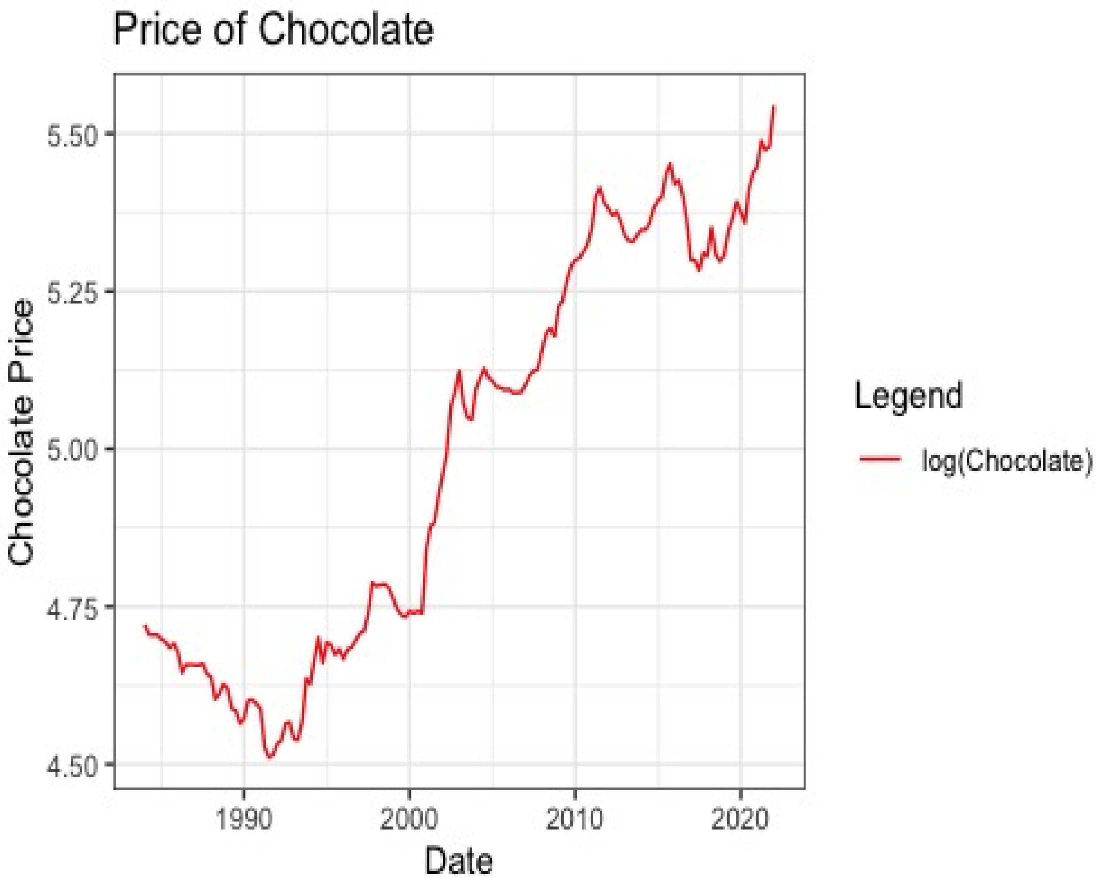


Figure 9: Chocolate Price.

We see there is a trend. Test this more appropriately with the ADF:

Augmented Dickey Fuller

Test Statistic	Critical Values	Levels
-2.18	-3.13	10%
	-3.43	5%
	-3.99	1%

As my test statistic is greater than my critical values, I fail to reject the null that a unit root is present. I therefore must first difference.

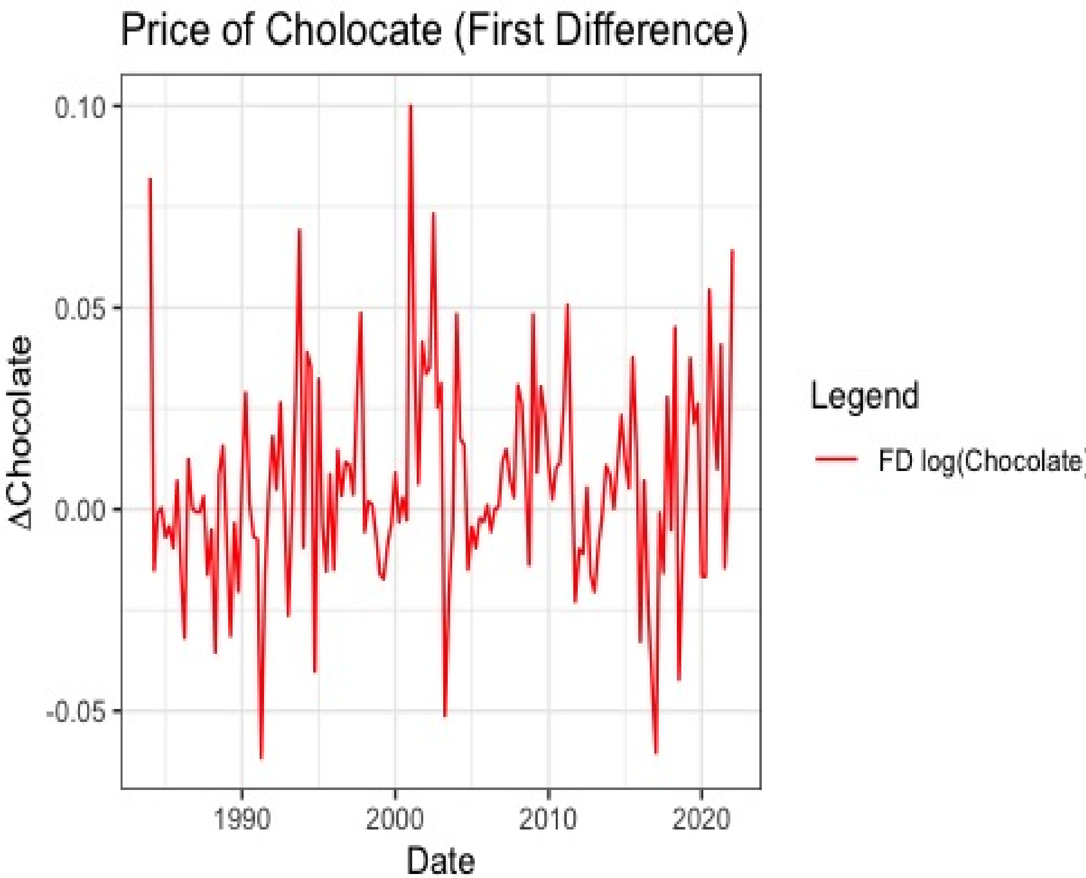


Figure 9: Δ Chocolate Price.

Augmented Dickey Fuller

Test Statistic	Critical Values	Levels
-7.17	-1.62	10%
	-1.95	5%
	-2.58	1%

The graph above seems like our new series is stationary, and the test statistics in our ADF test is below our critical value: we thus reject null that our time series has a unit root.

2.3 Question c

How many lags is required? Results show:

Results from R

	AIC	HQ	SBIC
Lags	1	1	1

So we require a VAR(1), or one lag for each variable in each equation. We COULD say that we should add more lags to create a model with higher explanatory power. But we witness a concept in machine learning called Bias-Variance trade-off: lower number of variables yield higher bias, higher number of variables increases the variance due to overfitting.

In matrix form, this looks like:

$$\begin{pmatrix} Growth_t \\ \Delta Chocolate_t \end{pmatrix} = \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} Growth_{t-1} \\ \Delta Chocolate_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{21} \end{pmatrix}$$

2.4 Question d

Here are the coefficients for the VAR:

	Dependent Variable:	
	$Growth_t$	$\Delta Chocolate_t$
$Growth_{t-1}$	-0.135^{*}	-0.191
	(0.081)	(0.180)
$\Delta Chocolate_{t-1}$	0.009	0.225^{***}
	(0.037)	(0.081)
intercept	0.007^{***}	0.006^{**}
	(0.001)	(0.002)
Observations	153	153
R^2	0.018	0.052
Adjusted R^2	0.005	0.039
Residual Std. Error (df = 150)	0.011	0.025
F Statistic (df = 2; 150)	1.384	4.107^{**}
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$	

Unlike our AR(p) from the prior, where the sum of our ρ's have to be less than one for stationarity, our eigenvalues have to be within the unit circle: less than one in absolute value.

$$det(B - \lambda I)$$

Where B is our matrix of coefficients, I is a n by n identity matrix, and λ is what we are solving for which can be computed by the determinant. Our $\lambda's$ are 0.22 and 0.13: our VAR is stationary.

From the table above, the dependent variable's lagged coefficient is significant, but the other is not. This may mean that including this extra variable yields overfitting.

2.5 Question e

Plotting the Impulse Response Functions:

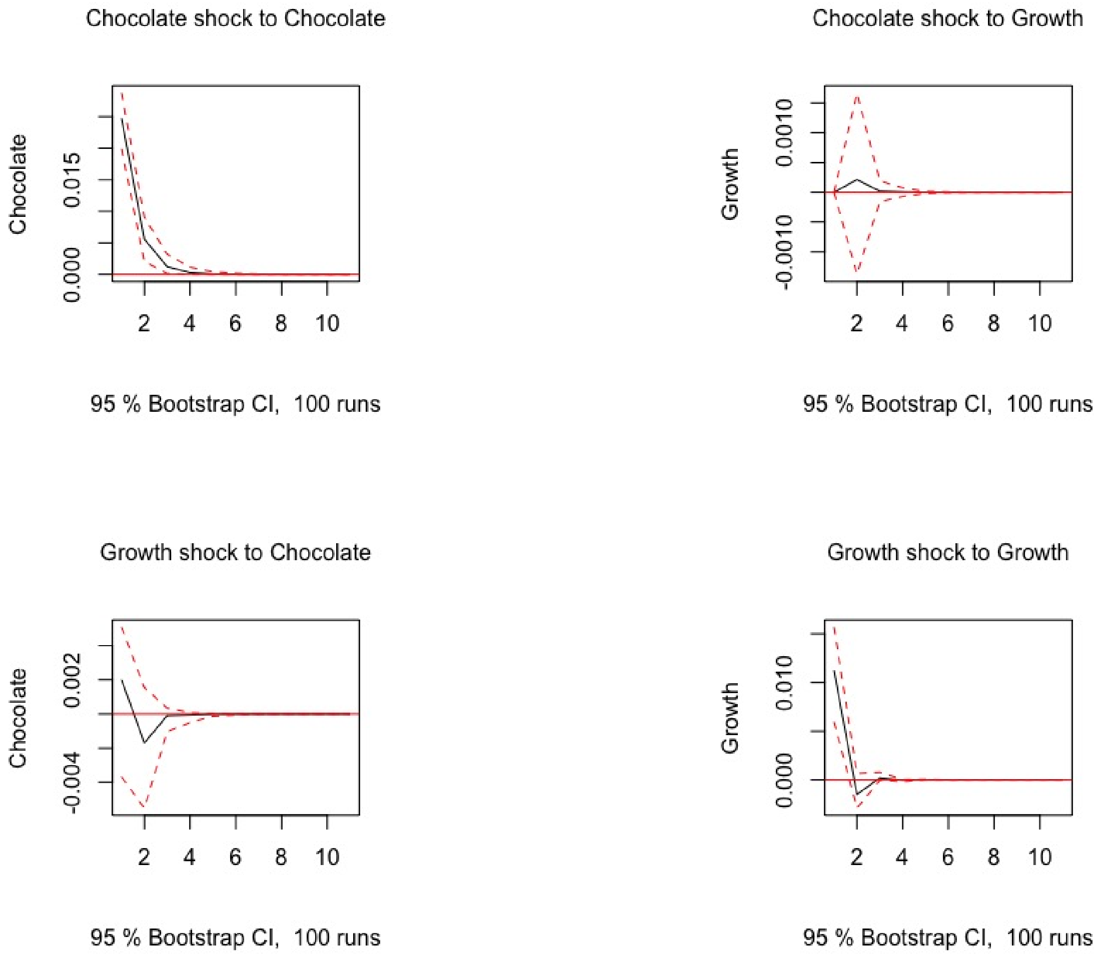


Figure 10: Impulse Response Functions

From the above, we see that each individual shock on its own variable (i.e. Growth to Growth) has a hump-shaped effect: it increases by a certain unit and falls toward stable position.

However, the shocks across variables are statistically insignificant: we know this because the confidence intervals, or red-dotted error bands contain the number zero. As a result, they have no effect on each other. This can be justified in the following subquestion.

2.6 Question f

The purpose of Granger Causation is to consider whether the inclusion of another variable help generate a better forecast. Thus, unlike the name, there is no causality. I perform the test in R: null is the new variable does not Granger Cause our Y variable; the alternative says it does.

Granger Causation

Cause	Effect	p-value
$\Delta Chocolate$	Growth	0.8164
$Growth$	$\Delta Chocolate$	0.2886

From the above, we fail to reject the null that the new variables in each multivariate equation does not Granger Cause our dependent variable: A univariate one is preferred.

2.7 Question g

I first plot the how well the VAR fits to the actual data:

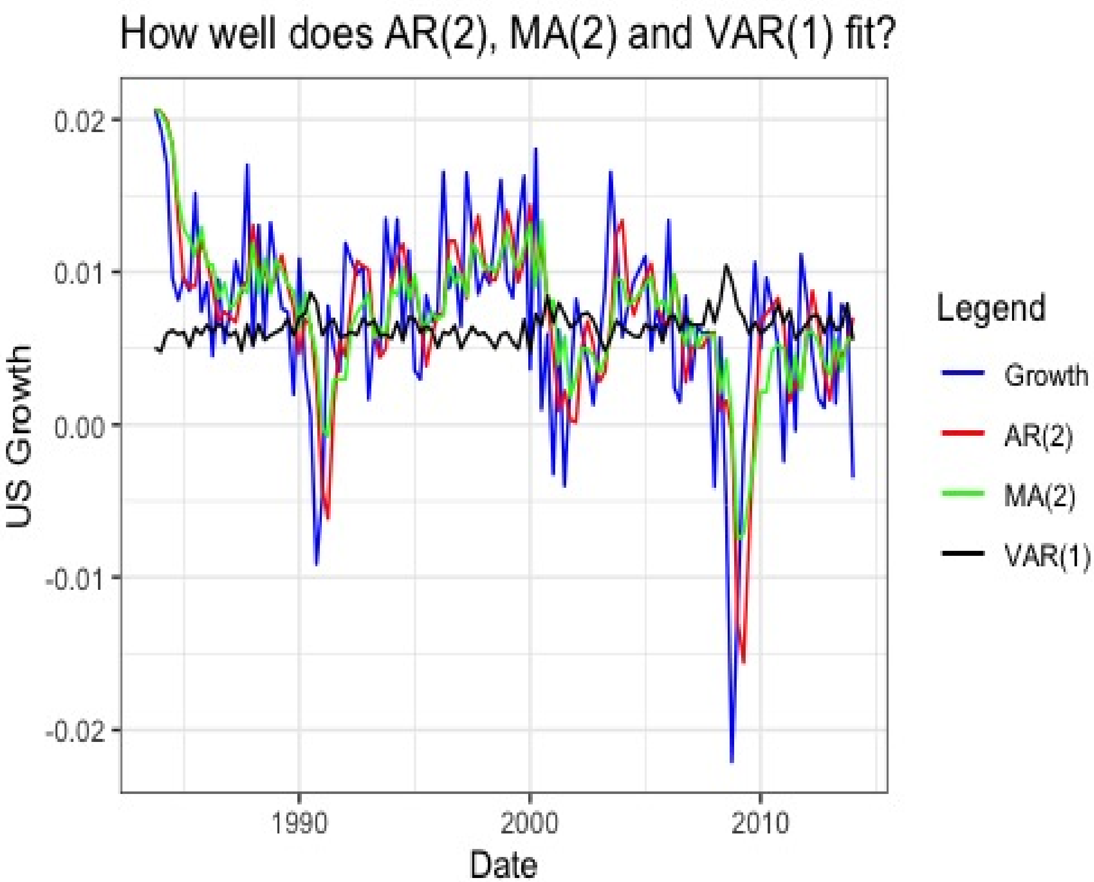


Figure 11: Fitted AR(2), MA(2) and VAR(1)

The VAR seems to have a time lag on understanding the dynamics: it increases in 2008 and drops afterwards whilst the actual series has a significant drop. Moreover, it does not capture the series as well as the AR(2) and MA(2).

The forecast for the VAR(1) is thus:

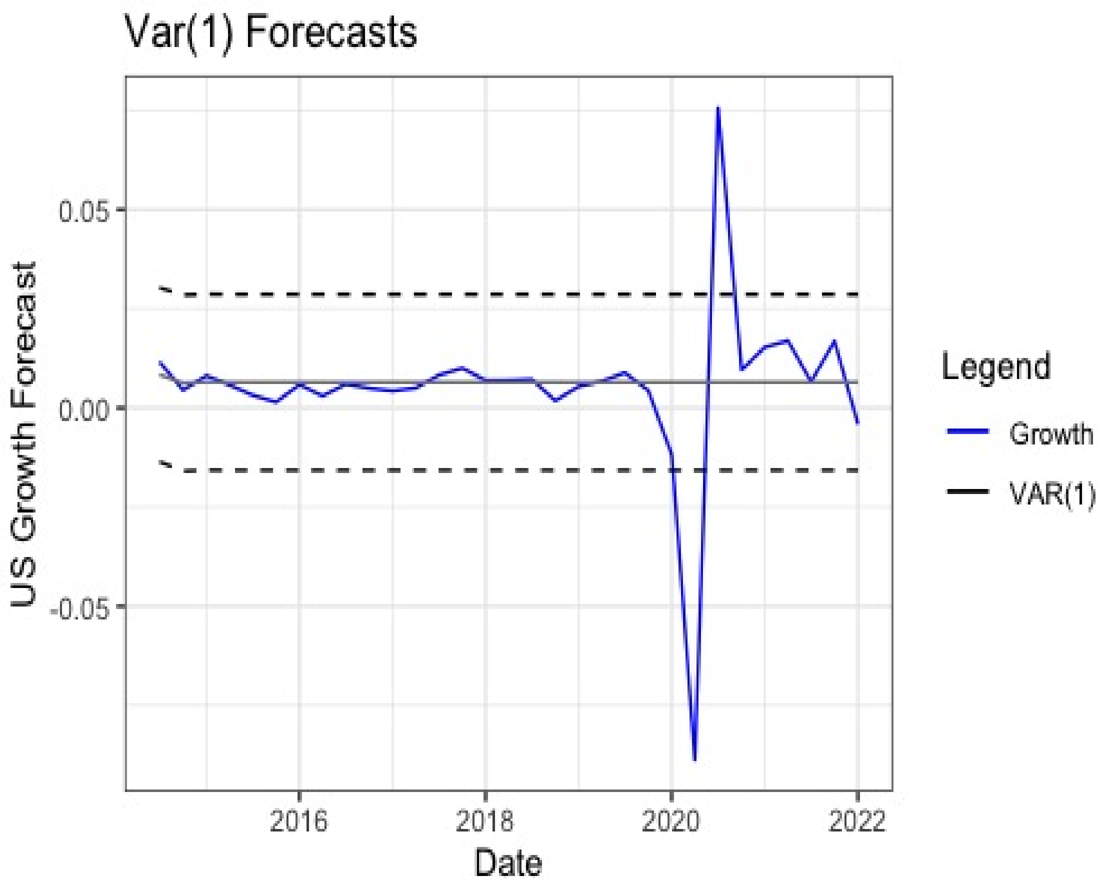


Figure 12: Forecasts VAR(1)

Like the AR(2) and MA(2), it is mean reverting, implying our model is stationary. It forecasts quite well on average.

2.8 Question h

To see whether the univariate or multivariate models do better, it is best to plot them on the one graph:

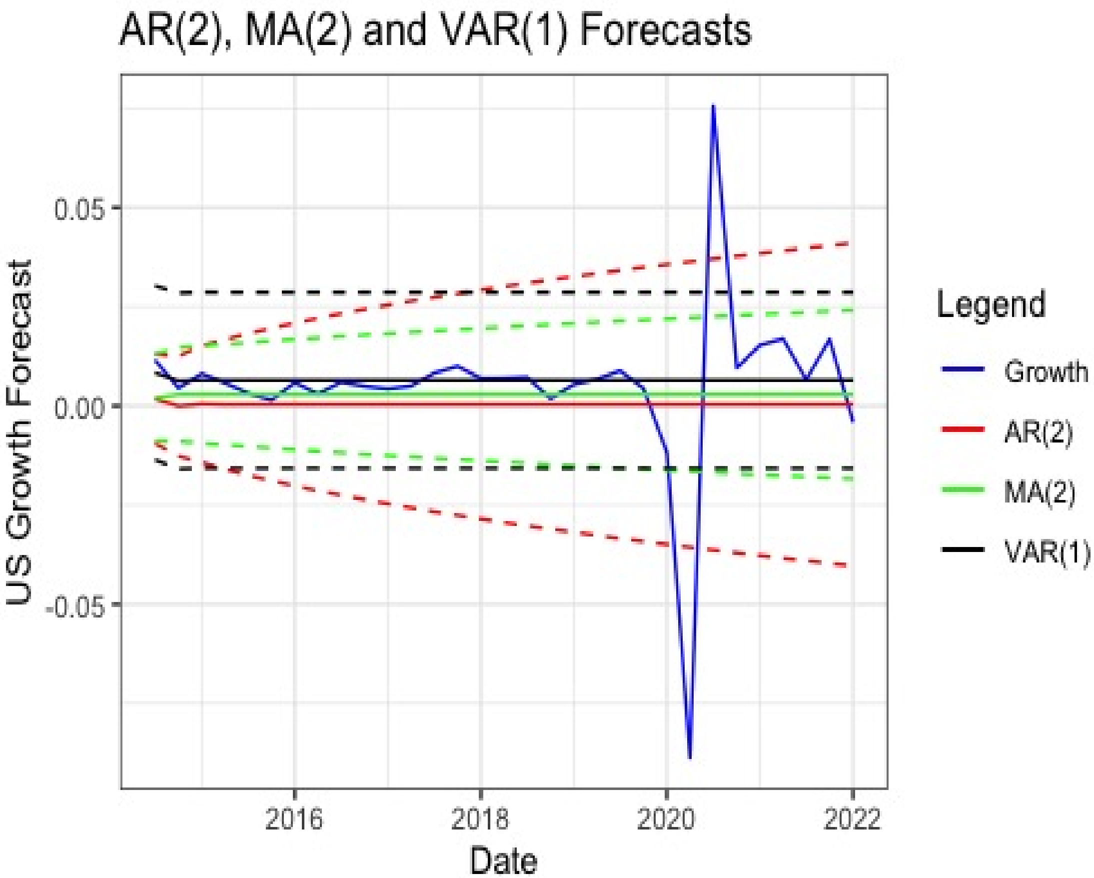


Figure 13: Forecasts AR(2), MA(2) and VAR(1)

They are all mean reverting. But note that the confidence interval for VAR(1) are larger to those of the univariate models especially at the beginning: why? We have more variables to explain and thus our variance increases: recall our machine learning bias-variance trade-off from earlier.

If I were choosing the best model, I would pick my AR(2) as it takes account of dynamics that an MA(2) does not, and also does not overfit the model as the VAR(1) here does.

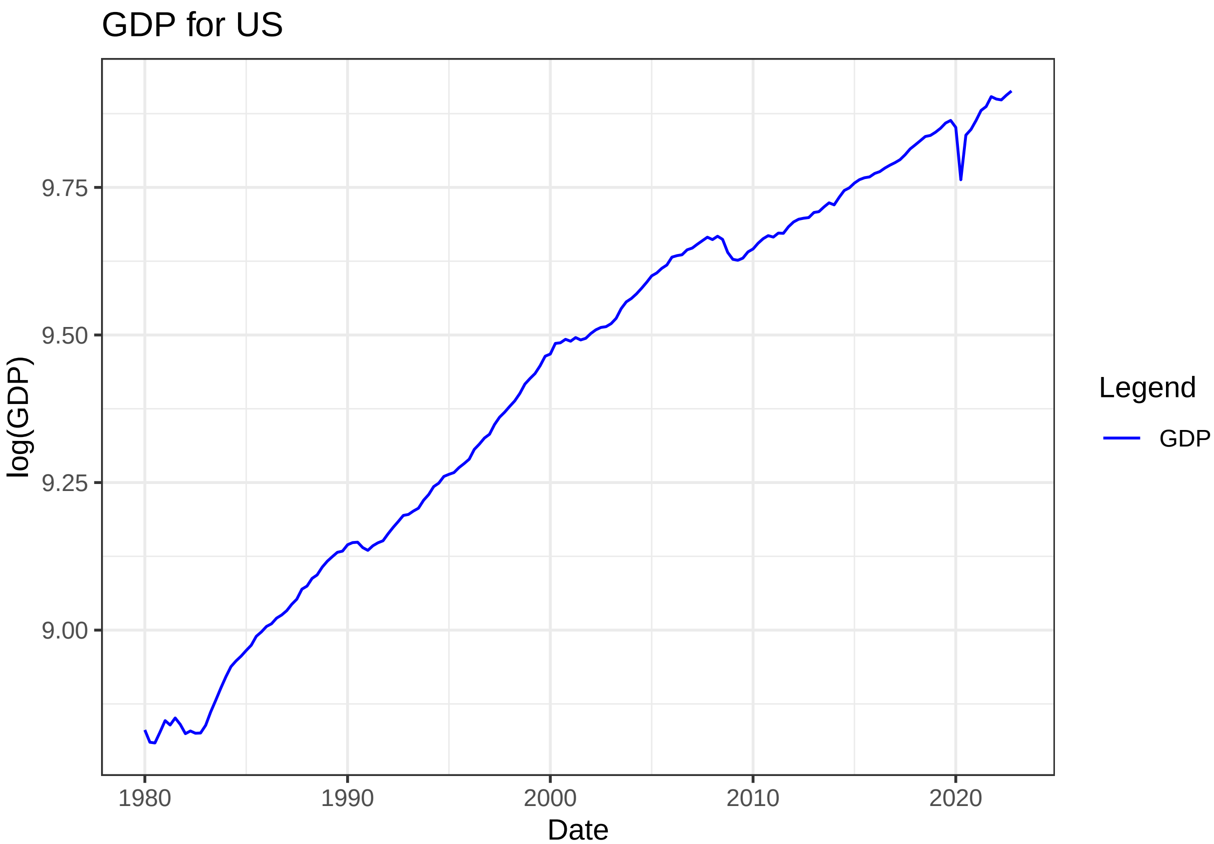
3 References

Messerli F. H (2012) 'Chocolate consumption, cognitive function, and Nobel laureates.' *The New England Journal of Medicine*. Vol 367. Pp. 1562 - 1564.

Time Series

Wilson Tai

```r
##loading packages
library(haven)
library(lubridate)
library(tidyverse)
library(tsibble)
library(urca)
library(ggplot2)
library(forecast)
library(stargazer)
library(vars)
library(cowplot)
library(gridExtra)

##inserting data and renaming
USdata <- read_dta("/Users/wilsontai/Downloads/assignment1.dta")
USdata <- USdata[, c(1, 2, 8)]
colnames(USdata)[2] = "GDP"
colnames(USdata)[3] = "Chocolate"
attach(USdata)

##PART 1
##Question a
USdata  <- USdata %>%
  mutate(date=ymd(datestr), logGDP = log(GDP)) %>%  #assign time component and log the GDP
  as_tsibble(index=date)

cols <- c("GDP" = "blue")

ggplot(data=USdata, aes(x=date)) +
  theme_bw() +
  geom_line(aes(y=logGDP, color = "GDP")) +
  scale_colour_manual(name="Legend",values=cols) +
  labs(x = "Date", y = "log(GDP)", title = "GDP for US")
```




```r
##Question b
summary(ur.df(USdata$logGDP, type = "trend"))
```

```
##
## ###############################################
## # 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
## -0.095563 -0.002769  0.000843  0.003789  0.061310
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.1972481  0.1516221   1.301    0.195
## z.lag.1     -0.0212094  0.0171266  -1.238    0.217
## tt           0.0001104  0.0001137   0.971    0.333
## z.diff.lag  -0.0703893  0.0761612  -0.924    0.357
##
## Residual standard error: 0.01122 on 166 degrees of freedom
## Multiple R-squared:  0.02977,    Adjusted R-squared:  0.01223
```

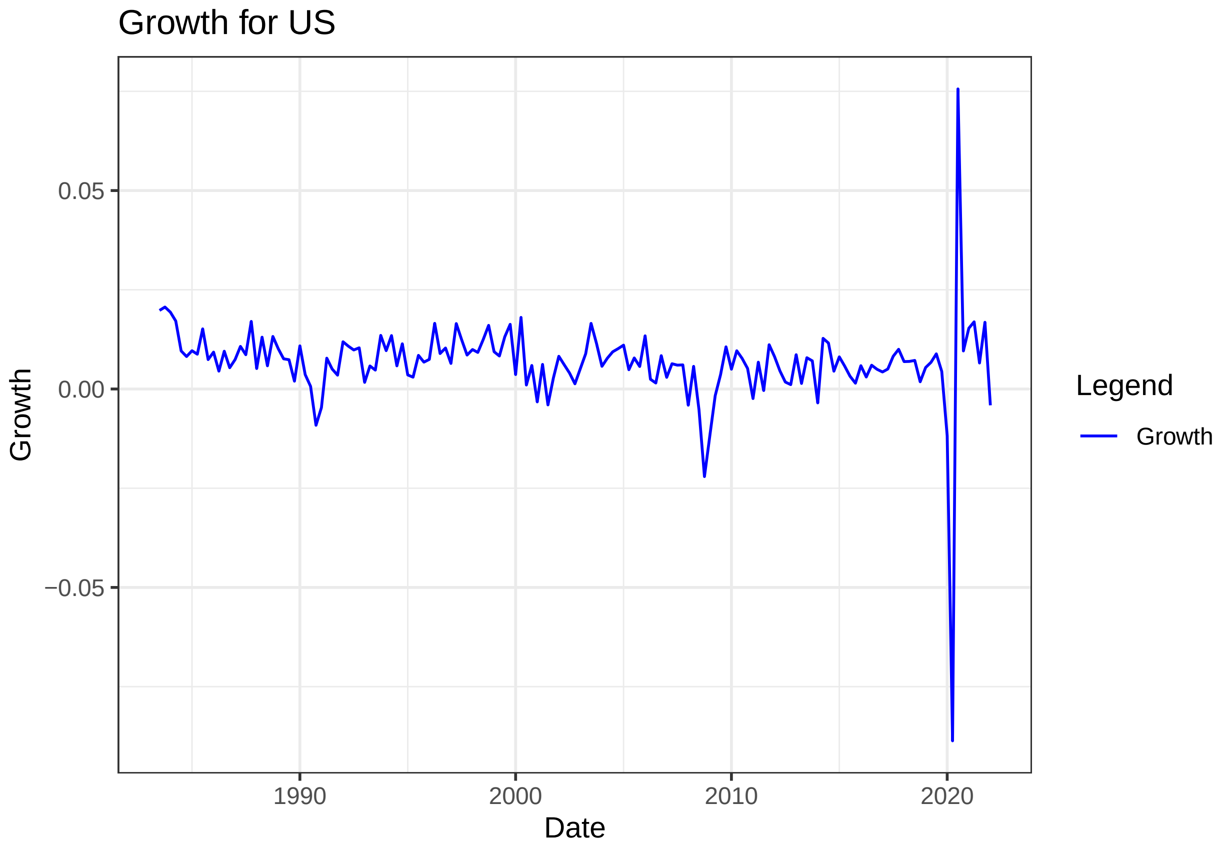
```
## F-statistic: 1.698 on 3 and 166 DF,  p-value: 0.1695
##
##
## Value of test-statistic is: -1.2384 18.0257 2.0924
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau3 -3.99 -3.43 -3.13
## phi2  6.22  4.75  4.07
## phi3  8.43  6.49  5.47
```

```
##null: unit root: fail to reject. Solution is to first difference

USdata <- USdata %>%
  mutate(growth = difference(logGDP)) %>% drop_na() ##growth

summary(ur.df(USdata$growth)) ##reject null of unit root
```

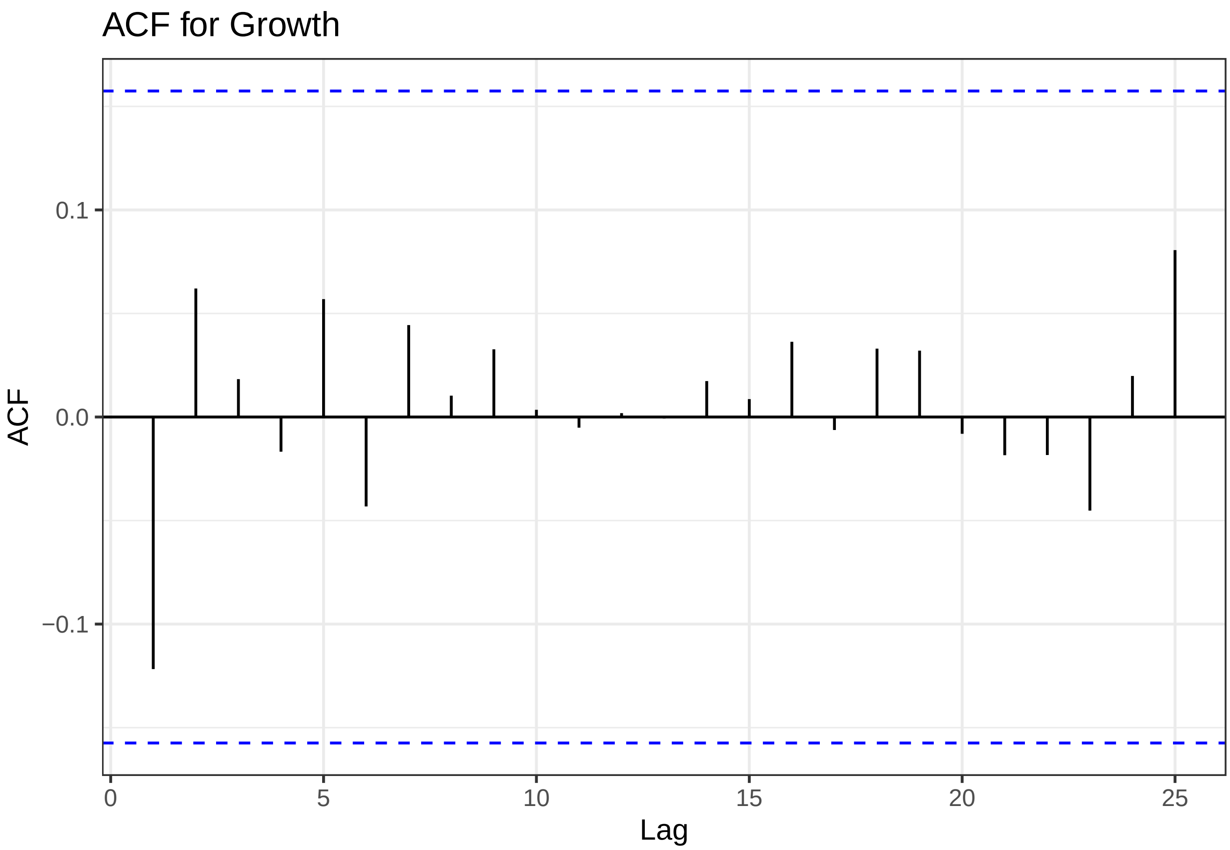
```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.088640  0.000889  0.004187  0.007152  0.088461
##
## Coefficients:
##            Estimate Std. Error t value Pr(>|t|)
## z.lag.1    -0.61047    0.09964  -6.127 7.45e-09 ***
## z.diff.lag -0.28225    0.07747  -3.643 0.000369 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01237 on 151 degrees of freedom
## Multiple R-squared:  0.472,  Adjusted R-squared:  0.465
## F-statistic:  67.5 on 2 and 151 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -6.1267
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau1 -2.58 -1.95 -1.62
```

```
cols <- c("Growth" = "blue")
ggplot(data=USdata, aes(x=date)) +
  theme_bw() +
  geom_line(aes(y=growth, color="Growth")) +
```

```
  scale_colour_manual(name="Legend",values=cols) +
  labs(x = "Date", y = "Growth", title = "Growth for US")
```




```
##Question c
ggAcf(USdata$growth, lag.max = 25, main = "ACF for Growth") + theme_bw()
```

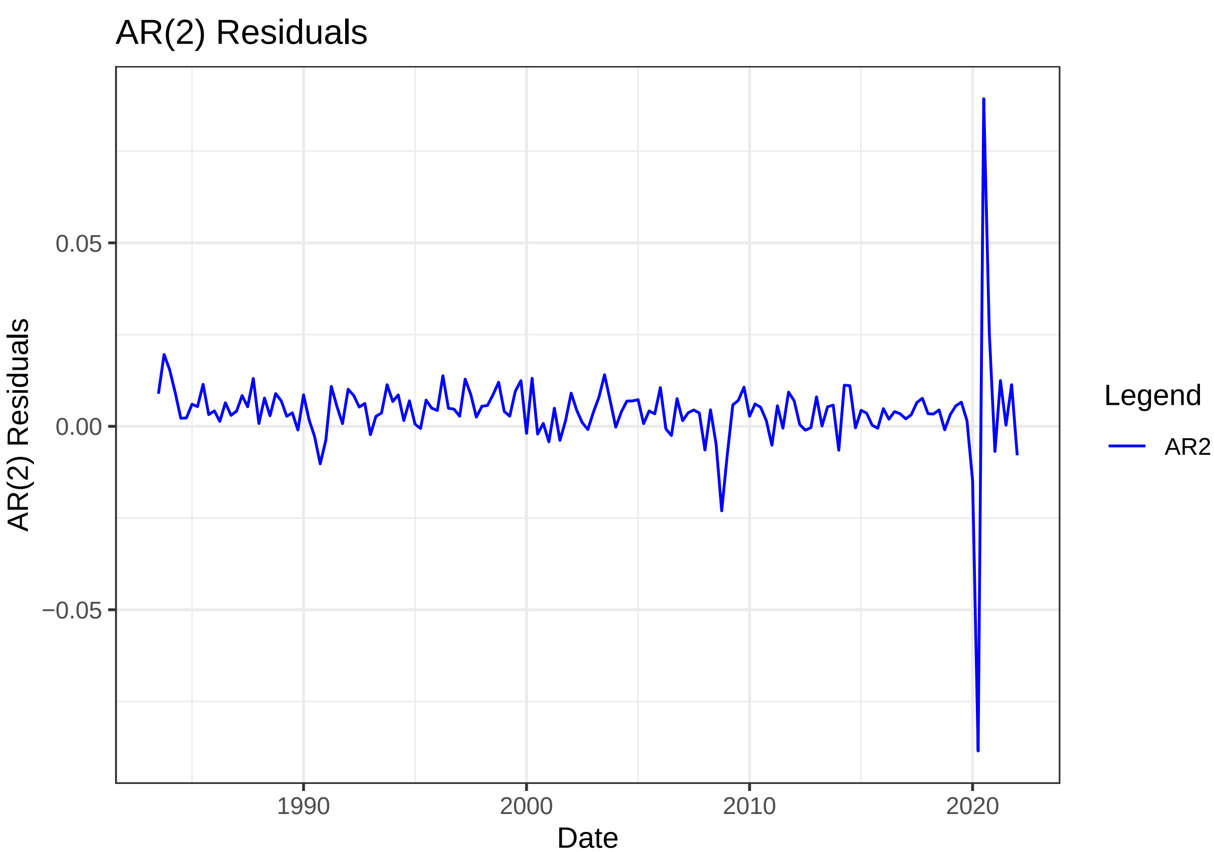



```
#Question d
ar1 <- arima(USdata$logGDP, c(1,1,0)) ##ARIMA(1,1,0) #first difference included I(1)
ar2 <- arima(USdata$logGDP, c(2,1,0)) ##ARIMA(2,1,0)
ma1 <- arima(USdata$logGDP, c(0,1,1)) ##ARIMA(0,1,1)
ma2 <- arima(USdata$logGDP, c(0,1,2)) ##ARIMA(0,1,2)

##Question e
stargazer(ar1, ar2, ma1, ma2, type = "latex")
```

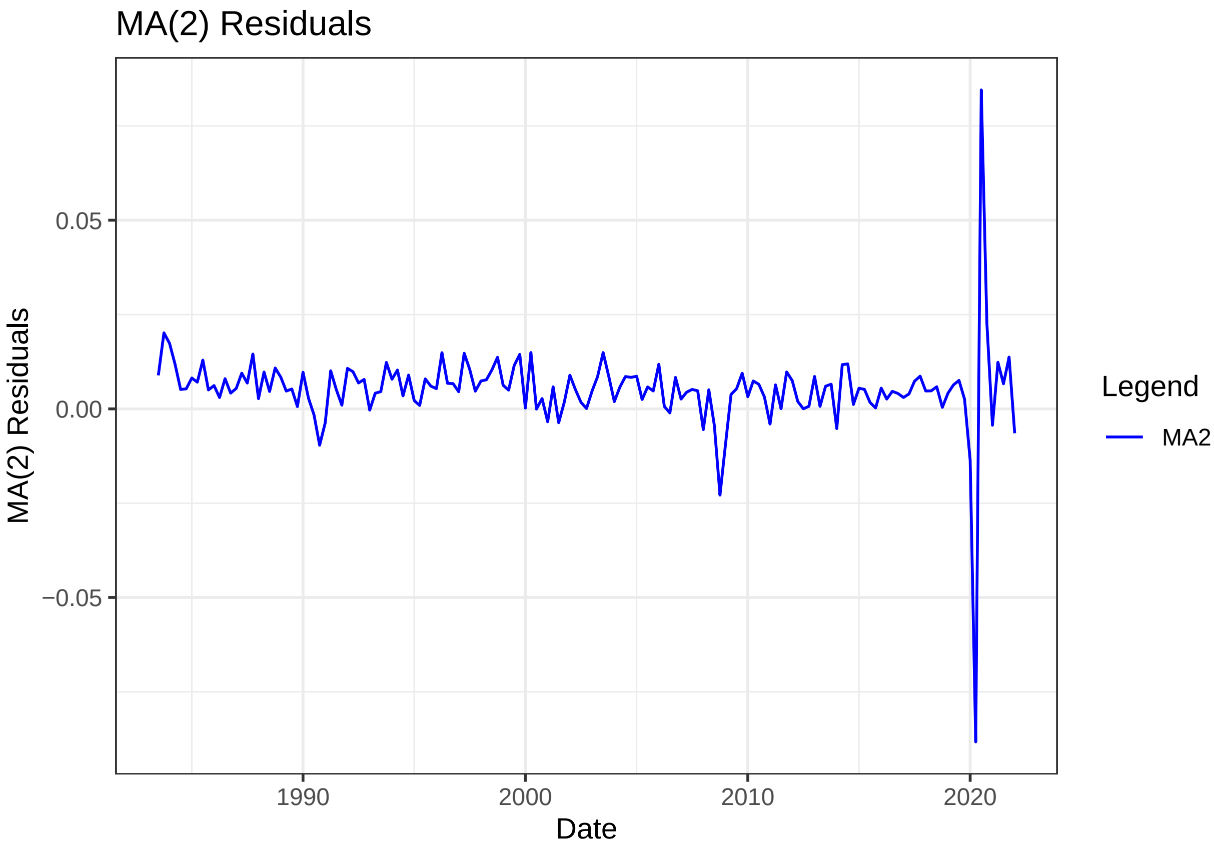
```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Mon, Mar 20, 2023 - 13:48:44
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lcccc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{4}{c}{\textit{Dependent variable:}} \\
## \cline{2-5}
## \\[-1.8ex] &  &  &  &  \\
## \\[-1.8ex] & (1) & (2) & (3) & (4)\\
## \hline \\[-1.8ex]
##  ar1 & 0.158$^{**}$ & 0.117 &  &  \\
##   & (0.080) & (0.078) &  &  \\
##   & & & & \\
##  ar2 &  & 0.278$^{***}$ &  &  \\
```

```
##   &  & (0.078) &  &  \\
##   & & & & \\
##  ma1 &  &  & 0.106 & 0.068 \\
##   &  &  & (0.067) & (0.082) \\
##   & & & & \\
##  ma2 &  &  &  & 0.214$^{***}$ \\
##   &  &  &  & (0.072) \\
##   & & & & \\
## \hline \\[-1.8ex]
## Observations & 154 & 154 & 154 & 154 \\
## Log Likelihood & 451.734 & 457.784 & 451.070 & 455.214 \\
## $\sigma^{2}$ & 0.0002 & 0.0002 & 0.0002 & 0.0002 \\
## Akaike Inf. Crit. & $-$899.469 & $-$909.569 & $-$898.141 & $-$904.429 \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{4}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
```

```
##plots
cols <- c("AR2" = "blue")
ggplot(data=USdata, aes(x=date)) +
  theme_bw() +
  geom_line(aes(y=ar2$residuals, color="AR2")) +
  scale_colour_manual(name="Legend",values=cols) +
  labs(x = "Date", y = "AR(2) Residuals", title = "AR(2) Residuals")
```



```r
cols <- c("MA2" = "blue")
ggplot(data=USdata, aes(x=date)) +
  theme_bw() +
  geom_line(aes(y=ma2$residuals, color="MA2")) +
  scale_colour_manual(name="Legend",values=cols) +
  labs(x = "Date", y = "MA(2) Residuals", title = "MA(2) Residuals")
```




```r
Box.test(ar2$residuals, lag = 10, type = "Ljung")
```

```
##
##  Box-Ljung test
##
## data:  ar2$residuals
## X-squared = 11.642, df = 10, p-value = 0.3097
```

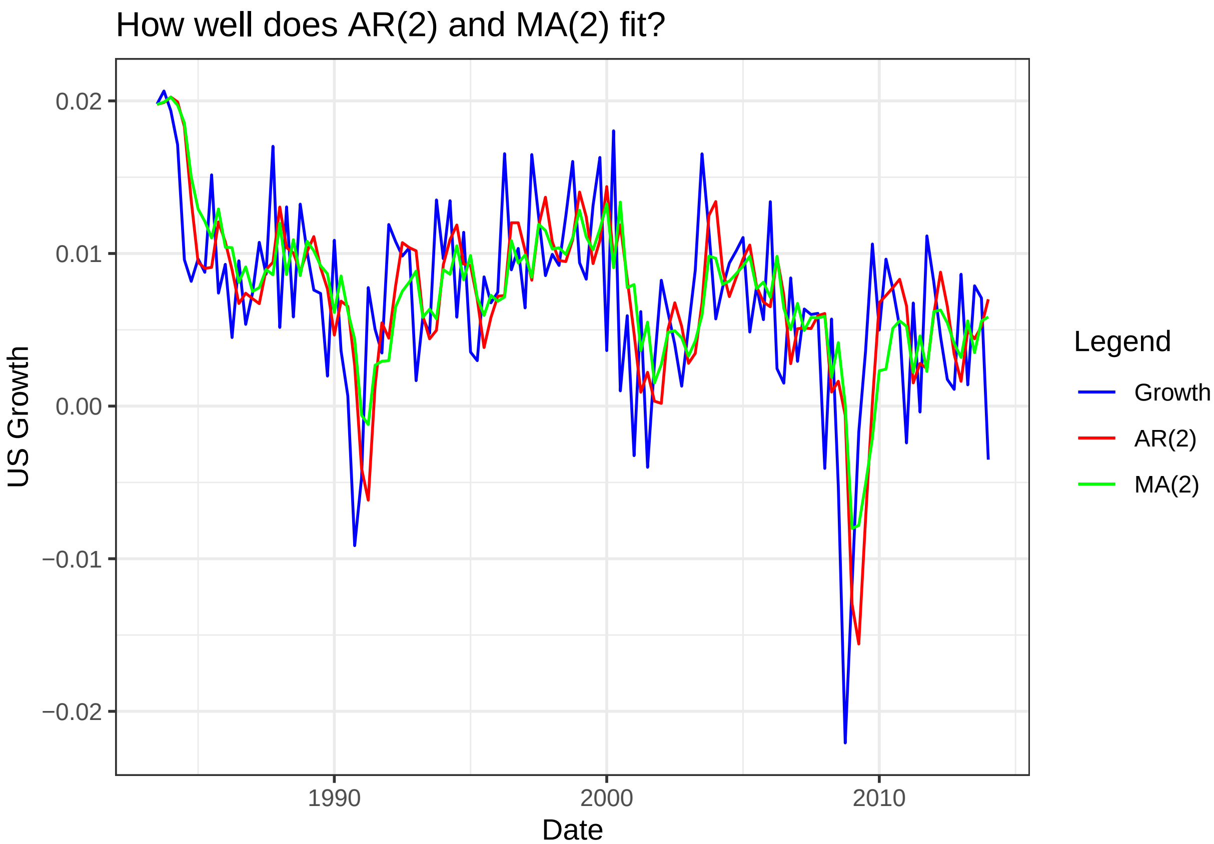
```r
Box.test(ma2$residuals, lag = 10, type = "Ljung")
```

```
##
##  Box-Ljung test
##
## data:  ma2$residuals
## X-squared = 8.7718, df = 10, p-value = 0.5539
```

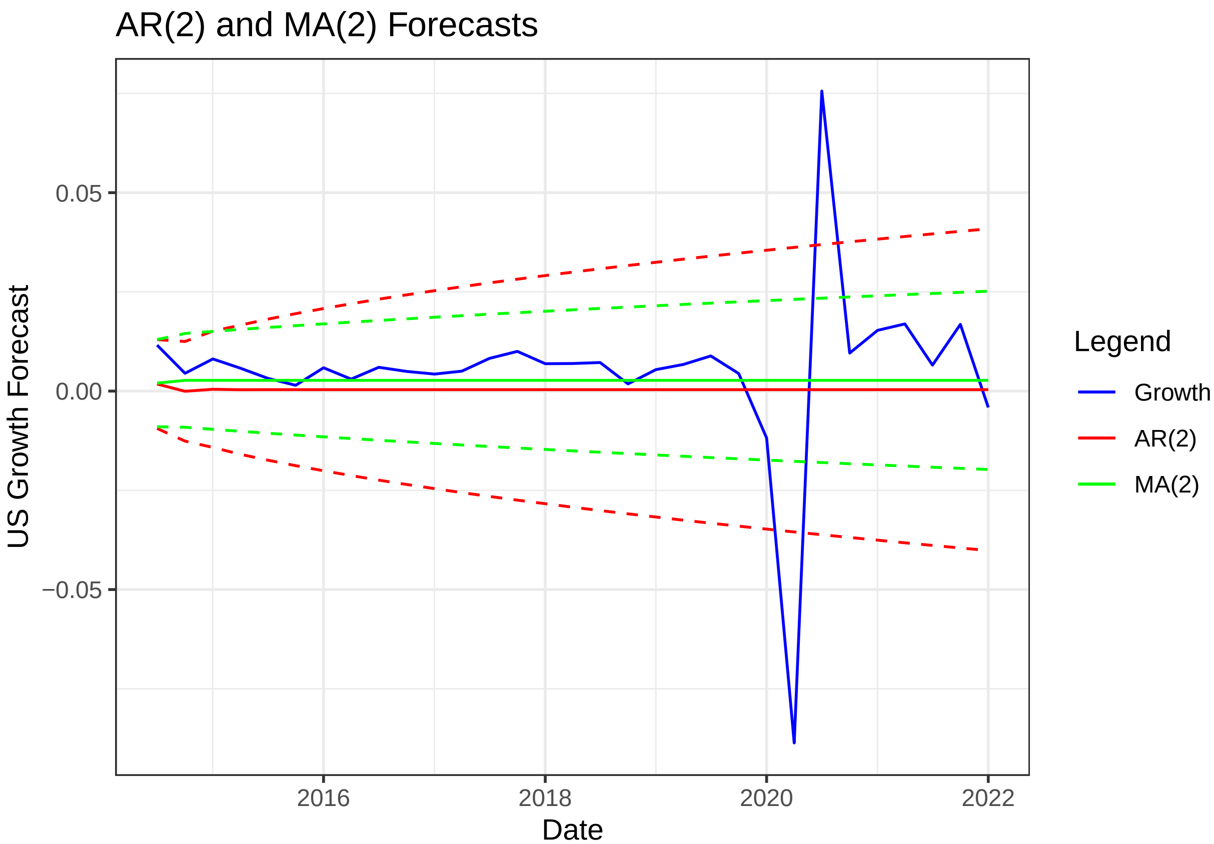
```r
##Question f
#Train/Test rule of thumb: 80/20 split
154*0.8
```

```
## [1] 123.2
```

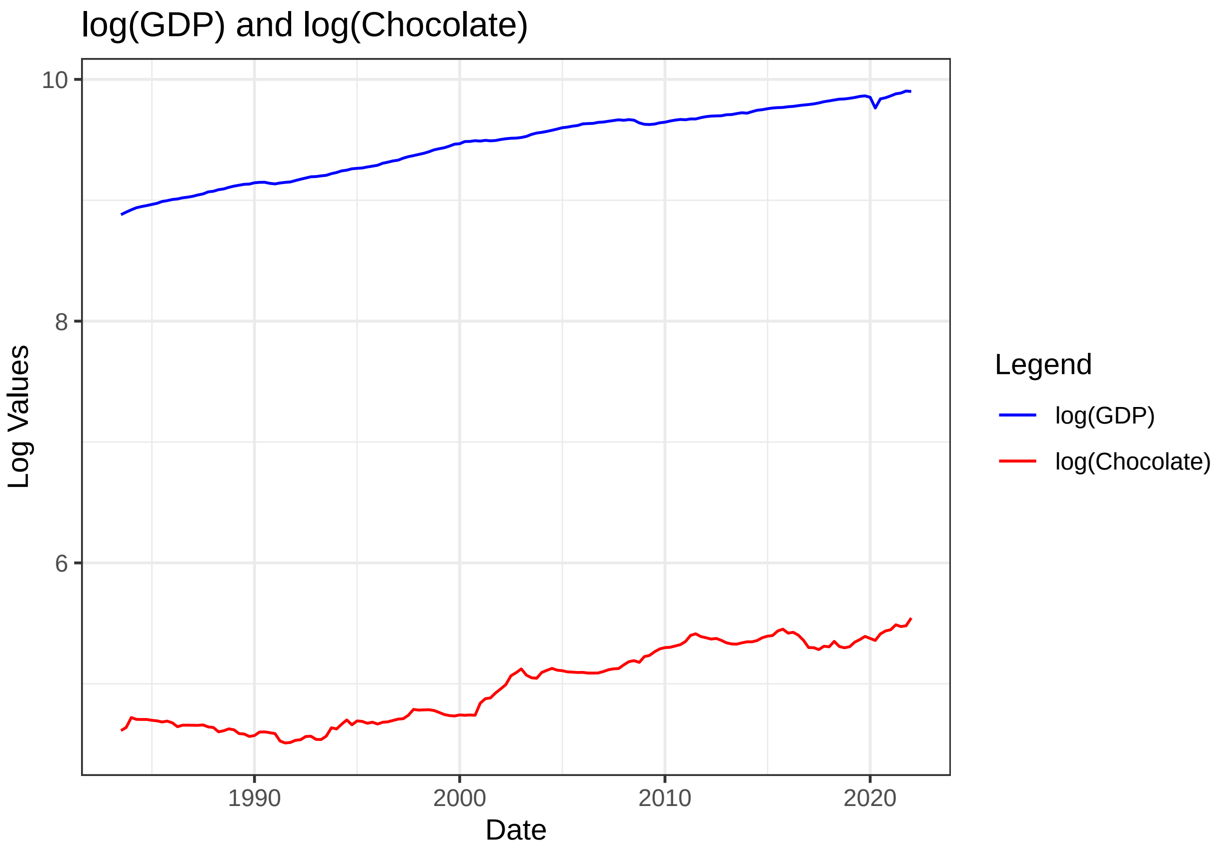
```r
train <- filter(USdata ,date < as.Date('2014-04-01'))
test = filter(USdata, date>= as.Date('2014-07-01'))

##y hats
ar2 <- arima(train$growth ,order = c(2,1,0))
ar2fit <- fitted(ar2)
ar2pred <- predict(ar2, n.ahead = nrow(test))
ma2 <- arima(train$growth ,order = c(0,1,2))
ma2fit <- fitted(ma2)
ma2pred <- predict(ma2, n.ahead = nrow(test))

##upper and lower bounds of forecasts
test$ar2pred <- ar2pred$pred
test$ar2low <- ar2pred$pred-1.96*ar2pred$se
test$ar2high <- ar2pred$pred + 1.96*ar2pred$se

test$ma2pred <- ma2pred$pred
test$ma2low <- ma2pred$pred-1.96*ma2pred$se
test$ma2high <- ma2pred$pred + 1.96*ma2pred$se

train$ar2fit <- ar2fit
train$ma2fit <- ma2fit


cols = c("Growth" = "blue", "AR(2)" = "red", "MA(2)" = "green")
ggplot(data = train, aes(x = date)) +
  geom_line(aes(y=growth, color="Growth"))+
  geom_line(aes(y=ar2fit, color="AR(2)")) +
  geom_line(aes(y=ma2fit, color="MA(2)")) +
  scale_colour_manual(name="Legend",values=cols) +
  theme_bw() +
  labs(x="Date",
       y="US Growth", title= "How well does AR(2) and MA(2) fit?")
```



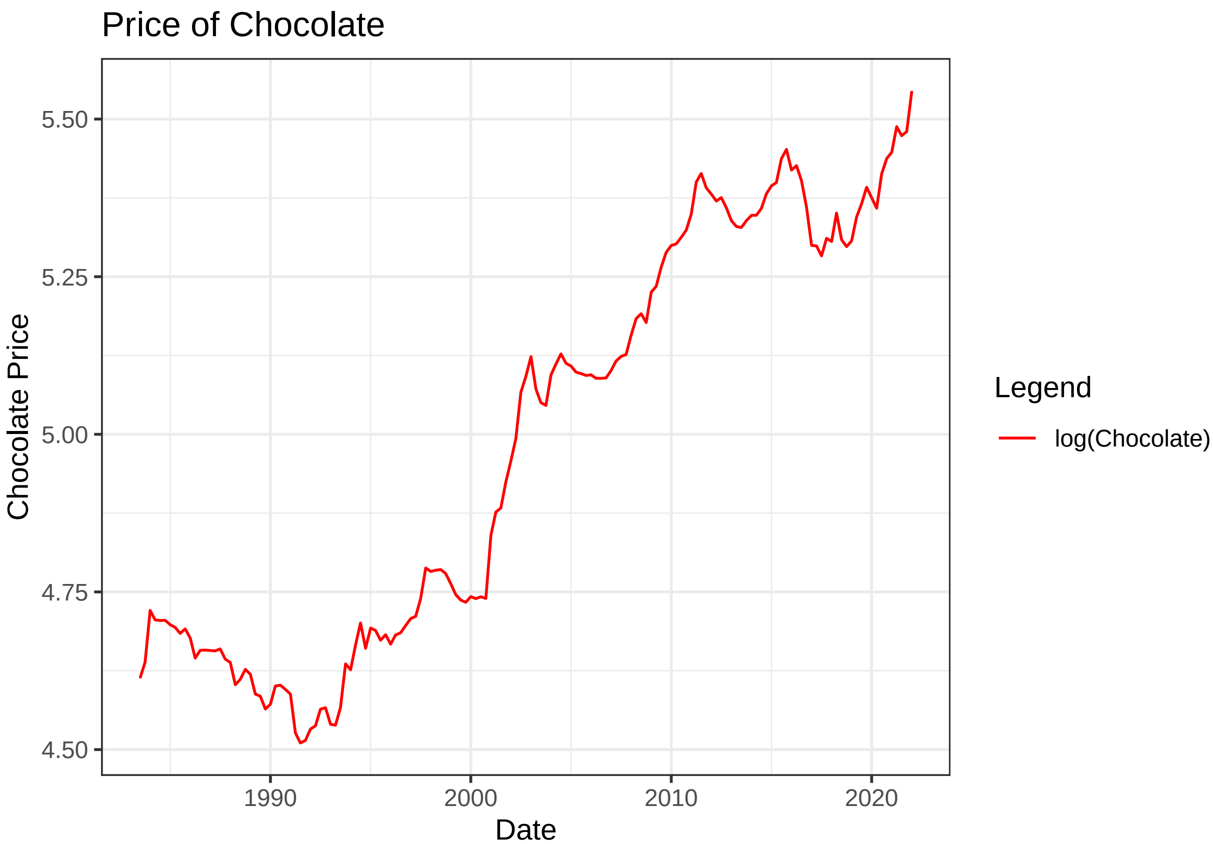

```
ggplot(data=test, aes(x=date)) +
  geom_line(aes(y=growth, color="Growth"))+
  geom_line(aes(y=ar2pred, color="AR(2)"))+
  geom_line(aes(y=ar2low), color="red", linetype = "dashed")+
  geom_line(aes(y=ar2high), color="red", linetype = "dashed") +
  geom_line(aes(y=ma2pred, color="MA(2)"))+
  geom_line(aes(y=ma2low), color="green", linetype = "dashed") +
  geom_line(aes(y=ma2high), color="green", linetype = "dashed") +
  theme_bw() +
  scale_colour_manual(name="Legend",values=cols) +
  labs(x="Date",
       y="US Growth Forecast", title= "AR(2) and MA(2) Forecasts")
```




```r
##question g
##this section requires me to comment on forecasting power



#############
##PART 2
attach(USdata)
##Question a
##comment on how chocolate could impact GDP
cols = c("log(GDP)" = "blue", "log(Chocolate)" = "red")
ggplot(data=USdata, aes(x=date)) +
  theme_bw() +
  geom_line(aes(y=log(Chocolate), color="log(Chocolate)")) +
  geom_line(aes(y=logGDP, color="log(GDP)")) +
  scale_colour_manual(name="Legend",values=cols) +
  labs(x = "Date", y = "Log Values", title = "log(GDP) and log(Chocolate)")
```




```r
##question b
cols = c("log(Chocolate)" = "red")
ggplot(data=USdata, aes(x=date)) +
  theme_bw() +
  geom_line(aes(y=log(Chocolate), color="log(Chocolate)")) +
  scale_colour_manual(name="Legend",values=cols) +
  labs(x = "Date", y = "Chocolate Price", title = "Price of Chocolate")
```




```r
##Question c
summary(ur.df(log(USdata$Chocolate), type = "trend")) ##null: unit root: fail to reject. Solution is to
```

```
##
## ###############################################
## # 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
## -0.065356 -0.011962 -0.001064  0.010458  0.088682
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.1809593  0.0826088   2.191  0.03004 *
## z.lag.1     -0.0404215  0.0185413  -2.180  0.03082 *
## tt           0.0003221  0.0001340   2.403  0.01750 *
## z.diff.lag   0.2252387  0.0802536   2.807  0.00568 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

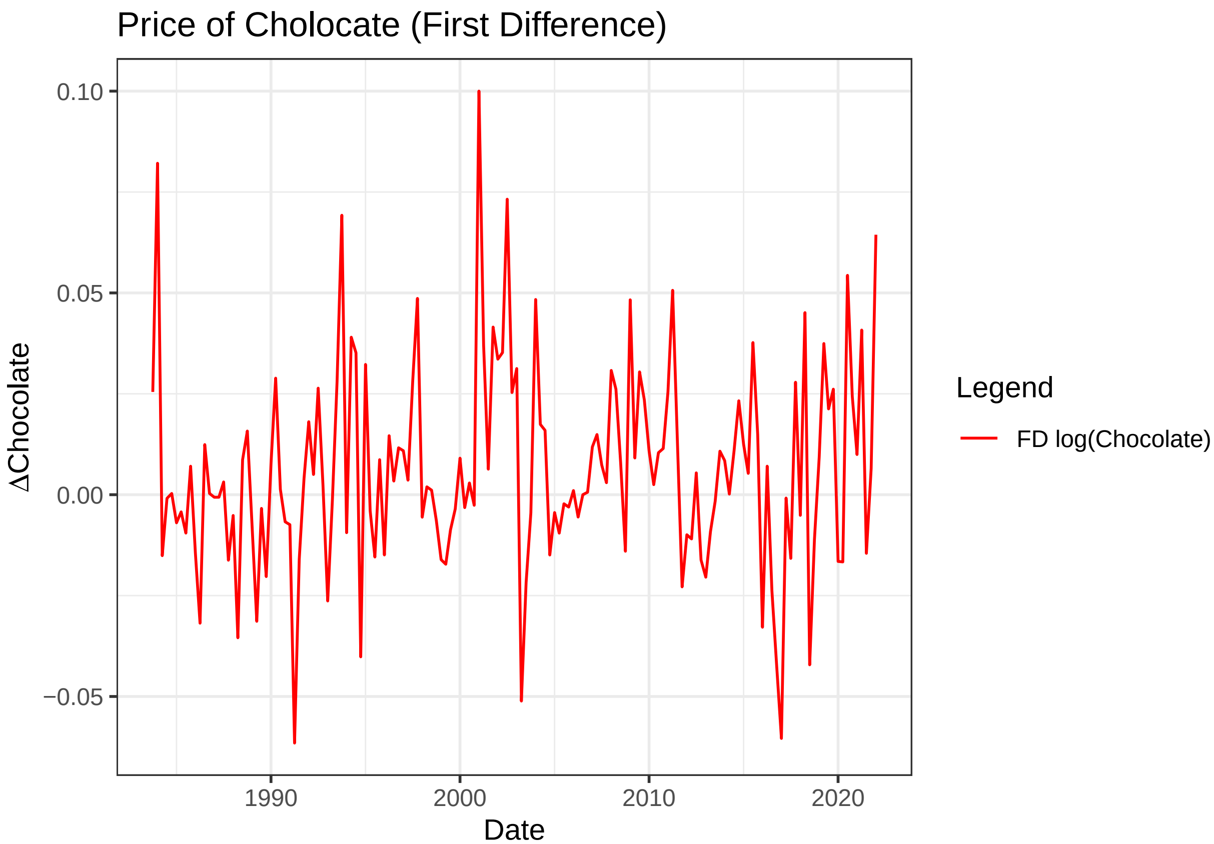
```
## Residual standard error: 0.02453 on 149 degrees of freedom
## Multiple R-squared:  0.08077,    Adjusted R-squared:  0.06226
## F-statistic: 4.364 on 3 and 149 DF,  p-value: 0.005611
##
##
## Value of test-statistic is: -2.1801 3.7272 2.9178
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau3 -3.99 -3.43 -3.13
## phi2  6.22  4.75  4.07
## phi3  8.43  6.49  5.47
```

```
USdata <- USdata %>%
  mutate(Coco = difference(log(Chocolate))) %>% drop_na() ##Chocolate FD

summary(ur.df(USdata$Coco)) ##reject null there is a unit root
```

```
##
## ###############################################
## # 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
## -0.059883 -0.007503  0.001126  0.015551  0.100351
##
## Coefficients:
##            Estimate Std. Error t value Pr(>|t|)
## z.lag.1    -0.70508    0.09838  -7.167 3.23e-11 ***
## z.diff.lag -0.07826    0.08034  -0.974    0.332
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0245 on 150 degrees of freedom
## Multiple R-squared:  0.3923, Adjusted R-squared:  0.3842
## F-statistic: 48.41 on 2 and 150 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -7.1672
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau1 -2.58 -1.95 -1.62
```

```
cols = c("FD log(Chocolate)" = "red")
ggplot(data=USdata, aes(x= date)) +
  theme_bw() +
  geom_line(aes(y=Coco, color="FD log(Chocolate)")) +
```

```
  scale_colour_manual(name="Legend",values=cols) +
  labs(x = "Date", y = expression(Delta*Chocolate),
       title = "Price of Cholocate (First Difference)")
```




```
#Question c
VARDATA <- ts(cbind(USdata$growth,USdata$Coco),names = c("growth","Coco"))
print(VARselect(VARDATA,type="const", lag.max = 5)) #1 lag required according to ICs
```

```
## $selection
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      1      1      1      1
##
## $criteria
##                    1             2             3             4             5
## AIC(n) -1.638028e+01 -1.635542e+01 -1.631104e+01 -1.628878e+01 -1.624409e+01
## HQ(n)  -1.633113e+01 -1.627351e+01 -1.619636e+01 -1.614134e+01 -1.606389e+01
## SC(n)  -1.625931e+01 -1.615381e+01 -1.602879e+01 -1.592588e+01 -1.580056e+01
## FPE(n)  7.693799e-08  7.887741e-08  8.246444e-08  8.433400e-08  8.820909e-08
```

```
#Question d
var1 <- VAR(VARDATA ,p=1 , type="const")
print(summary(var1))
```

```
##
## VAR Estimation Results:
## =========================
## Endogenous variables: growth, Coco
## Deterministic variables: const
```

```
## Sample size: 153
## Log Likelihood: 821.695
## Roots of the characteristic polynomial:
## 0.2201 0.1305
## Call:
## VAR(y = VARDATA, p = 1, type = "const")
##
##
## Estimation results for equation growth:
## =======================================
## growth = growth.l1 + Coco.l1 + const
##
##            Estimate Std. Error t value Pr(>|t|)
## growth.l1 -0.135127   0.081248  -1.663   0.0984 .
## Coco.l1    0.008543   0.036754   0.232   0.8165
## const      0.007377   0.001066   6.920 1.22e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.01123 on 150 degrees of freedom
## Multiple R-Squared: 0.01812, Adjusted R-squared: 0.005029
## F-statistic: 1.384 on 2 and 150 DF,  p-value: 0.2537
##
##
## Estimation results for equation Coco:
## =====================================
## Coco = growth.l1 + Coco.l1 + const
##
##            Estimate Std. Error t value Pr(>|t|)
## growth.l1 -0.191046   0.179696  -1.063  0.28942
## Coco.l1    0.224680   0.081289   2.764  0.00643 **
## const      0.005925   0.002358   2.513  0.01304 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.02483 on 150 degrees of freedom
## Multiple R-Squared: 0.05191, Adjusted R-squared: 0.03927
## F-statistic: 4.107 on 2 and 150 DF,  p-value: 0.01835
##
##
##
## Covariance matrix of residuals:
##           growth      Coco
## growth 1.260e-04 2.245e-05
## Coco   2.245e-05 6.164e-04
##
## Correlation matrix of residuals:
##         growth    Coco
## growth 1.00000 0.08054
## Coco   0.08054 1.00000
```

```r
stargazer::stargazer(var1$varresult$growth, var1$varresult$Coco)
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Mon, Mar 20, 2023 - 13:48:45
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lcc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{2}{c}{\textit{Dependent variable:}} \\
## \cline{2-3}
## \\[-1.8ex] & \multicolumn{2}{c}{y} \\
## \\[-1.8ex] & (1) & (2)\\
## \hline \\[-1.8ex]
##  growth.l1 & $-$0.135$^{*}$ & $-$0.191 \\
##   & (0.081) & (0.180) \\
##   & & \\
##  Coco.l1 & 0.009 & 0.225$^{***}$ \\
##   & (0.037) & (0.081) \\
##   & & \\
##  const & 0.007$^{***}$ & 0.006$^{**}$ \\
##   & (0.001) & (0.002) \\
##   & & \\
## \hline \\[-1.8ex]
## Observations & 153 & 153 \\
## R$^{2}$ & 0.018 & 0.052 \\
## Adjusted R$^{2}$ & 0.005 & 0.039 \\
## Residual Std. Error (df = 150) & 0.011 & 0.025 \\
## F Statistic (df = 2; 150) & 1.384 & 4.107$^{**}$ \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{2}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
```

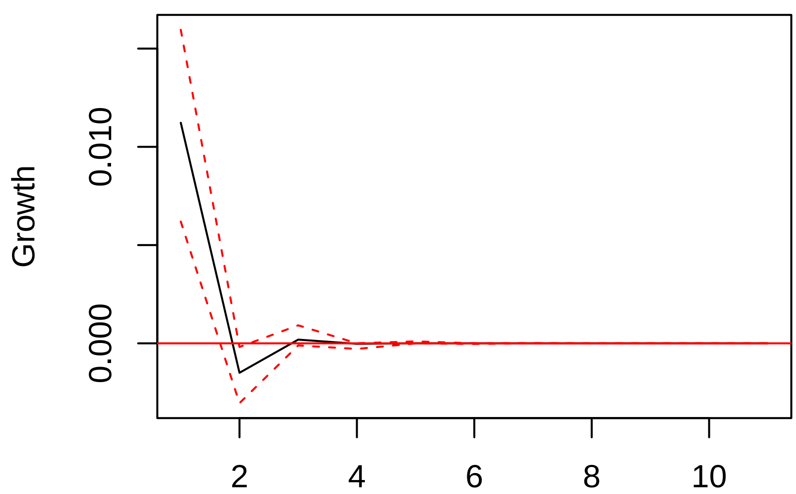
```r
summary(var1)
```

```
##
## VAR Estimation Results:
## =========================
## Endogenous variables: growth, Coco
## Deterministic variables: const
## Sample size: 153
## Log Likelihood: 821.695
## Roots of the characteristic polynomial:
## 0.2201 0.1305
## Call:
## VAR(y = VARDATA, p = 1, type = "const")
##
##
## Estimation results for equation growth:
## =======================================
```

```
## growth = growth.l1 + Coco.l1 + const
##
##            Estimate Std. Error t value Pr(>|t|)
## growth.l1 -0.135127   0.081248  -1.663   0.0984 .
## Coco.l1    0.008543   0.036754   0.232   0.8165
## const      0.007377   0.001066   6.920 1.22e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.01123 on 150 degrees of freedom
## Multiple R-Squared: 0.01812, Adjusted R-squared: 0.005029
## F-statistic: 1.384 on 2 and 150 DF,  p-value: 0.2537
##
##
## Estimation results for equation Coco:
## =====================================
## Coco = growth.l1 + Coco.l1 + const
##
##            Estimate Std. Error t value Pr(>|t|)
## growth.l1 -0.191046   0.179696  -1.063  0.28942
## Coco.l1    0.224680   0.081289   2.764  0.00643 **
## const      0.005925   0.002358   2.513  0.01304 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.02483 on 150 degrees of freedom
## Multiple R-Squared: 0.05191, Adjusted R-squared: 0.03927
## F-statistic: 4.107 on 2 and 150 DF,  p-value: 0.01835
##
##
##
## Covariance matrix of residuals:
##           growth      Coco
## growth 1.260e-04 2.245e-05
## Coco   2.245e-05 6.164e-04
##
## Correlation matrix of residuals:
##         growth    Coco
## growth 1.00000 0.08054
## Coco   0.08054 1.00000
```

```r
#Qeustion e
##Impulse Response Functions
p1 <- irf(var1, impulse = "growth", response = "growth", n.ahead = 10, boot = TRUE)
plot(p1, ylab = "Growth", main = "Growth shock to Growth")
```

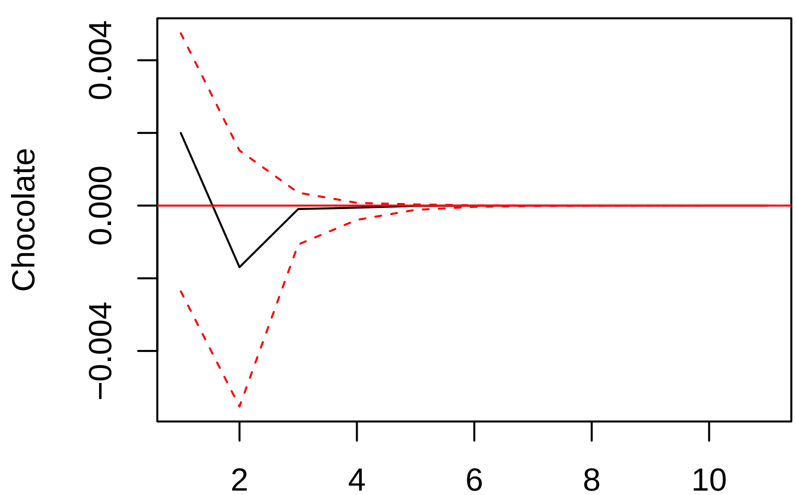
Growth shock to Growth



95 % Bootstrap CI, 100 runs

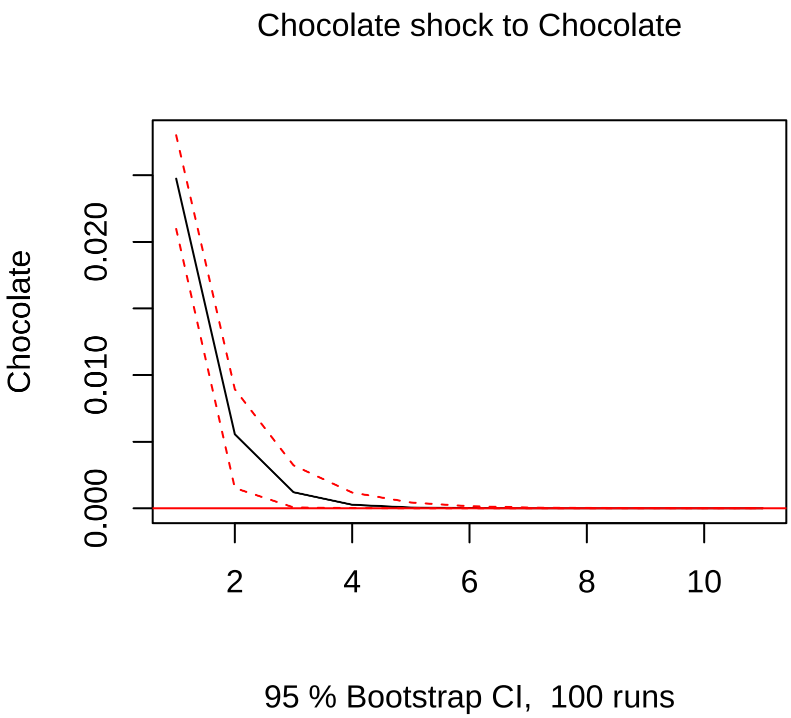
```
p2 <- irf(var1, impulse = "growth", response = "Coco", n.ahead = 10, boot = TRUE)
plot(p2, ylab = "Chocolate", main = "Growth shock to Chocolate")
```

Growth shock to Chocolate

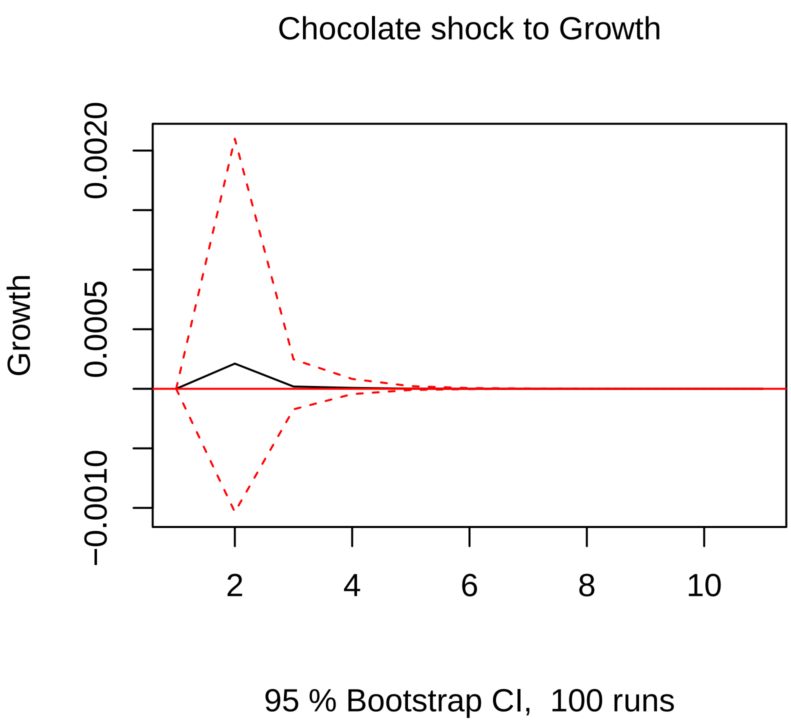


95 % Bootstrap CI, 100 runs

```
p3 <- irf(var1, impulse = "Coco", response = "Coco", n.ahead = 10, boot = TRUE)
plot(p3, ylab = "Chocolate", main = "Chocolate shock to Chocolate")
```




```r
p4 <- irf(var1, impulse = "Coco", response = "growth", n.ahead = 10, boot = TRUE)
plot(p4, ylab = "Growth", main = "Chocolate shock to Growth")
```




```r
##Question f
##Granger Causality
print(causality(var1, cause="Coco")$Granger)
```

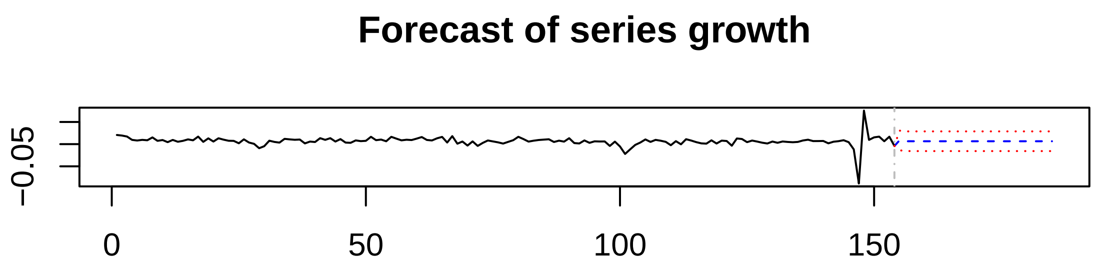
```
##
##  Granger causality H0: Coco do not Granger-cause growth
```

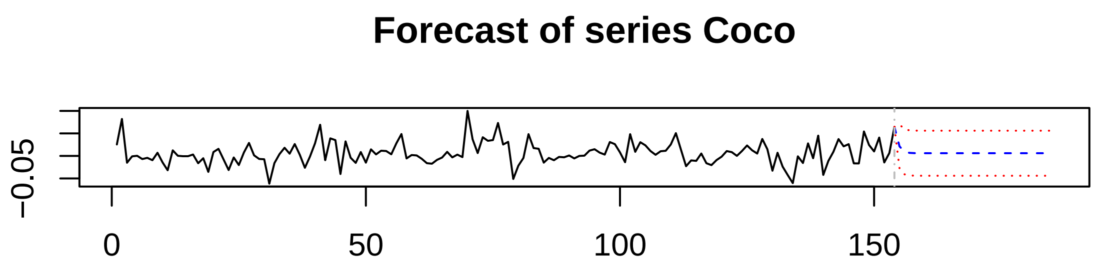
```
##
## data:  VAR object var1
## F-Test = 0.054028, df1 = 1, df2 = 300, p-value = 0.8164
```

```
print(causality(var1, cause="growth")$Granger)
```

```
##
##  Granger causality H0: growth do not Granger-cause Coco
##
## data:  VAR object var1
## F-Test = 1.1303, df1 = 1, df2 = 300, p-value = 0.2886
```

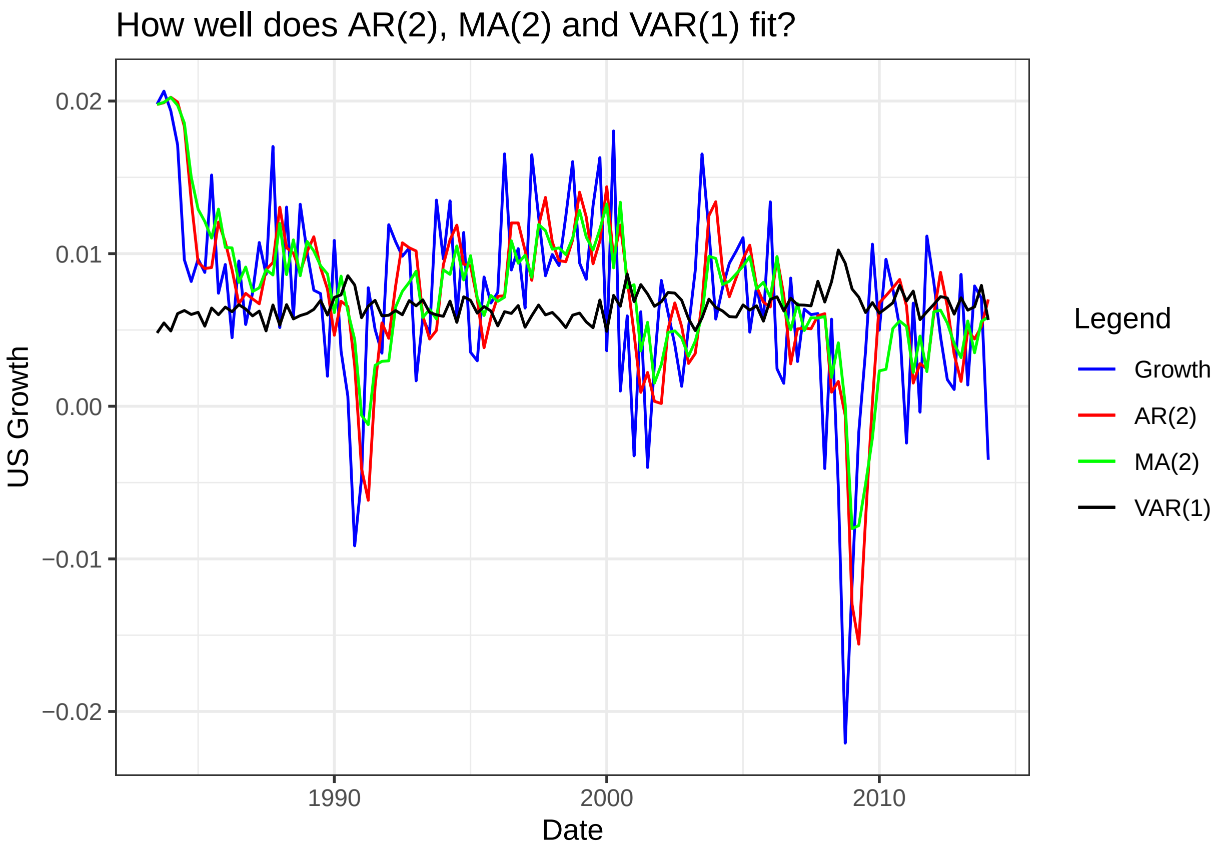
```
##Question g
##Var Forecasts
Varforecasts <- predict(var1, n.ahead = nrow(test))
Varfitted <- fitted(var1)
plot(Varforecasts)
```



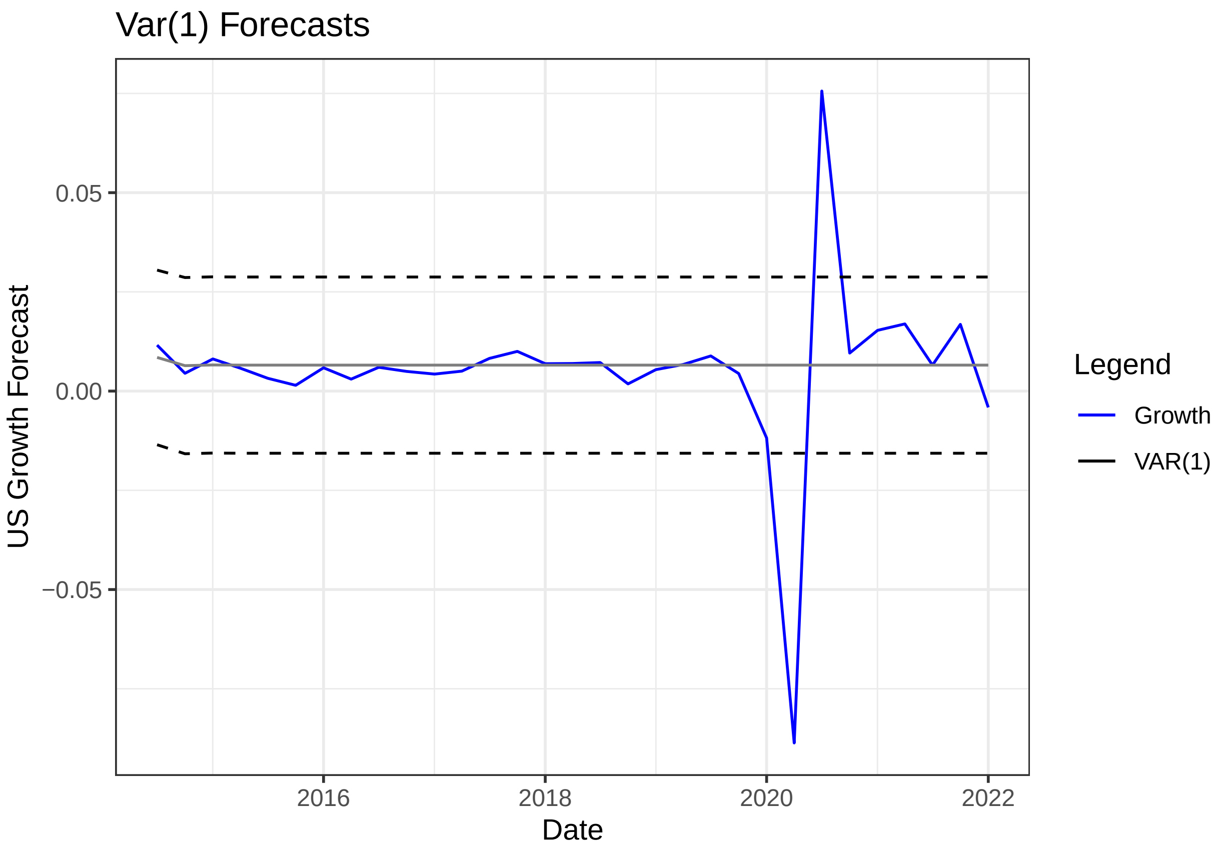



```
test$Varfc <- Varforecasts$fcst$growth[,1]
test$Varlow <- Varforecasts$fcst$growth[,2]
test$Varhigh <- Varforecasts$fcst$growth[,3]



train$Varfit <- Varfitted[1:123, 1]



cols = c("Growth" = "blue", "AR(2)" = "red", "MA(2)" = "green", "VAR(1)" = "black")
ggplot(data = train, aes(x = date)) +
  geom_line(aes(y=growth, color="Growth"))+
  geom_line(aes(y=ar2fit, color="AR(2)")) +
  geom_line(aes(y=ma2fit, color="MA(2)")) +
  geom_line(aes(y=Varfit, color="VAR(1)")) +
```

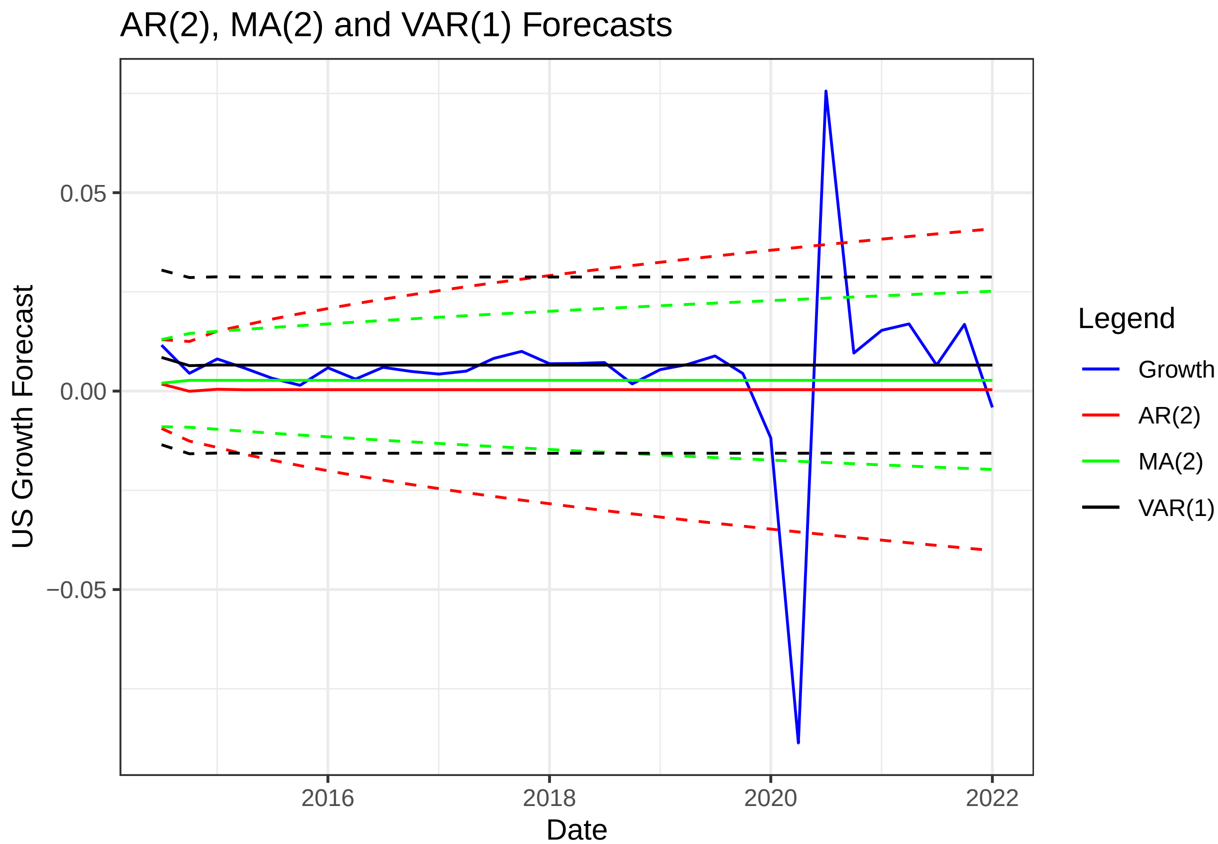
```
  scale_colour_manual(name="Legend",values=cols) +
  theme_bw() +
  labs(x="Date",
       y="US Growth", title= "How well does AR(2), MA(2) and VAR(1) fit?")
```




```
cols = c("Growth" = "blue", "VAR(1)" = "black")
ggplot(data=test, aes(x=date)) +
  geom_line(aes(y=growth, color="Growth")) +
  geom_line(aes(y=Varfc, color="Var(1)")) +
  geom_line(aes(y=Varlow), color="black", linetype = "dashed") +
  geom_line(aes(y=Varhigh), color="black", linetype = "dashed") +
  scale_colour_manual(name="Legend",values=cols) +
  theme_bw() +
  labs(x="Date",
       y="US Growth Forecast", title= "Var(1) Forecasts")
```




```
cols = c("Growth" = "blue", "AR(2)" = "red", "MA(2)" = "green", "VAR(1)" = "black")
ggplot(data=test, aes(x=date)) +
  geom_line(aes(y=growth, color="Growth"))+
  geom_line(aes(y=ar2pred, color="AR(2)"))+
  geom_line(aes(y=ar2low), color="red", linetype = "dashed")+
  geom_line(aes(y=ar2high), color="red", linetype = "dashed") +
  geom_line(aes(y=ma2pred, color="MA(2)"))+
  geom_line(aes(y=ma2low), color="green", linetype = "dashed") +
  geom_line(aes(y=ma2high), color="green", linetype = "dashed") +
  geom_line(aes(y=Varfc, color="VAR(1)")) +
  scale_colour_manual(name="Legend",values=cols) +
  geom_line(aes(y=Varlow), color="black", linetype = "dashed") +
  geom_line(aes(y=Varhigh), color="black", linetype = "dashed") +
  theme_bw() +
  labs(x="Date",
       y="US Growth Forecast", title= "AR(2), MA(2) and VAR(1) Forecasts")
```


AR(2), MA(2) and VAR(1) Forecasts
US Growth Forecast
0.05
0.00
−0.05
2016
2018
2020
2022
Date
Legend
Growth
AR(2)
MA(2)
VAR(1)