Q1.Discuss briefly what order you should call these functions in to estimate an Arima model.

R functions: ndiffs, nsdiffs, diff, tsdisplay, autoplot, monthdays, BoxCox, Arima, checkresiduals

If we want to estimate the ARIMA model, I think we should call the above functions in the following order. First, we need to check the trend and seasonality of the time series data by analyzing the graph. First, we use auto plot to view the simple graph, to first view the trend of the data.

If we suspect that the variance of the data is unstable, we can use the boxcox function to transform the data to better fit the ARIMA model.And if our time series contains date information such as months, we can use the monthdays function to handle seasonality.

We then call the nsdiffs function to estimate the number of seasonal differences, then call ndiffs to estimate the number of first-order differences needed, and then use diff to differentiate the function to make the data stationary.

And then use tsdisplay to get more in-depth analysis. These functions generate a picture of the time series data to understand whether the data contains trends and seasonality in PCF and PACF.

What's more,if the ACF is still decaying slowly,we may need to difference one or more time.

According to the results of ACF and PACF,we can form some reasonable hypothesis of ARIMA's lag,and fit them into ARIMA model by using ARIMA function.And choose the result with lowest AICc,which is our optimal lag of ARIMA

Finally, the checkresiduals function is used to check whether the residual of the ARIMA model meets the requirements of stationarity and independence. This helps verify the applicability of the model.If the P-value is greater than 0.05,then pass the residuals test,which means we choose the right lag of ARIMA model.

Q2.

1.ARIMA(0,0,1)

2. ARIMA(0,0,1)
3. ARIMA(0,0,1)(0,0,1)

Q3.

From the table of Johansen procedure's result, it can be deduced that there's no cointegrating relationships among the data.We can tell from the second part,which is the value of teststatistic and critical value.As for the row of r>=1,if the test statistic of this row is greater than the value of 10pct,5pct or 1pct,we can say that the data have at least one cointegrating relationship,but unluckily,the test statistic for row r>=1 is lower than 10pct ,which means we can not say that the data has at least one cointegrating relationship at the significant level of 10%.

Then let's take a look of the second row which means there's no cointegrating relationships among the data,the test statistic of r=0 is greater than 1pct significant value,which means we can conclude that there's no cointegrating relationships among the data at the significant level of 1%.

All in all ,after summarizing the above analysis,we can say that there's no cointegrating relationships among the data.

Q4.Regardless of your answer to B3, outline an appropriate course of action if there are no cointegrating relationships. i.e. what model should we then build, should we render the data stationary, outline how we go about building such a model

If we determine that there is no cointegration relationship between the data, then the ECM and VECM models will not be introduced. For each individual time series data, we can consider using ARIMA model, but the premise of using ARIMA model is to make the data stable. We can observe the ACF diagram of the data or use the ndiff function to determine whether the data needs difference to make the data stable.

If we need to build an ARIMA model to analyze time series data, the easiest way is to use the autoarima formula in R, and then use checkresiduals to check if the residuals fit the normal distribution.

5. Regardless of your answer to B3, outline an appropriate course of action if:

a. There is exactly 1 cointegrated relationship

b. We are unable to reject hypothesis that there is more than 1 cointegrating relationship

a.If there is exactly 1 cointegrated relationship,then it means the data is not stationary,the ARIMA model is no longer available.We can consider using the ECM model.The ECM model is suitable for short term one cointegration relationship data.

b.If we are unable to reject the hypothesis of more than 1 cointegration relationship,which means the data contains one or more cointegration relationship,for this condition it's more suitable to apply VECM model.VECM model is especially useful when the data contains series of time series data and it remains one or more cointegration relationship.

6.

Covariance stationary means that the mean, variance, and correlation coefficients of the data are constant, while the data should show an approximate normal distribution.

From observing the time plot of each 5 graph,the only two that are close to covariance stationary is B8d and B8e,because they has the constant mean which is close to 0,and volatility is stable,and it doesn't has clearly time nor seasonal trend.So compare to other graphs,the B8a is the closest to covariance stationary.