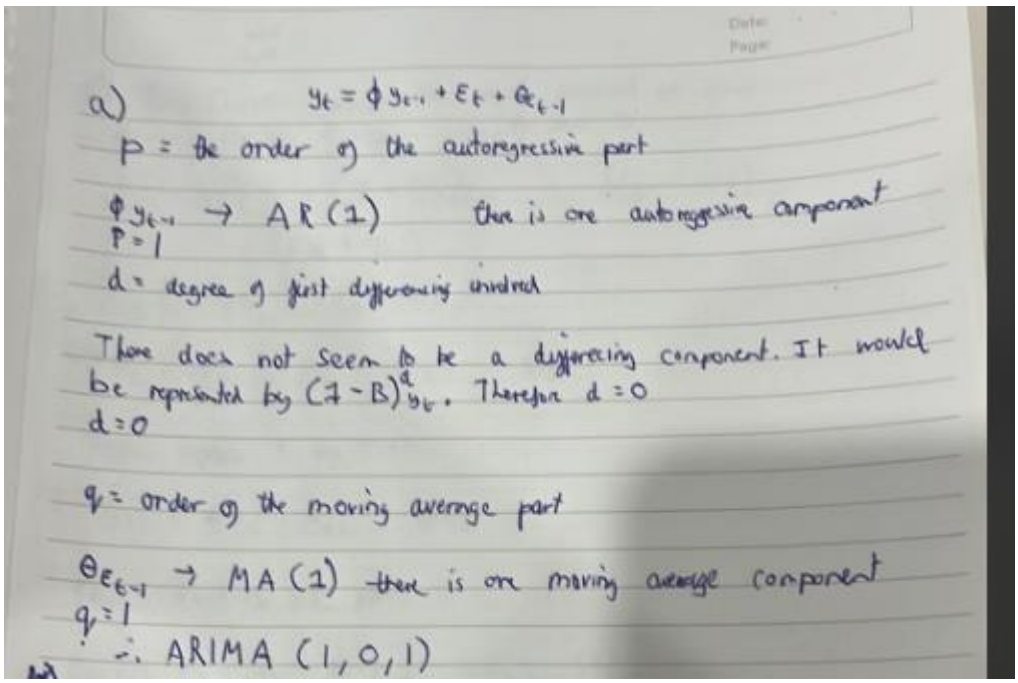
STA8122 Time Series Assignment 3

Aidan Van Klaveren

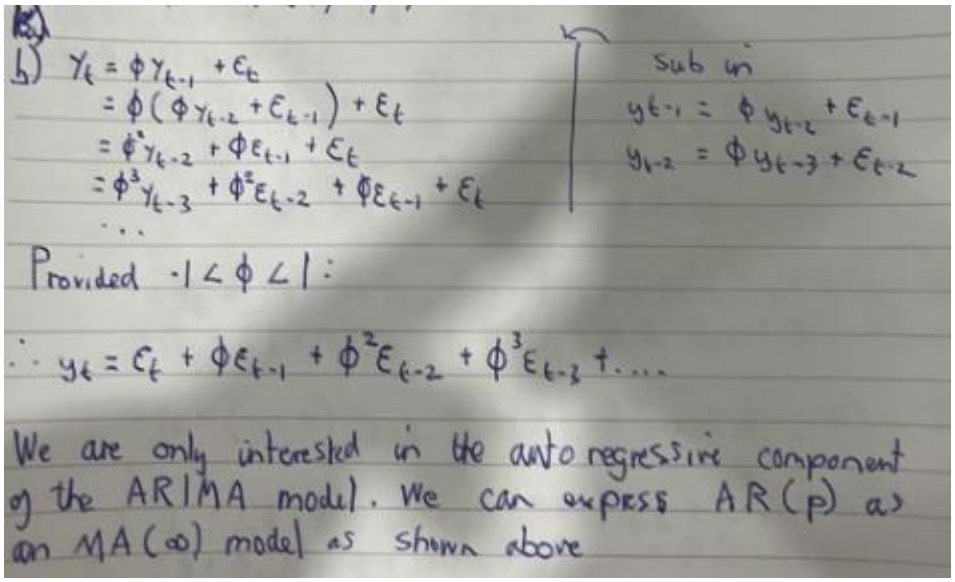
44588070

Question 1

a) What sort of ARIMA(p,d,q) model is stated in Equation (1)



b) Express this ARIMA model as a MA(infinity) model



c) Express yt+1|t, the point forecast for yt+1, as a function of y1,…yt, φ and θ.]

c) Express $y_{T+1|T}$, the point forecast for y_{T+1} as a function of $y_1, \ldots, y_T, \phi$ and θ

$$(1-\phi B)Y_t = (1+\theta B)\varepsilon_t \qquad (\varepsilon_t \sim WN)$$

$$[BY_t = Y_{t-1}]$$

$$\Rightarrow Y_t - \phi Y_{t-1} = \varepsilon_t + \theta \varepsilon_{t-1}$$

$$\Rightarrow Y_t = \phi Y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1} \quad (1)$$

Step2 replace t by $T+h$

$$Y_{T+1} = \phi Y_T + \varepsilon_{T+1} + \theta \varepsilon_T$$

Step 3: Find the AR form

By (1), $\varepsilon_t = Y_t - \phi Y_{t-1} - \theta \varepsilon_{t-1}$ (2)

Apply (2) iteratively

$$\varepsilon_t = Y_t - \phi Y_{t-1} - \theta \varepsilon_{t-1}$$
$$= Y_t - \phi Y_{t-1} - \theta (Y_{t-1} - \phi Y_{t-2} - \theta \varepsilon_{t-2})$$
$$= Y_t - \phi Y_{t-1} - \theta Y_{t-1} + \theta \phi Y_{t-2} + \theta^2 \varepsilon_{t-2}$$
$$= Y_t - (\phi+\theta) Y_{t-1} + \theta \phi Y_{t-2} + \theta^2 \varepsilon_{t-2}$$
$$= Y_t - (\phi+\theta) Y_{t-1} + \theta \phi Y_{t-2} + \theta^2 (Y_{t-2} - \phi Y_{t-3} - \theta \varepsilon_{t-3})$$
$$= Y_t - (\phi+\theta) Y_{t-1} + \theta \phi Y_{t-2} + \theta^2 Y_{t-2} + \theta^2 (-\phi Y_{t-3} - \theta \varepsilon_{t-3})$$
$$= Y_t - (\phi+\theta) Y_{t-1} + \theta(\phi+\theta) Y_{t-2} + \theta^2 (-\phi Y_{t-3} - \theta \varepsilon_{t-3})$$
$$= \ldots$$
$$= Y_t - (\phi+\theta) Y_{t-1} + \theta(\phi+\theta) Y_{t-2} + \theta^2(\phi+\theta) Y_{t-3} + \ldots$$
$$Y_t = (\phi+\theta) Y_{t-1} - \theta(\phi+\theta) Y_{t-2} + \theta^2(\phi+\theta) Y_{t-3} \ldots + \varepsilon_t$$

Step 4 Replace future Y by current and previous Y & replace epsilon

$$Y_{T+1} = \phi Y_T + \varepsilon_{T+1} + \theta \varepsilon_T$$
$$= \phi Y_T + \varepsilon_{T+1} + \theta [Y_T - (\phi+\theta) Y_{T-1} + \theta(\phi+\theta) Y_{T-2} - \theta^2(\phi+\theta) Y_{T-3} + \ldots]$$

Step 5 Replace future epsilon by zero and get the prediction

$$\hat{Y}_{T+1|T} = \phi Y_t + \theta [Y_T - (\phi+\theta) Y_{T-1} + \theta(\phi+\theta) Y_{T-2} - \theta^3(\phi+\theta) Y_{t-3} + \ldots]$$

Question 2

a) Generate a time plot and find a suitable Box-Cox transformation for the data

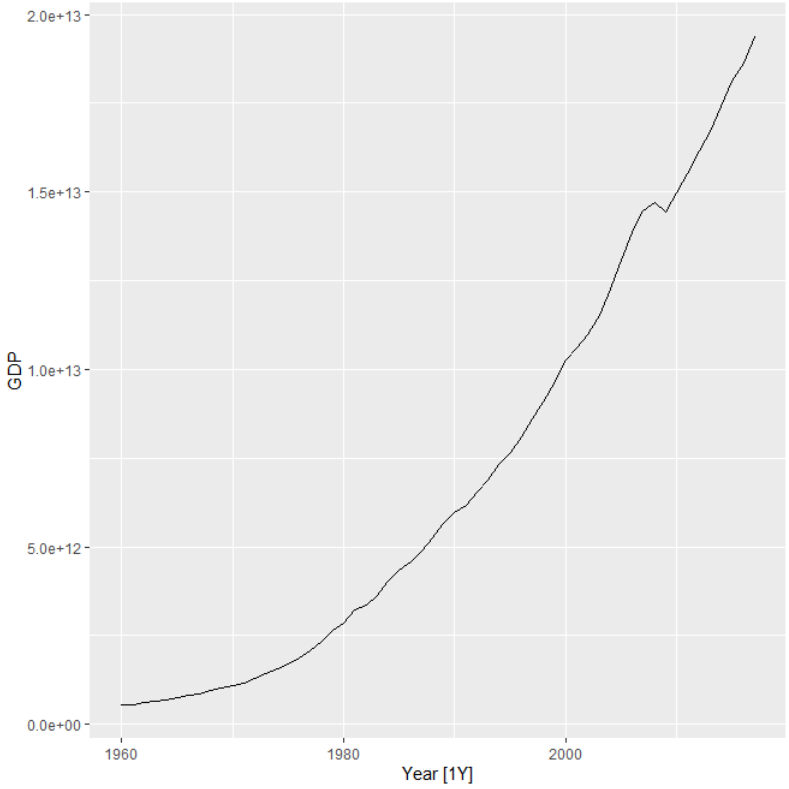
```
#load in the data global_economy
global_economy <- tsibbledata::global_economy

#create a new table for just USA data
us_economy <- global_economy %>%
  filter(Code == "USA") %>%
  as_tsibble(index = Year)

#can see the GDP numbers here
us_economy$GDP
```

The above code loads in the data and creates a new table with just data from the USA

```
#time series plot of GDP
us_economy %>%
  autoplot(GDP)
```



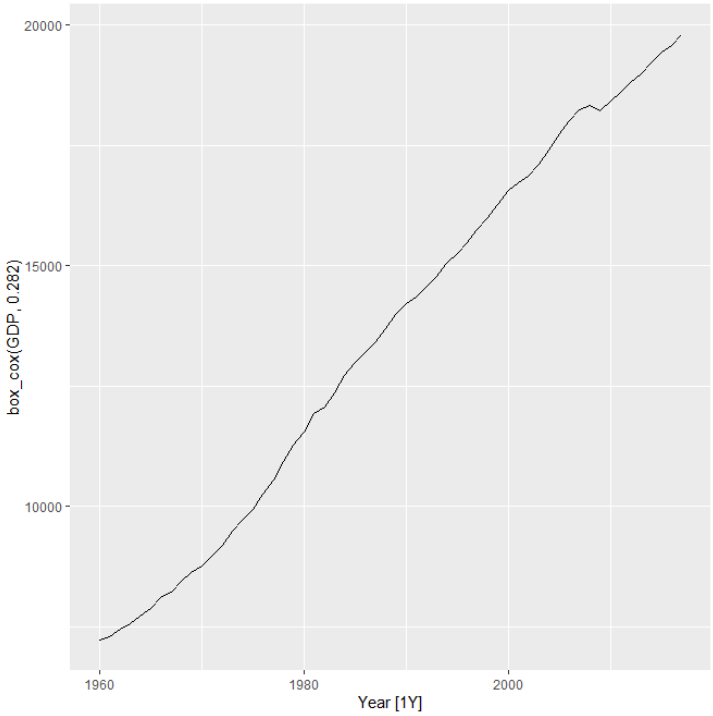
The Gross Domestic Product of the USA can be seen from the following graph. GDP looks like it is increasing exponentially over the period with a drop off point around 2008 which can be explained by the global financial crisis. After that period, it continues to increase rapidly.

```
#find suitable box-cox transformation
us_economy %>%
  features(GDP, guerrero)
```

```
# A tibble: 1 x 2
  Country       lambda_guerrero
  <fct>                   <dbl>
1 United States           0.282
```

The Guerrero method finds a suitable number to input into the Box-Cox transformation, represented by lamda. The recommended lambda to use 0.282.

```
#time series with transformation
us_economy %>%
  autoplot(box_cox(GDP, 0.282))
```



After fitting the variable with the recommended box-cox transformation value. We can see the results in the above chart. The line appears a lot smoother and linear. The effect of the decrease in 2008 can also be seen to be lessened.

b) Fit a ARIMA model to the Box-Cox transformed data. Report the fit with report

```
#fit ARIMA model with box-cox transformed data
fit <- us_economy %>%
  model(
    arima = ARIMA(box_cox(GDP,0.282))
  )
report(fit)
```

```
Series: GDP
Model: ARIMA(1,1,0) w/ drift
Transformation: box_cox(GDP, 0.282)

Coefficients:
          ar1  constant
      0.4586  118.3757
s.e.  0.1198    9.5203

sigma^2 estimated as 5497:  log likelihood=-325.42
AIC=656.83   AICc=657.29   BIC=662.96
```

The fitted model uses the base ARIMA model which has auto selected and ARIMA(1,1,0) as the fitting with drift. The model selected includes differencing (1-B)^1 and AR(1).The coefficient for the autoregressive component is shown with the constant. We can also see the estimated variance and log likelihood. The AICc score for the model is also calculated which we can use to compare against other models.

c) Write down the codes to fit ARIMA(0,1,0) w/ drift, ARIMA(0,1,1) w/ drift, and ARIMA(1,1,1) w? drift to the transformed data as well

```
#fit codes for models ARIMA(0,1,0) w/drift, ARIMA(0,1,1) w/drift and ARIMA(1,1,1) w/drift
fit1 <- us_economy %>%
  model(
    arima010 = ARIMA(box_cox(GDP,0.282) ~ pdq(0,1,0)),
    arima011 = ARIMA(box_cox(GDP,0.282) ~ pdq(0,1,1)),
    arima111 = ARIMA(box_cox(GDP,0.282) ~ pdq(1,1,1))
  )
glance(fit1) %>% arrange(AICc)
```

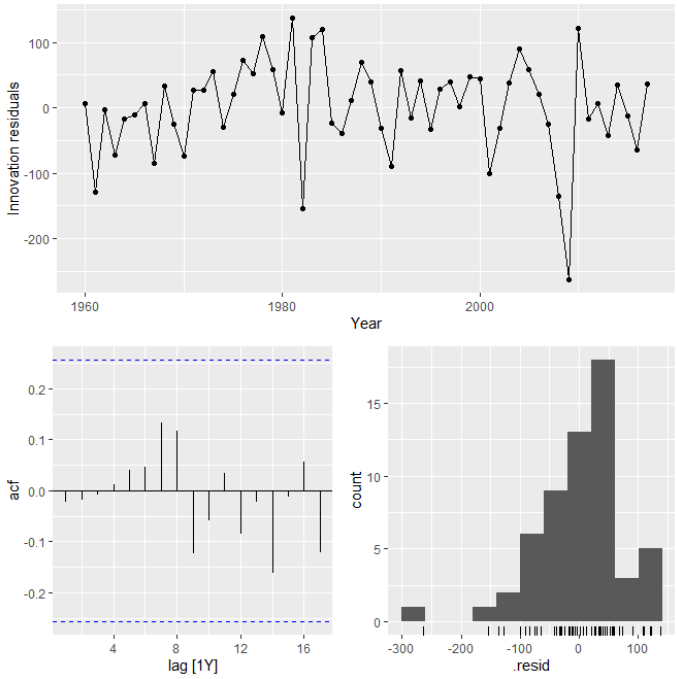
```
# A tibble: 3 x 9
  Country       .model   sigma2 log_lik   AIC  AICc   BIC ar_roots  ma_roots
  <fct>         <chr>     <dbl>   <dbl> <dbl> <dbl> <dbl> <list>    <list>
1 United States arima011  5708.    -326.  659.  659.  665. <cpl [0]> <cpl [1]>
2 United States arima111  5599.    -325.  659.  660.  667. <cpl [1]> <cpl [1]>
3 United States arima010  6797.    -332.  668.  668.  672. <cpl [0]> <cpl [0]>
```

The results from the models can be seen from the table. The table is ordered from smallest AICc to largest. The smaller the AICc the better the model.

d) Among the models in the previous two questions, choose what you think is the best model according to AICc, and describe this models' residual diagnostics

From the models in part b and c, we can see that 657.29 is the lowest AICc. We would therefore choose the ARIMA(1,1,0) with drift as the best model. We need to examine the residuals to make sure.

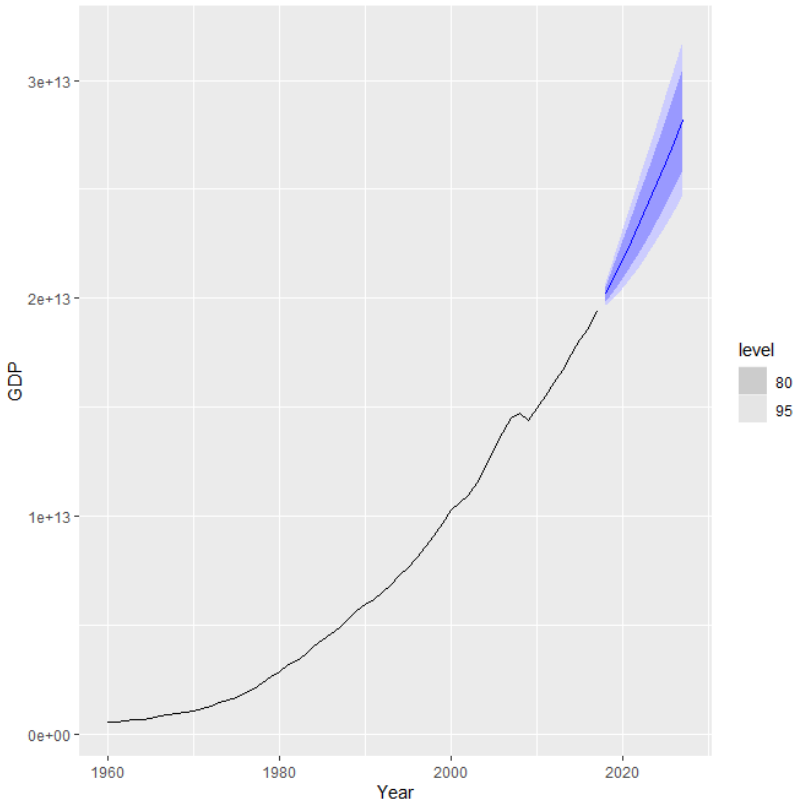
```
#Examine residuals of ARIMA(1,1,0) w/ drift
fit %>%
  gg_tsresiduals()
```



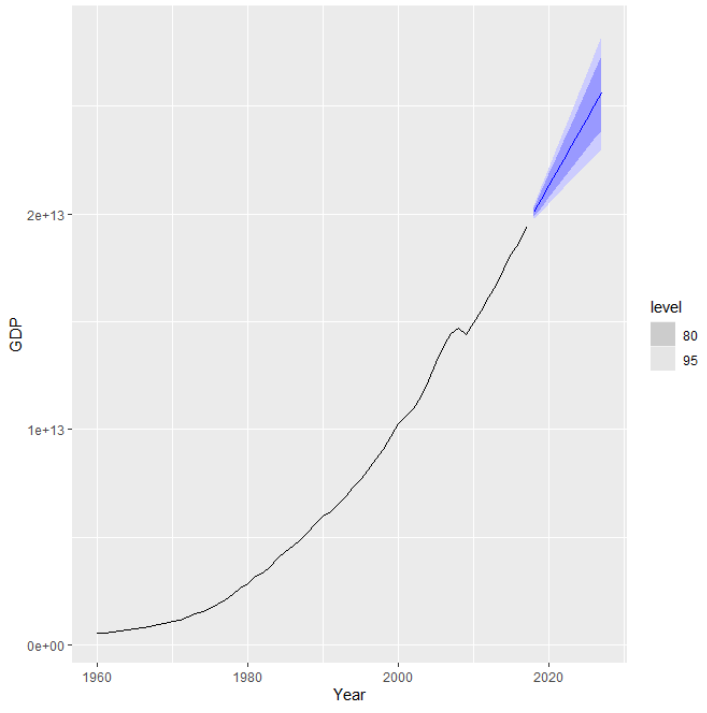
The residuals for the selected model are shown above. The acf plot has no spikes meaning the data follows a white noise process. The distribution of the residuals is not normal however it is not alarming enough to be considered a poor estimation. Some leniencies can be used.

e) Produce a 10-year forecast of the previous best model to the Box-Cox transformed data and include a forecast in a plot. Also, produce a 10-year forecast using ARIMA() on the non-transformed data and include the forecast in another plot. Compare these two forecasts.

```
#10 year forecast for box-cox transformed data
fit %>%
  forecast(h= "10 years") %>%
  autoplot(us_economy)
```




```r
#10 year forecast using ARIMA() on non transformed data
fit3 <- us_economy %>%
  model(
    arima = ARIMA(GDP)
  )
fit3 %>%
  forecast(h= "10 years") %>%
  autoplot(us_economy)
```



We can see from the two outputs there is a slight difference in the prediction.

f) Compare the results with what you would obtain using ETS() on the non-transformed data.

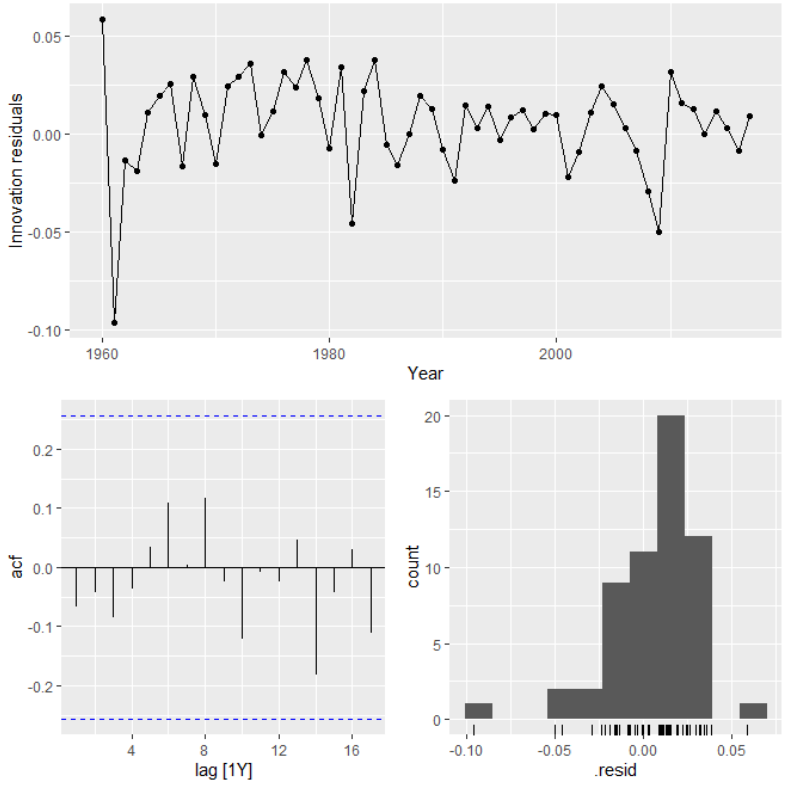
```
#ets model on GDP
us_economy %>%
  model(ets = ETS(GDP))
```

```
# A mable: 1 x 2
# Key:     Country [1]
  Country                ets
  <fct>                <model>
1 United States <ETS(M,A,N)>
```

\

The ETS model recommends using a multiplicative error, additive trend and no seasonality.

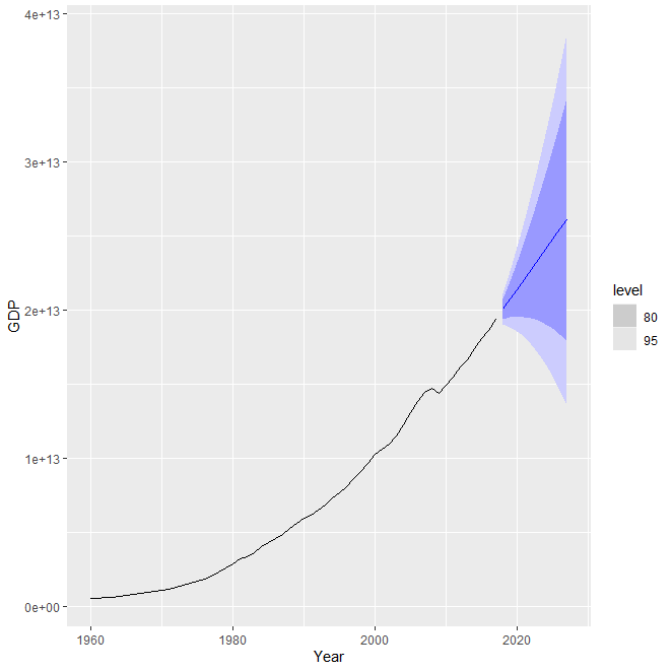
```
#residual diagnostic check for ETS model
us_economy %>%
  model(ets = ETS(GDP)) %>%
  gg_tsresiduals()
```



The autocorrelation has no significant spikes and the residuals do not appear to be normally distributed but are not alarming enough to be considered inappropriate.

```r
#forecast 10 years
us_economy %>%
  model(ets = ETS(GDP)) %>%
  forecast(h= "10 years") %>%
  autoplot(us_economy)
```

-



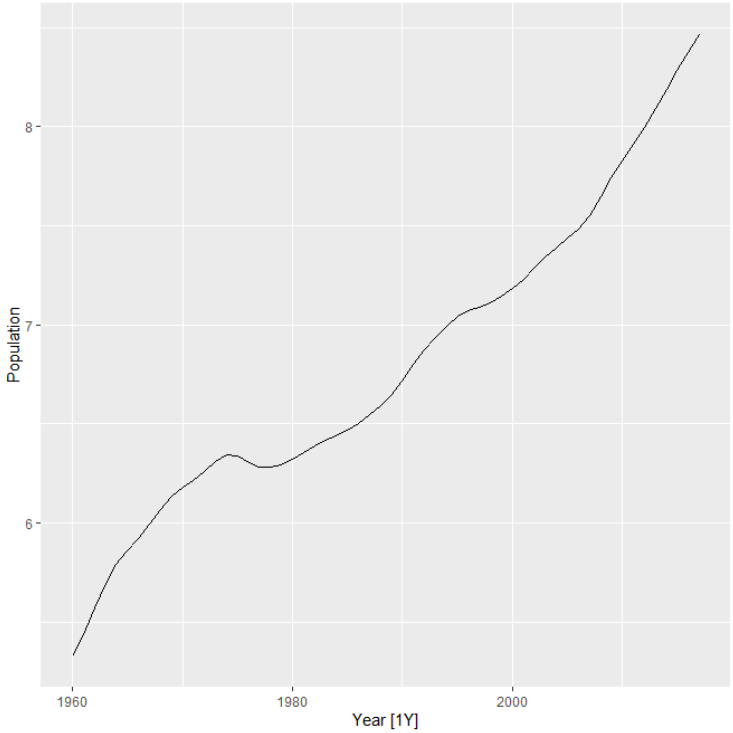
The above plot displays the prediction intervals from the ETS model. The prediction intervals appear to be significantly larger than the ARIMA model. The ARIMA model seems like a better choice.

Question 3

a) Produce a time plot of the Swiss population data

```r
#produce a time plot of the data
swiss_pop <- global_economy %>%
  filter(Country == "Switzerland") %>%
  mutate(Population = Population / 1e6) %>%
  as_tsibble(index = Year)
swiss_pop$Population

swiss_pop %>%
  autoplot(Population)
```



The time series for Swiss population can be seen in the graph.

b) What sort of ARIMA is the model

$$y_t = c + y_{t-1} + \phi_1(y_{t-1} - y_{t-2}) + \phi_2(y_{t-2} - y_{t-3}) + \phi_3(y_{t-3} - y_{t-4}) + \varepsilon_t$$

b)

p = the autoreggresive part

$\phi_1(y_{t-1} - y_{t-2}) + \phi_2(y_{t-2} - y_{t-3}) + \phi(y_{t-3} - y_{t-4})$

. There are 3 differenced autoregressive components

$p = 3$

d = degree of first differencing involved

$d = 1$ represented by the subtraction of the previous data point from the latest

q = order of moving average part

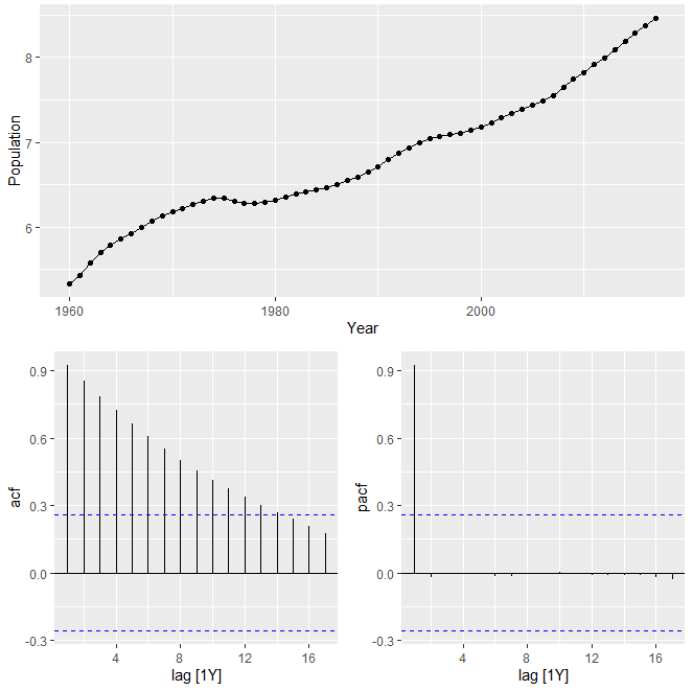
There are no moving average components

$q = 0$

∴ ARIMA(3, 1, 0)

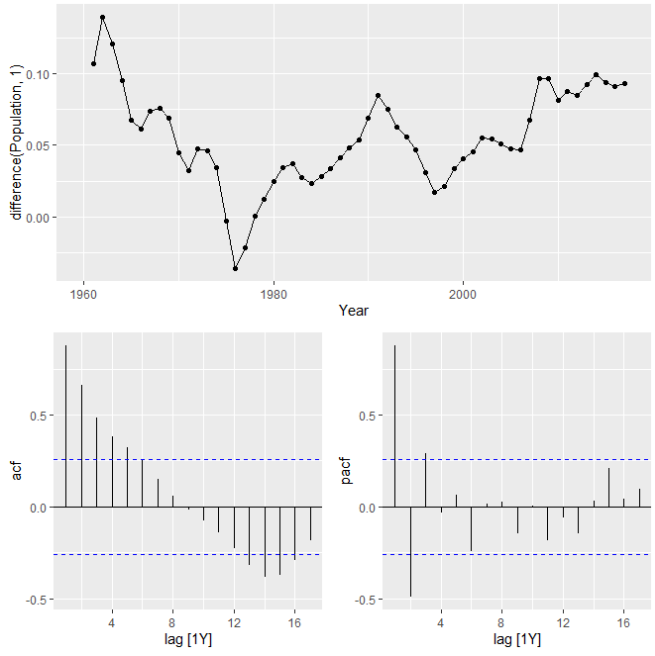
c) Generate the ACF and PACF plots of the original and differenced time series. Based on these plots, explain why the model in the previous question is reasonable.

```
#Generate ACF and PACF plots on original series
swiss_pop %>% gg_tsdisplay(Population, plot_type="partial")
```



Looks like a ARIMA(13,0,0) could be used or a ARIMA(0,0,1) on non-differenced data

```
#generate ACF and PACF plots on differenced series
swiss_pop %>% gg_tsdisplay(difference(Population,1), plot_type="partial")
```



Looks like an ARIMA(4,1,0) could be used as the later spikes around lag 14 in the ACF seem to be slowing down. However, we cannot see past lag 16, it could diminish making it suitable. An ARIMA(0,1,2) or ARIMA(0,1,3) could also be used as the 3 spike is only just significant in PACF.

The model in part b is reasonable as it captures the major spikes within the ACF plot. Conducting multiple tests and ranking based on AICc scores could provide a more optimal model.

d) Calculate forecasts using r for the next 3 years

```
#calculate forecasts for next 3 years
first_year = (0.0053 + 8.47 + 1.64*(8.47-8.37) - 1.17*(8.37-8.28) + 0.45*(8.28-8.19))
second_year = (0.0053 + 8.5745 + 1.64*(8.5745-8.47) - 1.17*(8.47-8.37) + 0.45*(8.37-8.28))
third_year = (0.0053 + 8.67468 + 1.64*(8.67468-8.5745) - 1.17*(8.5745-8.47) + 0.45*(8.47-8.37))
forecasts <- matrix(c(first_year, second_year, third_year))
forecasts
```

```
         [,1]
[1,] 8.57450
[2,] 8.67468
[3,] 8.76701
```

e) Now use the forecast function in r to obtain forecasts from the same model, how are these different?

```
#calculate forecast using function
function_forecast <- swiss_pop %>%
  model(
    arima = ARIMA(Population ~ pdq(3,1,0))
  ) %>%
  forecast(h = "3 years")

view(function_forecast)
```

	Country	.model	Year	Population	.mean
1	Switzerland	arima	2018	*N(8.6, 0.00013)*	8.558549
2	Switzerland	arima	2019	*N(8.6, 0.001)*	8.647550
3	Switzerland	arima	2020	*N(8.7, 0.0033)*	8.731702

These forecasts use the exact population number from the table without rounding. This accounts for the difference between part d and part e estimates. Part e is more accurate.