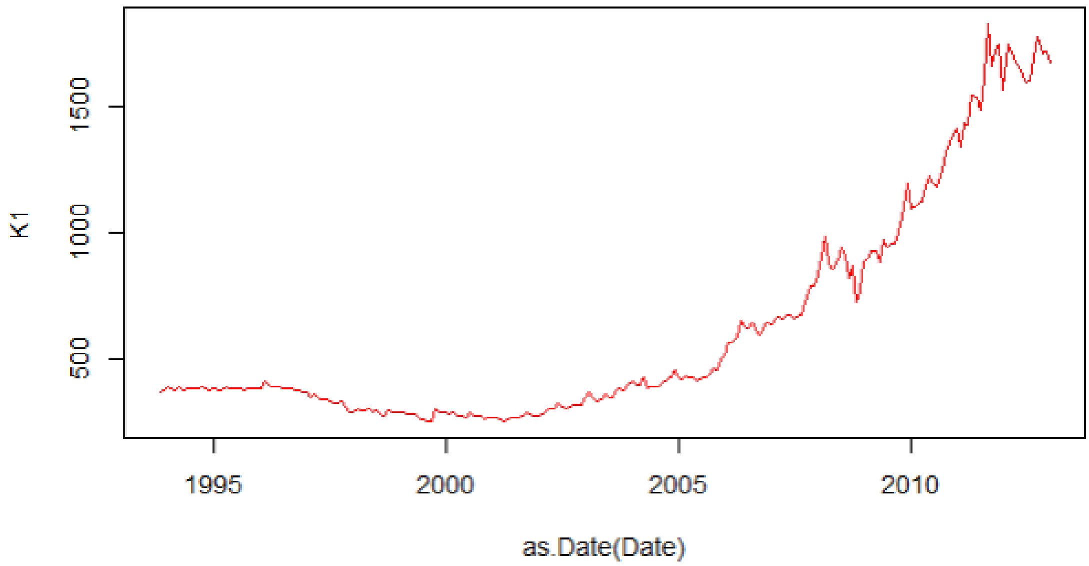
Student Name: Afnan Ur Rehman
Student Id: 21324930

ECOM6003 – PREDICTIVE ANALYTICS IN BUSINESS
ASSESSMENT 3 – PROJECT

1. **<u>The file energy-prices.csv contains time series data on monthly stock prices across four energy companies (K1 to K4) from November 1993 to January 2013.</u>**

(a) Plot the prices for K1 and K2 and calculate the summary statistics. Provide commentary on both the plot and the summary statistics.

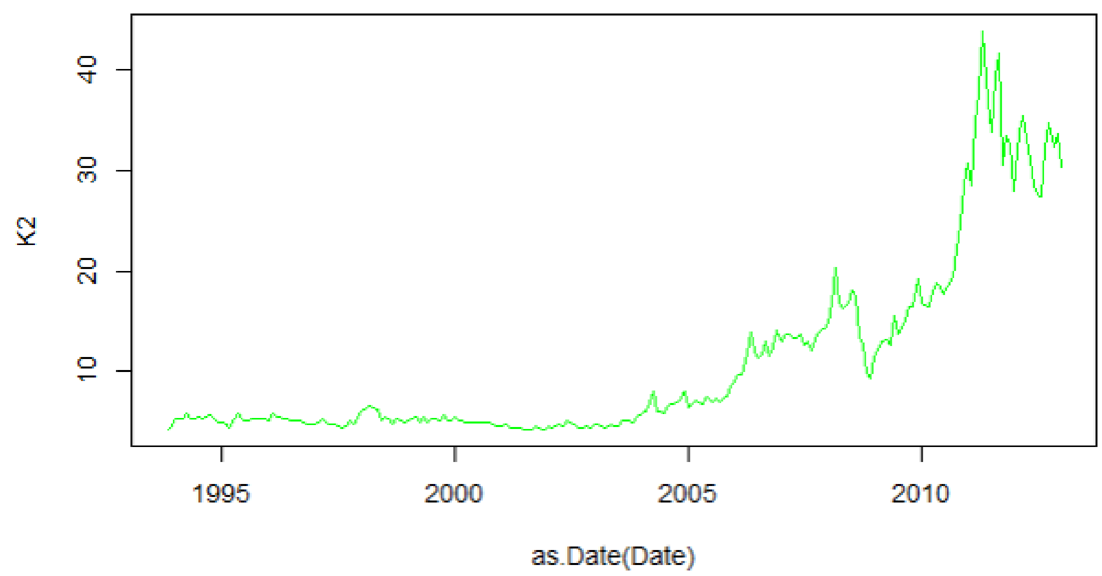
```
     Date                  K1               K2               K3               K4
 Length:231         Min.   : 253.8   Min.   : 4.19   Min.   : 337.3   Min.   : 115.2
 Class :character   1st Qu.: 316.5   1st Qu.: 5.00   1st Qu.: 417.2   1st Qu.: 189.8
 Mode  :character   Median : 392.0   Median : 5.79   Median : 679.5   Median : 305.0
                    Mean   : 619.1   Mean   :10.83   Mean   : 867.1   Mean   : 350.8
                    3rd Qu.: 825.7   3rd Qu.:13.67   3rd Qu.:1245.0   3rd Qu.: 436.8
                    Max.   :1826.1   Max.   :43.80   Max.   :2232.5   Max.   :1045.0
```



Summary Statistics:
K1 has a minimum value of 253.8 and a maximum value of 1826.1 which suggests the extensive growth over the period of 1995 to 2010. K1 has a mean of 619.1 and the median of 392. K2 has prices range from 4.19 to 43.80 and have magnitude much lower in comparison to k1. K2 has a median of 5.79 and a mean of 1083. There is high variation in the values between K1 and K2 and thus we will apply logarithmic transformation on K1 to reduce the variance and normalize the range between K1 and K2.

Plot Analysis:
The plot of K1 and K2 shows the log transformed K1 and raw data of K2. The log transformed K1 is represented using the red line. It can be seen that after 2005, the

stock price indicates a long-term upward trend with high growth. The variation is broad with noticeable troughs and peaks, which shows the periods of high volatility which could be due to some economic event or due to some policy changes. K2 is represented by the green line which shows less volatile and more steady growth patterns as compared to K1. K2 energy prices after 2005 with some minor corrections, indicating a noteworthy upward trend at the same time as K1, which suggests external economic factors, policy implications or similar markets influenced both stocks after 2005.



K1 and K2, both shows the long term growth over the period of 1995 to 2010, where K2 is seemed to face more high volatility and regular bullish corrections in the price levels. This could be a as result of a larger company expansion as compared to the other.

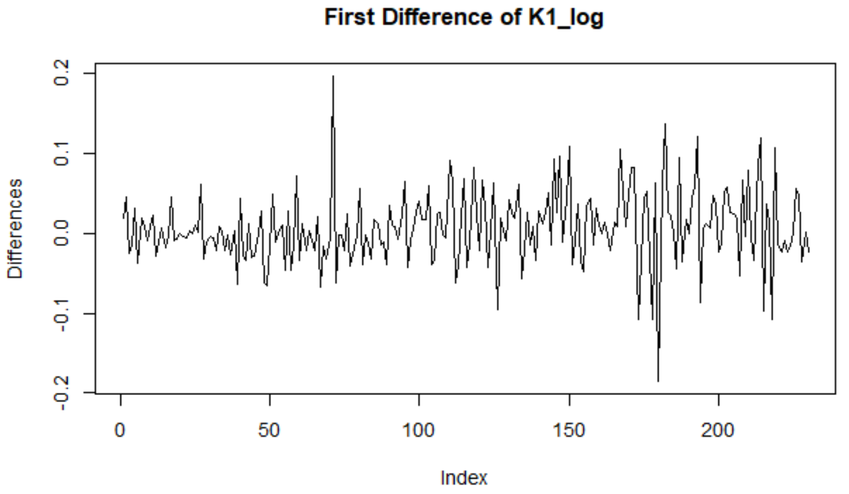
(b) Test for the presence of a unit root for K1 and K2 and provide commentary on the results. Determine the order of integration for each of these series.

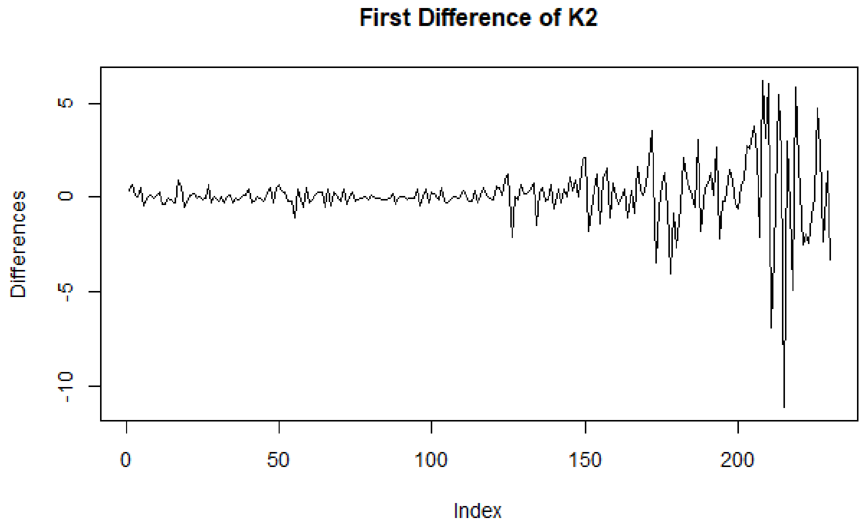
```
        Augmented Dickey-Fuller Test

data:  K1_log
Dickey-Fuller = -1.4954, Lag order = 6, p-value = 0.7879
alternative hypothesis: stationary


        Augmented Dickey-Fuller Test

data:  K2
Dickey-Fuller = -1.9321, Lag order = 6, p-value = 0.6044
alternative hypothesis: stationary
```

Variable	Dickey-Fuller Statistic	Lag Order	p-value	Conclusion
K1_log	-1.4954	6	0.7879	Fail to reject the null hypothesis: non-stationary
K2	-1.9321	6	0.6044	Fail to reject the null hypothesis: non-stationary

ADF test results are summarized in the table above, K1_log and K2, both datasets are non-stationary, as p-values of both variables are above the 0.05 threshold. The ADF test leads to conclusion that unit root null hypothesis cannot be rejected for both variables. K1 and K2 both show non stationary behaviour which means that the statistical properties like variance and mean are not constant with time. The non stationarity could be due to cycles, trends or structural changes in time series. The stock price for non-stationary timeseries implies that the price levels in past do not predict well for the future price levels.

Order of Integration:





The First Difference of K1 shows the fluctuations in the plot around the mean value of zero, without any visible trend in plot, this suggests that after applying the first difference, the series may have become stationary.

First Difference of K2 shows the similar properties as to first difference of logarithm of K1, where the first differenced K2 plot shows the fluctuation of dataset around the mean of zero while it shows the spikes toward the end of the plot indicating the periods of high volatility and stagnant growth. The plot shows no persistent drifts or trends which indicates that after applying first difference, the time series has become potentially stationary. We can further investigate this by applying the ADF test on the first difference time series of log_K1 and K2 as shown in the figure below.

```
Warning in adf.test(diff(ata$K1_log, 1), alternative = "stationary") :
  p-value smaller than printed p-value

	Augmented Dickey-Fuller Test

data:  diff(ata$K1_log, 1)
Dickey-Fuller = -5.8442, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary

Warning in adf.test(diff(ata$K2, 1), alternative = "stationary") :
  p-value smaller than printed p-value

	Augmented Dickey-Fuller Test

data:  diff(ata$K2, 1)
Dickey-Fuller = -5.2172, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
```

The results shows that after applying first difference on both non stationary variables K1_log and K2, the ADF test indicates the p-value for both test to be less that the threshold of 0.05, where both the p-value for K1_log and K2 has come out to be 0.01, which shows that the null hypothesis of unit root can be rejected for both K1_log and K2. This is aligned with what could be seen in the plot that after first differencing both series are stationary.
Thus, K1_log and K2 both are integrated in order 1, which means that for each series, after taking the first difference, resulting data is stationary.

(c) Use the Engle Granger cointegration test to determine if K1 and K2 are cointegrated.

```
Call:
lm(formula = K1_log ~ K2, data = ata)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.86036 -0.18683  0.04078  0.09820  0.51426 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 5.586184   0.022821  244.78   <2e-16 ***
K2          0.059748   0.001611   37.09   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2235 on 229 degrees of freedom
Multiple R-squared:  0.8573,	Adjusted R-squared:  0.8567 
F-statistic:  1376 on 1 and 229 DF,  p-value: < 2.2e-16


	Augmented Dickey-Fuller Test

data:  residuals_eg
Dickey-Fuller = -2.7757, Lag order = 6, p-value = 0.25
alternative hypothesis: stationary
```

The Engle Granger cointegration test has been applied using the model K1_log ~ K2 to test the cointegration of K1 and K2. The Intercept has come out to be 5.586184. The intercept is highly significant and has a p-value < 2e-16. The Slope (K2) has come out to be 0.059748 which is also highly significant variable with the p-value of < 2e-16. Residual Standard Error: 0.2235 on 229 freedom degrees while the adjusted R-square has come out to be 0.8567, signifying that around 85.67% of the variability in K1_log can be explained by K2.

The Augmented Dickey-Fuller Test on Residuals indicated the p-value of 0.25 which is greater than the threshold value of 0.05.

It could be Interpreted from the results that, from regression, the ADF test on residual has a p-value of 0.25, this indicates that we fail to reject the null hypothesis for unit root in the residuals. According to the Engle-Granger two-step method, the failure to reject the null hypothesis suggests the absence of stationary residuals.

The Engle Granger cointegration test suggests that K1_log and K2, both of these variables are not cointegrated, indicating that there is no long-term stable relationship between the logarithm of stock prices of K1 and K2 that exhibits for a longer period of time.

(d) Estimate an error correction model (for K1 and K2) and compare the findings with the results from (c).

Using the equation, we will estimate the error correction model after applying the first difference on K1_log and K2.

$$\Delta y_t = \alpha(y_{t-1} - \beta x_{t-1}) + \gamma \Delta x_t + \epsilon$$

The results of the model are shown in the figure below:

```
Call:
lm(formula = dK1 ~ residuals_eg_diff + dK2)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.15343 -0.02414 -0.00243  0.01831  0.18888 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)       0.004628   0.002425   1.908   0.0576 .  
residuals_eg_diff 0.025522   0.010927   2.336   0.0204 *  
dK2               0.017275   0.001489  11.605   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03669 on 227 degrees of freedom
Multiple R-squared:  0.3739,	Adjusted R-squared:  0.3684 
F-statistic: 67.79 on 2 and 227 DF,  p-value: < 2.2e-16
```

In the ECM model**,** ΔK1 has been taken as a Dependent variable, First difference of K1.

The variable residuals_eg_diff is the error correction term, lagged residuals from from cointegrating regression applied in part c above while the ΔK2 is the first difference series of K2.

The Regression Results suggest the Intercept coefficient of 0.004628 and the p-value has come out to be 0.0576, suggesting a non-significant, small constant term.

The residuals_eg_diff has a coefficient value of 0.025522 and p-value of 0.0204, indicates a positive and statistically significant relationship, suggesting that when K1 is below its equilibrium level comparative to K2, it inclines to do an upward correction towards equilibrium. The ΔK2 has a coefficient value of 0.017275 and p-value of < 2e-16, and is highly significant, representing a strong and positive short-term relationship among changes in K1 and K2.

The Multiple R-squared value has come out to be 0.3739, which indicates that around 37.39% of ΔK1 variation in values can be explained by the model.

The Adjusted R-squared of 0.3684, adjusted for the predictors number, confirms the explanatory power is not due only to the variables number. The F-statistics have come out to be 227 DF and 67.79 on 2, while the p-value has come out to be < 2.2e-16. The small p value suggests that the model is statistically significant.

Comparison:

The error correction term residuals_eg_diff has shown a statistical significance in the model which indicates that there is a long term stable relationship between both variables i.e., K1 and K2, even though the Engle-Granger test indicated that both variables are not cointegrated, which shows that there are some dynamics which are captured by the error correction model, but the simple cointegration test was unable to detect. ΔK2 strong impact indicates that K2 movements have immediate and direct impact on K1 changes. The moderate R-squared model shows that ECM captures a significant portion ΔK1 variability, and suggests that there could be other factors which are not included in model which might also influence K1 changes.

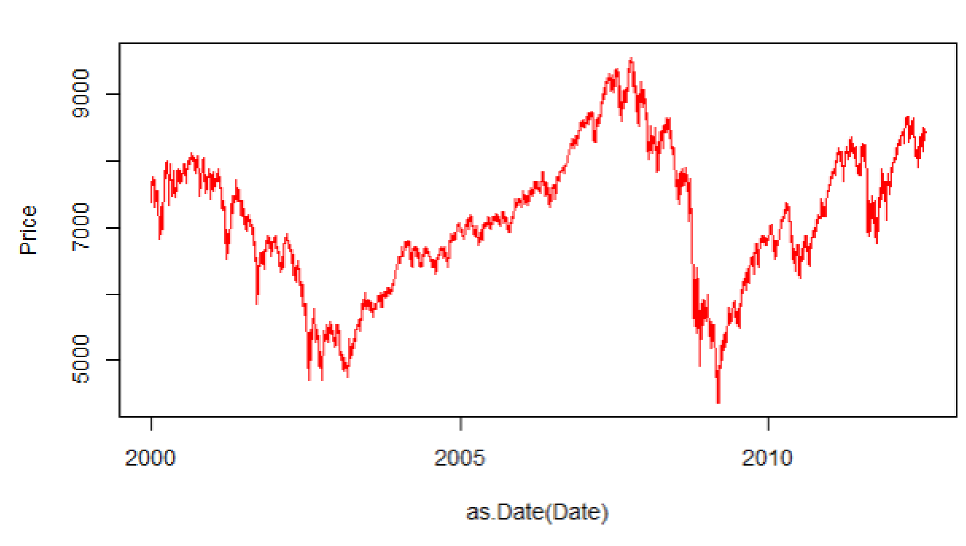
2. **(15 Marks) The file index.xlsx contains the prices of a popular index over time (Pt). Please upload this file and carry out the following tasks:**

(a) Plot the prices and calculate the descriptive statistics. Are the prices stationary? Provide evidence and commentary.

Summary Statistics:

```
      Date                Price     
 Min.   :2000-01-03   Min.   :4363  
 1st Qu.:2003-02-19   1st Qu.:6507  
 Median :2006-04-10   Median :7125  
 Mean   :2006-04-09   Mean   :7127  
 3rd Qu.:2009-05-27   3rd Qu.:7869  
 Max.   :2012-07-16   Max.   :9547  
```

The Descriptive Statistics show the price range of 4363 to 9547 from the period of 2000 to 2012. The price has a median of 7125 and Mean 7127. The mean and Median shows a symmetric distribution near the central values over time.



The plot shows the non-stationary price action with cyclic movement in price where the periods of down trend and upward trend along with high volatility are shown in the plot. The price shows a stagnant growth from 2003 to 2008 while sharp decline or correction in price in 2008 and 2009 with continuation in the upward trend from 2009 onwards.

```
        Augmented Dickey-Fuller Test

data:  Price
Dickey-Fuller = -2.0502, Lag order = 14, p-value = 0.5571
alternative hypothesis: stationary
```

The Stationarity could be Tested using the ADF test. The p-value of 0.5571 is higher than the minimum threshold of 0.05, which suggests that we fail to reject the null hypothesis of unit root indicating the data is nonstationary.

Commentary:

The high p-value of ADF test result shows that series price is non stationary which means that the dataset contains cyclic trends which impact variance and mean with time.

(b) Calculate the returns for the index using

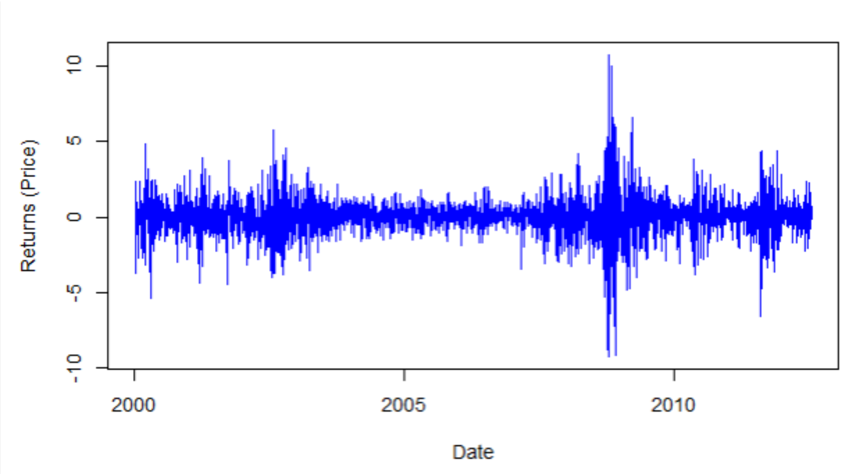
$$\ln\left(\frac{P_t}{P_{t-1}}\right) * 100$$

Plot the returns and calculate the descriptive statistics. Are the returns stationary? Provide evidence and commentary.

```
[1] -3.7084760  0.2253708  1.2054669  2.3361901  0.6573942 -0.8713789
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max. 
-9.236065 -0.549043  0.042127  0.002971  0.583762 10.698764 
```

The return formula is like taking the first difference of the variable, the return of price mostly is stationary. Taking the return of price has brought the Median and mean, close to zero, which is typical for data returns with long period.

The range among the maximum and minimum returns indicates significant volatility, is characteristic of data of financial time series.

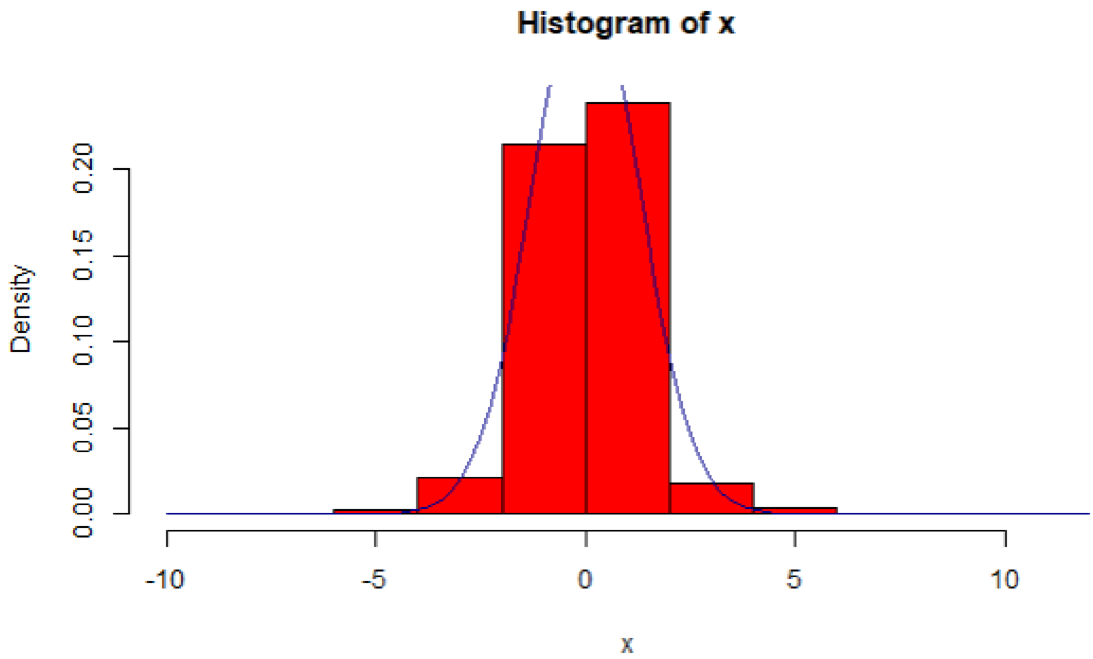


Plot of the Returns:
It shows significant variations in returns with spikes indicating high volatility periods. In financial markets this behavior is typical where returns can be influenced heavily by economic announcements, market events, and another factor.

```
        Augmented Dickey-Fuller Test

data:  ret_Price
Dickey-Fuller = -15.572, Lag order = 14, p-value = 0.01
alternative hypothesis: stationary
```

The augmented Dickey-Fuller (ADF) test result show that returns are stationary, where the p-value is at 0.01, below the common threshold 0.05, allows us to reject the null hypothesis suggesting that the price after taking return has become stationary.

(c) Construct a histogram of returns and assess if these are normally distributed. Provide evidence and commentary.

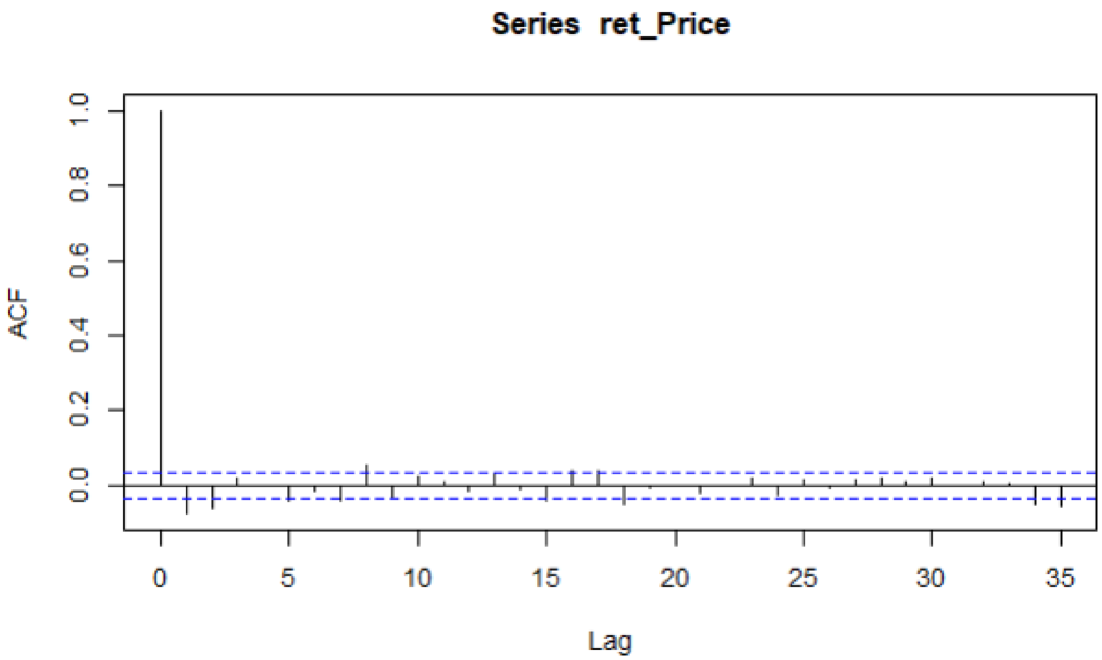


According to histogram, the distribution of price return shows a sharp peak, and the data is not symmetrically distributed around the mean. The histogram reflects extreme values (negative and positive both) more frequently which must be expected in normal distribution. The Blue curve overlaid indicates a normal distribution. Although, the distribution seems normally distributed, we can test it using the Shapiro Wilk Normality Test.

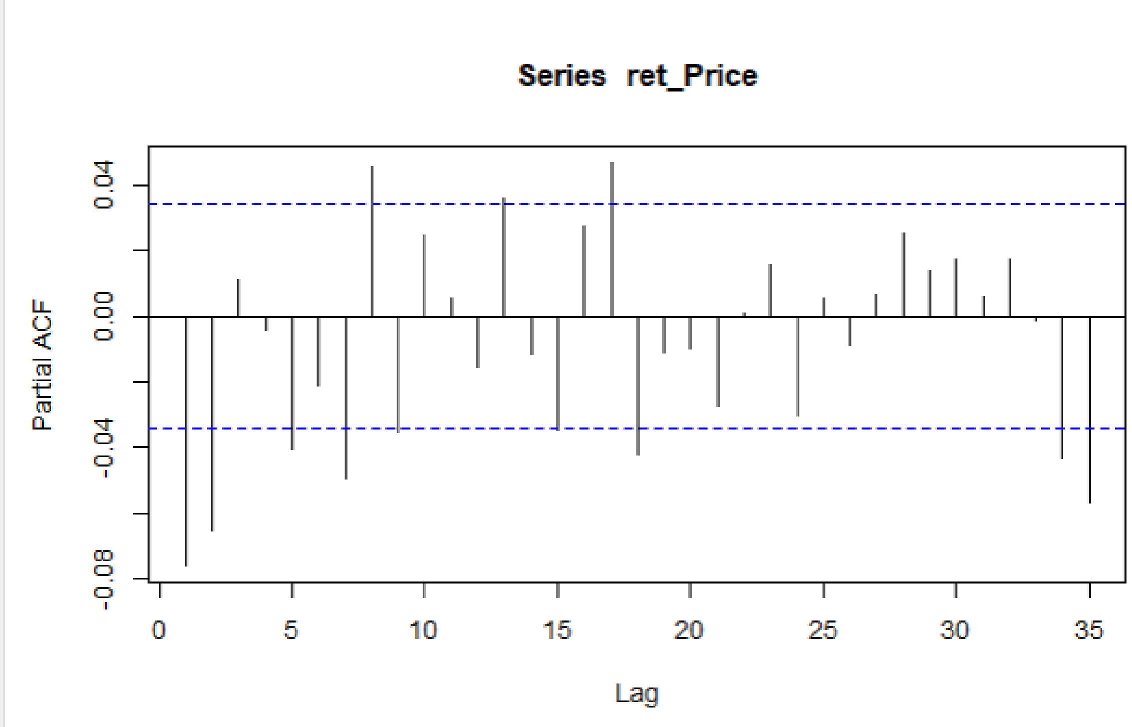
```
        Shapiro-Wilk normality test

data:  ret_Price
W = 0.90683, p-value < 2.2e-16
```

Shapiro-Wilk Normality Test indicates the p-value of < 2.2e-16. The small p-value suggests that the null hypothesis of normality is strongly rejected by Shapiro-Wilk test. This low p-value shows that there are no normally distributed returns.

(d) Plot the autocorrelation function (ACF) of the returns. Briefly describe this plot.



ACF plot for series returns indicates that at lag 0 there are significant spikes. The autocorrelations for subsequent lags quickly fall inside the confidence bounds which at any lag shows no significant autocorrelation. The absence of significant autocorrelation indicates that based on past values returns, the returns do not show linear predictability.



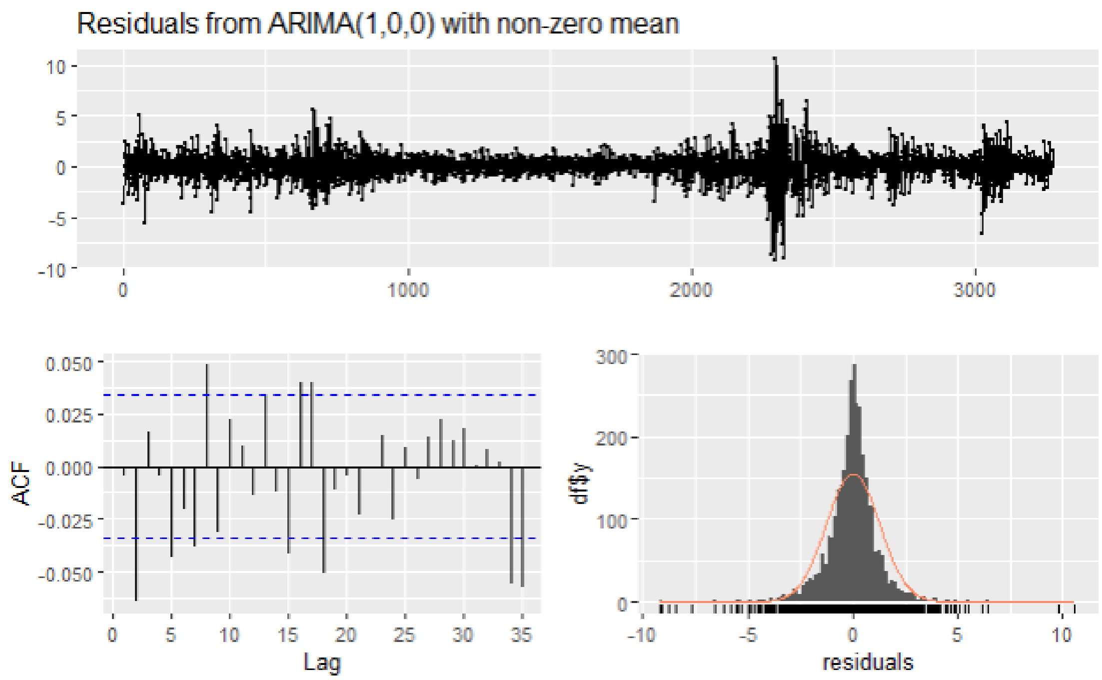
The returns Partial ACF plot indicates that all partial autocorrelations are inside the confidence bounds, showing no significant partial autocorrelation, at any lag. from the ACF plot. In the ACF plot, the lack of autocorrelation aligns with Augmented Dickey-Fuller test results previously conducted, supports the conclusion on stationary returns.

(e) Estimate an autoregressive (AR) models with lags 1, 2 and 3. Estimate a moving average (MA) models with lags 1, 2 and 3. Provide brief commentary on all models. Which of these models would you recommend and why?

Lag 1:

```
Series: price_ts
ARIMA(1,0,0) with non-zero mean

Coefficients:
          ar1    mean
      -0.0762  0.0031
s.e.   0.0175  0.0206

sigma^2 = 1.606:  log likelihood = -5413.32
AIC=10832.64   AICc=10832.65   BIC=10850.92

Training set error measures:
                        ME     RMSE       MAE MPE MAPE      MASE         ACF1
Training set -8.754185e-05 1.266821 0.8418888 NaN  Inf 0.6737676 -0.004809526
```

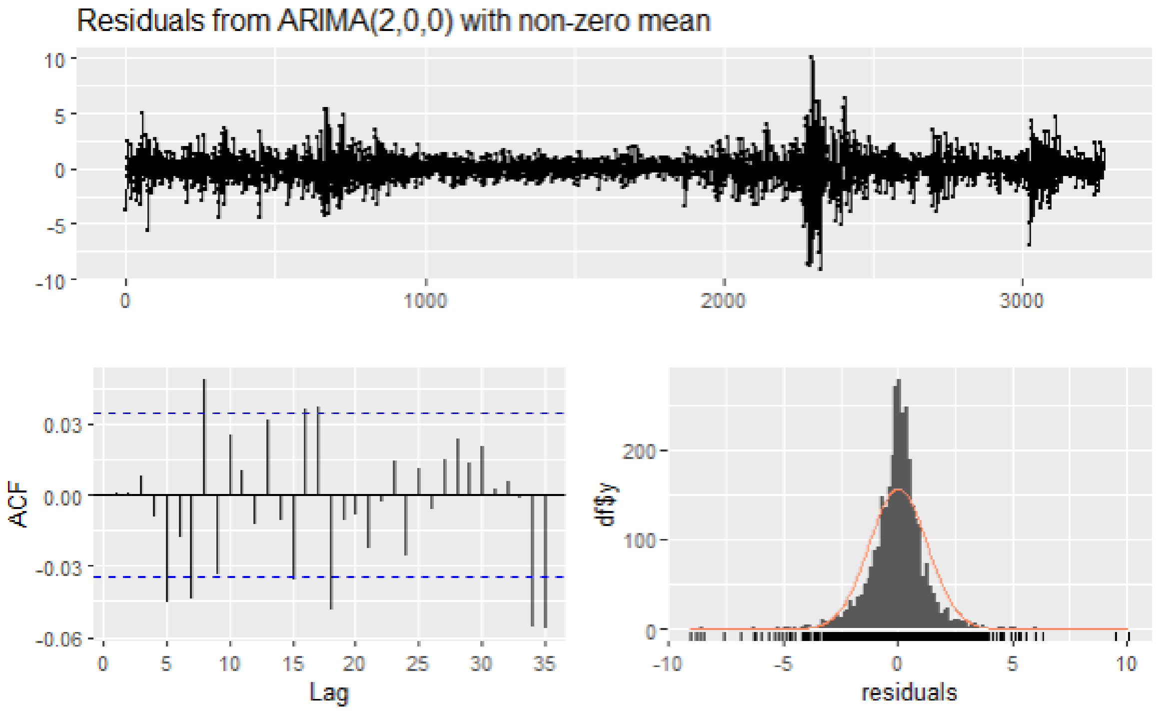
AR model with lag 1 has a coefficient of -0.0762 with the mean of 0.0031. The AIC came out to be 10832.64 and BIC has come out to be 10850.92. The model fit could be tested using the residual test.



Lag 2:

```
Series: price_ts
ARIMA(2,0,0) with non-zero mean

Coefficients:
          ar1      ar2    mean
      -0.0812  -0.0656  0.0031
s.e.   0.0175   0.0175  0.0193

sigma^2 = 1.599:  log likelihood = -5406.29
AIC=10820.58   AICc=10820.59   BIC=10844.95

Training set error measures:
                        ME     RMSE       MAE MPE MAPE      MASE         ACF1
Training set -0.0001487109 1.264098 0.8429978 NaN  Inf 0.6746552 0.0009003512
```

The AR Model with lag 2 has a coefficient of -0.0812 for AR1 and -0.0656 for AR2. The Mean has come out to be 0.0031. The AIC value of 10820.58 and BIC value of 10844.95 shows a slight improvement in model fit over the AR model with lag 1.



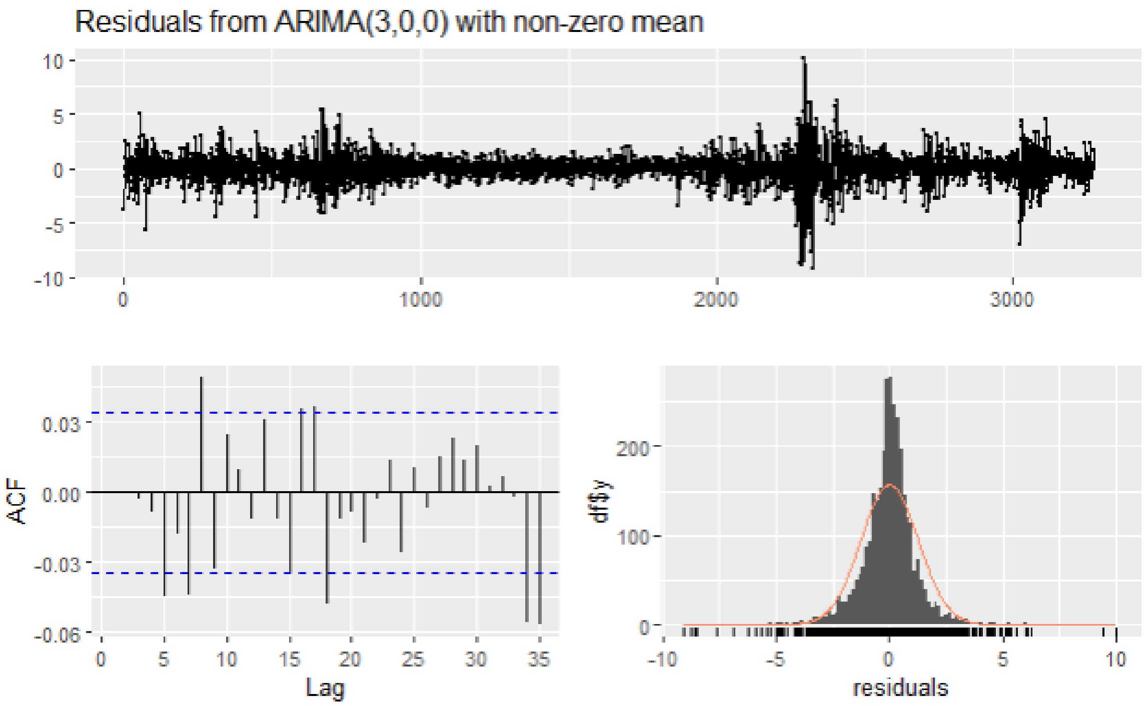
Lag 3:

```
Series: price_ts
ARIMA(3,0,0) with non-zero mean

Coefficients:
          ar1      ar2     ar3    mean
      -0.0805  -0.0647  0.0111  0.0031
s.e.   0.0175   0.0175  0.0175  0.0195

sigma^2 = 1.6:  log likelihood = -5406.09
AIC=10822.18   AICc=10822.2   BIC=10852.64

Training set error measures:
                        ME     RMSE      MAE MPE MAPE      MASE         ACF1
Training set -0.0001364372 1.264021 0.842832 NaN  Inf 0.6745224 0.0002269189
```

The AR model with lag 3 showing the coefficient of AR1 of -0.0805, AR2 of -0.0647 and AR3 of 0.0111 while the mean has come out to be 0.0031. The Standard Errors across all lags is Consistent. The AIC has come out to be 10822.18 and BIC has come out to be 10852.64, which is slightly higher as compared to the AR model with lag 2, suggesting that the third lag addition does not sufficiently improve model to justify the additional complexities.

AR (1) and AR (2) Residuals both show residuals around zero, and are broadly centered and randomly appeared to be scattered around the central value which indicated the good model fit with no obvious trends or patterns that are left unexplained.
The residual ACF for both models indicates no autocorrelations significance, which shows that model has captured a series of primary dynamics. The histogram residuals

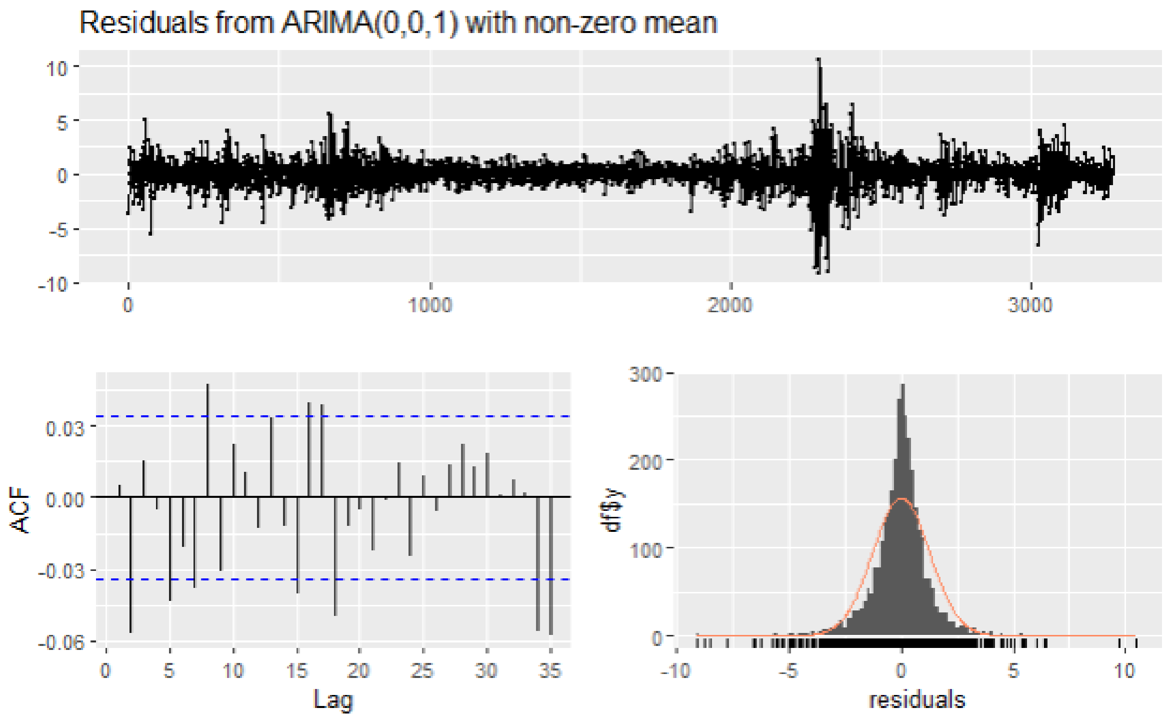
for AR with lag 1, 2 and 3 with a density plots show that there is approximately normal distribution of residuals, which is ideal for AR models.



Lag 1:

```
Series: price_ts
ARIMA(0,0,1) with non-zero mean

Coefficients:
          ma1    mean
      -0.0866  0.0030
s.e.   0.0185  0.0202

sigma^2 = 1.605:  log likelihood = -5412
AIC=10830   AICc=10830   BIC=10848.27

Training set error measures:
                        ME     RMSE       MAE MPE MAPE      MASE        ACF1
Training set -1.564475e-05 1.266308 0.8421425 NaN  Inf 0.6739707 0.005070276
```

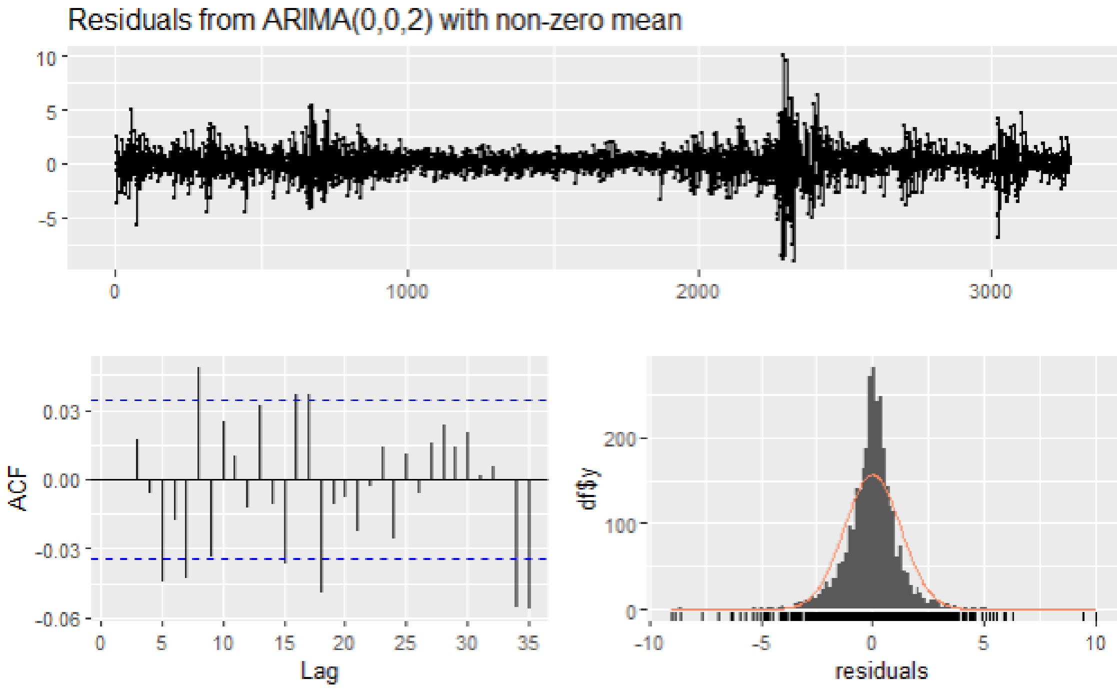
MA model with lag 1 has a coefficient of -0.0866 and a Mean of 0.0030. The AIC has come out to be 10830 and BIC is 10848.27.



Lag 2:

```
Series: price_ts
ARIMA(0,0,2) with non-zero mean

Coefficients:
          ma1      ma2   mean
      -0.0798  -0.0586  0.003
s.e.   0.0175   0.0178  0.019

sigma^2 = 1.6:  log likelihood = -5406.61
AIC=10821.22   AICc=10821.23   BIC=10845.59

Training set error measures:
                        ME     RMSE       MAE MPE MAPE      MASE          ACF1
Training set -4.978361e-05 1.264221 0.8430808 NaN  Inf 0.6747216 -0.0006668716
```

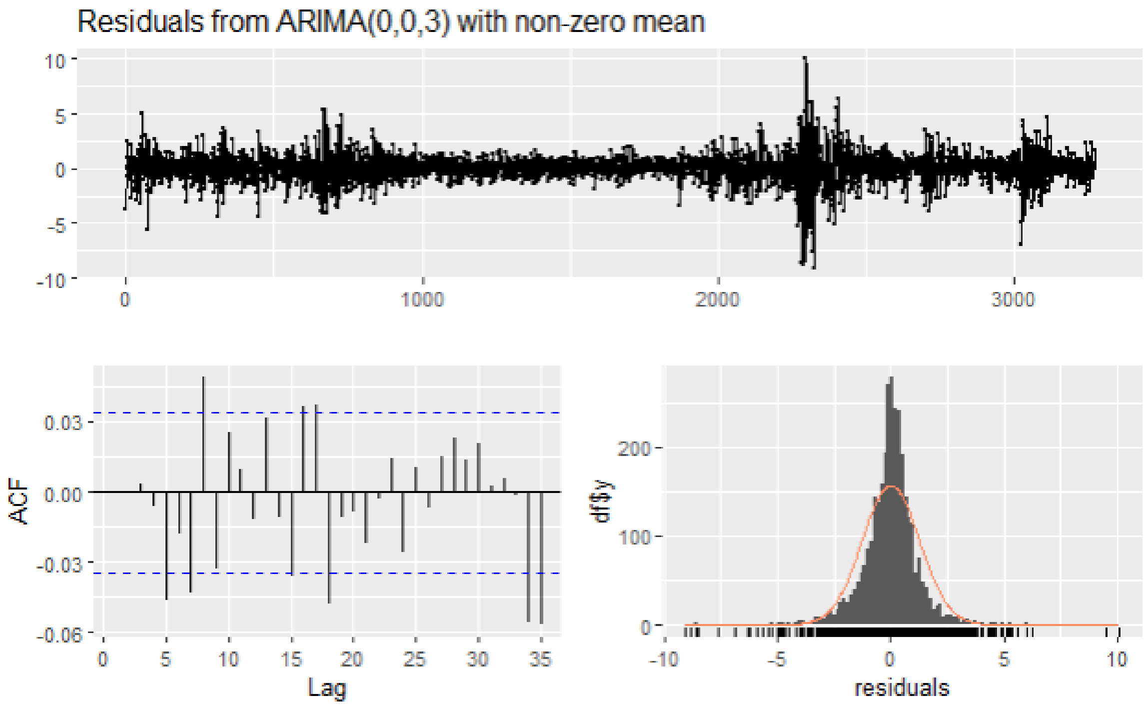
MA model with lag 2 is shown in the figure above, the coefficient MA1 is -0.0798, MA2 is -0.0586 with the mean of 0.0030. The AIC has come out to be 10821.22 while the BIC is 10845.59



Lag 3:

```
Series: price_ts
ARIMA(0,0,3) with non-zero mean

Coefficients:
          ma1      ma2     ma3    mean
      -0.0803  -0.0582  0.0144  0.0030
s.e.   0.0175   0.0178  0.0178  0.0194

sigma^2 = 1.6:  log likelihood = -5406.28
AIC=10822.56   AICc=10822.58   BIC=10853.03

Training set error measures:
                        ME     RMSE       MAE MPE MAPE      MASE         ACF1
Training set -4.118304e-05 1.264095 0.8429121 NaN  Inf 0.6745865 1.695918e-05
```

MA model with lag 3 has shown the coefficients, where MA1 is -0.0803, MA2 is -0.0582 and MA3 is 0.0144 while the mean has come out to be 0.0030. The model with lag 3 has AIC of 10822.56 while the BIC has come out to be 10853.03.



BIC and AIC values from MA (1) to MA (2) decrease, indicating that by doing so, the model fit has improved by second lag. For MA (2), AIC for MA (3) is slightly higher shows that third lag addition does not offer significant enough improvement in fit to defend the complexity of model.
Residuals from MA Models across all models, the residuals seem to fluctuate around zero, deprived of any apparent pattern. It shows good indication that major dynamics of data are captured by models. The ACF of residuals plot shows no significant autocorrelations, implying that there is no leftover residuals pattern that the models failed to capture

AR Models:

- AR (1) model has BIC of 10850.92 and AIC of 10832.64, having one significant negative coefficient, which at lag 1 suggests some negative autocorrelation.
- AR (2) shows improvement, with BIC of 10844.95 and AIC of 10820.58. It contains two significant negative coefficients, showing a complex lag structure in the data captures more dynamics.
- AR (3) BIC and AIC slightly increase as compared to AR (2).

MA Models:

- MA (1) with BIC of 10848.27 and AIC of 10830, Features a single negative coefficient significantly. from the previous period, this model accounts for shock effects
- MA (2) with lower AIC and BIC values, 10821.22 and 10845.59, Improves upon MA(1), suggests better model fit, involves two significant negative coefficients showing the impact of shocks extends with two periods.
- MA (3) as compared to MA(2), show Slightly higher AIC and BIC.

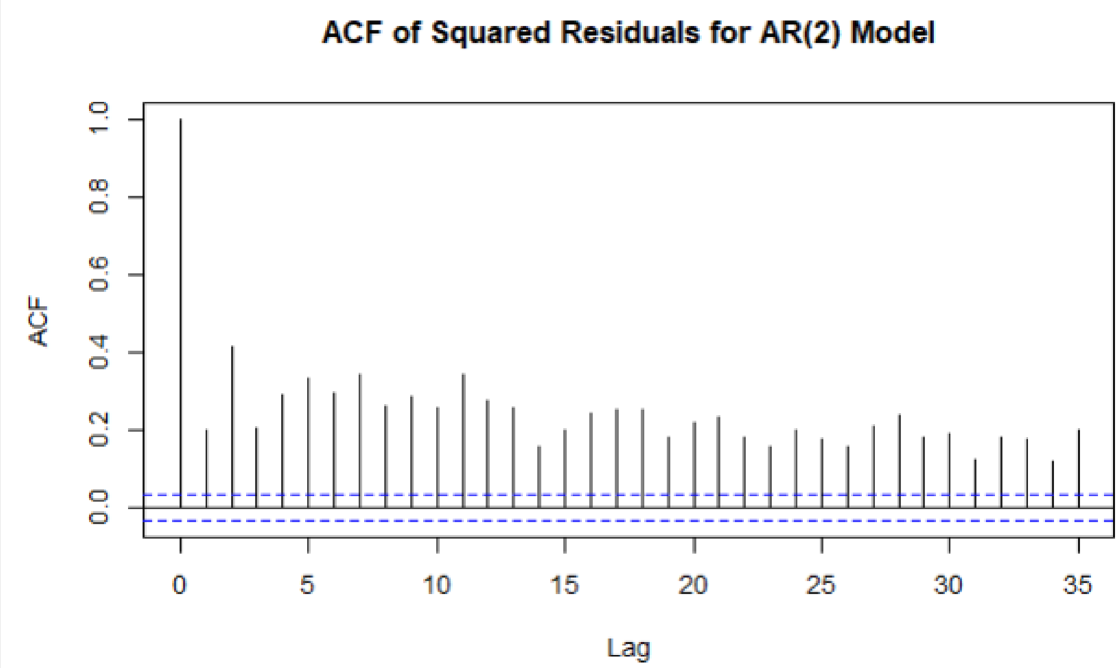
Residual Analysis:

Generally, the residuals show a random pattern around zero, across both types of models, with no significant autocorrelation in the ACF plots, showing effective data's autocorrelation capture structure by both models.

Model Recommendation:

MA (2) model is a better choice, based on the AIC/BIC values stands out as the most appropriate, showing that it offers a good balance between model fit and complexity and captures the immediate impact of shocks and their leftover, often critical in data of financial time series. Although AR (2) has low BIC and AIC values, the model MA (2) ability to model directly the process of noise may provide a better framework for forecasting after dealing with potentially unstable financial data. The residual model of MA is slightly whiter indicates randomness than that of AR model, indicating that series of dynamics for model MA (2) must be captured more effectively.

(f) Plot the ACF for the squared residuals of the AR (2) model. Test for the presence of GARCH effects by regressing the squared residuals over different lag lengths i.e. Regress $\hat{e}_t^2$ against $\hat{e}_{t-1}^2$, $\hat{e}_{t-2}^2$, $\hat{e}_{t-3}^2$, $\hat{e}_{t-4}^2$ and $\hat{e}_{t-5}^2$ (5-day lag) Hint: Are the coefficients significant, what is the significance of the overall regression?



ACF of Squared Residuals:

For squared residuals, the ACF plot at multiple lags shows significant autocorrelations. volatility clustering is followed in this pattern, where small changes follow small changes, and large changes tend to be followed by large changes. The presence of significant autocorrelation shows that in the data, AR (2) model does not completely capture the conditional heteroscedasticity.

```
Call:
lm(formula = lag0 ~ lag1 + lag2 + lag3 + lag4 + lag5, data = lagged_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-33.646  -0.913  -0.539   0.048  79.276 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.473727   0.085228   5.558 2.94e-08 ***
lag1        0.048760   0.017036   2.862  0.00423 ** 
lag2        0.316012   0.016966  18.626  < 2e-16 ***
lag3        0.004257   0.017843   0.239  0.81146    
lag4        0.100247   0.016964   5.909 3.79e-09 ***
lag5        0.232331   0.017015  13.655  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.376 on 3259 degrees of freedom
Multiple R-squared:  0.249,	Adjusted R-squared:  0.2478 
F-statistic: 216.1 on 5 and 3259 DF,  p-value: < 2.2e-16
```

At multiple lags, especially lag2, lag4, lag5, the regression outputs display significant coefficients, showing that current volatility significantly predicted by past volatility. Adjusted R-squared value shows that in squared residuals nearly 24.78% of the variability is described by their own previous values, which is substantial for unpredictability modeling. The p-value has come out to be < 2.2e-16, The highly significant F-statistic, confirms the efficacy of model in analyzing the relationship between the lags of squared residuals.

(g) Estimate a GARCH (1,1) model the using the Normal distribution. Perform diagnostic checks (ACF of residuals/ squared residuals) and provide commentary. Estimate the GARCH (1,1) model using the *t*-distribution. Compare the results across both models.

```
*---------------------------------*
*          GARCH Model Fit        *
*---------------------------------*

Conditional Variance Dynamics
-----------------------------------
GARCH Model      : sGARCH(1,1)
Mean Model       : ARFIMA(0,0,0)
Distribution     : norm

Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.039066    0.014551   2.6848 0.007257
omega   0.014913    0.002930   5.0898 0.000000
alpha1  0.090337    0.008734  10.3427 0.000000
beta1   0.898994    0.009156  98.1829 0.000000

Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.039066    0.013108   2.9804 0.002879
omega   0.014913    0.004634   3.2185 0.001289
alpha1  0.090337    0.012690   7.1189 0.000000
beta1   0.898994    0.012123  74.1562 0.000000

LogLikelihood : -4624.622
```

```
Information Criteria
------------------------------------

Akaike       2.8310
Bayes        2.8384
Shibata      2.8310
Hannan-Quinn 2.8336

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic  p-value
Lag[1]                      6.096 0.013552
Lag[2*(p+q)+(p+q)-1][2]     7.580 0.008359
Lag[4*(p+q)+(p+q)-1][5]     9.131 0.015337
d.o.f=0
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic  p-value
Lag[1]                      6.055 0.013863
Lag[2*(p+q)+(p+q)-1][5]    10.437 0.007108
Lag[4*(p+q)+(p+q)-1][9]    12.479 0.014099
d.o.f=2
```

```
Weighted ARCH LM Tests
------------------------------------
             Statistic Shape Scale P-Value
ARCH Lag[3]    0.2333 0.500 2.000  0.6291
ARCH Lag[5]    0.9222 1.440 1.667  0.7562
ARCH Lag[7]    2.5614 2.315 1.543  0.5999

Nyblom stability test
------------------------------------
Joint Statistic:  0.8717
Individual Statistics:
mu     0.2465
omega  0.1114
alpha1 0.1998
beta1  0.1329

Asymptotic Critical Values (10% 5% 1%)
Joint Statistic:         1.07 1.24 1.6
Individual Statistic:    0.35 0.47 0.75

Sign Bias Test
------------------------------------
```

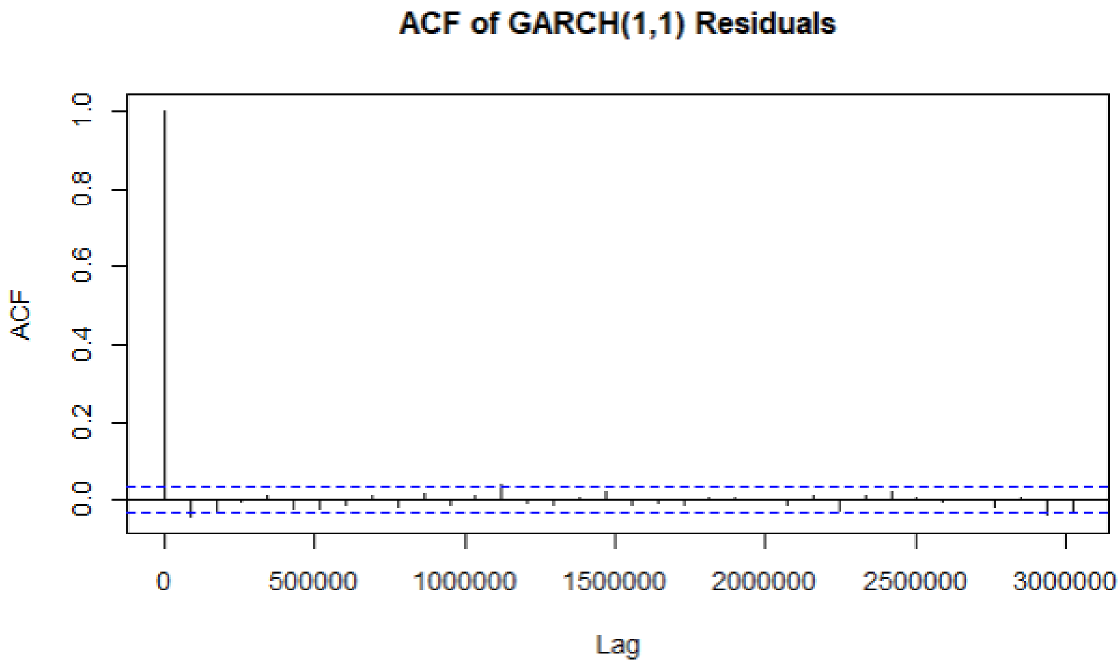
```
Sign Bias Test
------------------------------------


Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20     108.5    1.477e-14
2    30     134.4    1.633e-15
3    40     168.5    4.912e-18
4    50     169.6    3.363e-15


Elapsed time : 0.719069
```

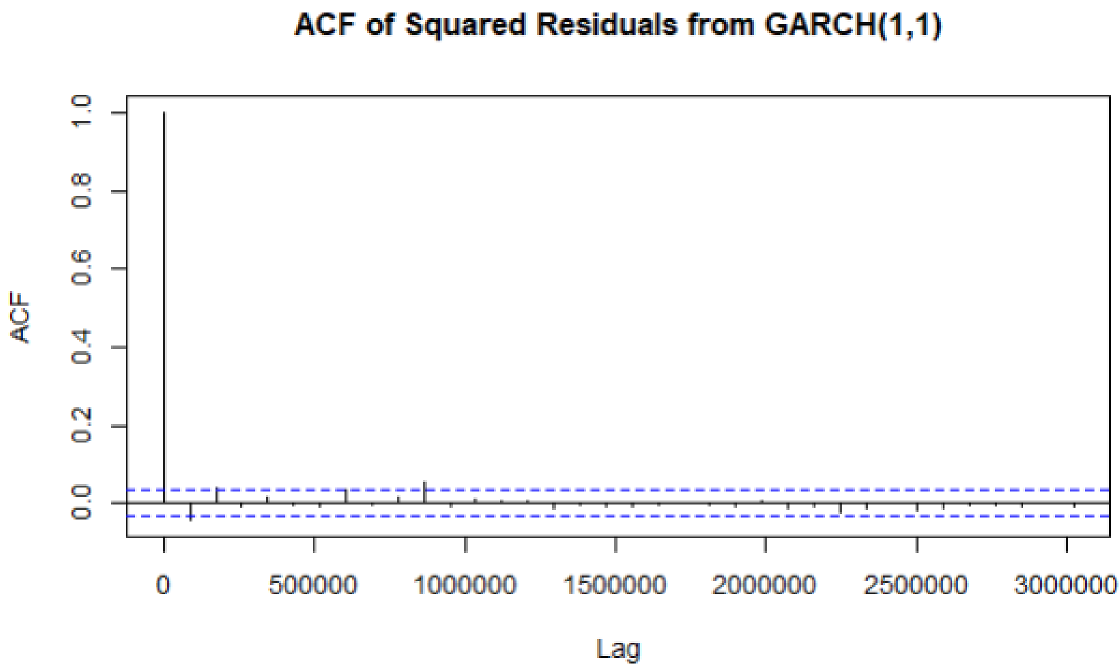
	t-value <dbl>	prob <dbl>	sig <chr>
Sign Bias	2.017794	0.0436944217	**
Negative Sign Bias	1.145273	0.2521801675	
Positive Sign Bias	2.315059	0.0206714015	**
Joint Effect	18.415737	0.0003610066	***

For past conditional variances and for past squared residuals each, GARCH (1,1) shows a standard GARCH model with one lag for normally distributed standardized residuals. The Mean Model indicates ARFIMA (0,0,0), no moving or autoregressive average components which are modeled directly in mean equation. Mean (mu) indicates a small, slightly positive but consistent upwards series drift. Omega (ω) suggests that when past values are zero, the baseline changes. The positive value significantly suggests a non-trivial baseline instability. Alpha (α1) value of 0.0903 shows that around 9% of the volatility is due to the last shock period. Beta (β1) has a high value of 0.8990 which indicates that over time, volatility shocks are highly persistent, with past variance nearly 90%, persisting into the future. All of the parameters are highly significant parameters with p-value of < 0.01, which indicates and confirms the reliability model estimates. The log likelihood value of -

4624.622 indicates a strong fit. The Lower values of BIC and AIC are better. Given the data complexity, the values here suggest a comparatively efficient model.

ACF of residuals/ squared residuals:





```
*---------------------------------*
*          GARCH Model Fit        *
*---------------------------------*

Conditional Variance Dynamics
-----------------------------------
GARCH Model     : sGARCH(1,1)
Mean Model      : ARFIMA(0,0,0)
Distribution    : std

Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.053217    0.013588   3.9163 0.000090
omega   0.009877    0.003018   3.2725 0.001066
alpha1  0.092128    0.010725   8.5898 0.000000
beta1   0.904712    0.010167  88.9839 0.000000
shape   6.714222    0.807712   8.3126 0.000000

Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.053217    0.011838   4.4952 0.000007
omega   0.009877    0.002944   3.3546 0.000795
alpha1  0.092128    0.012182   7.5624 0.000000
beta1   0.904712    0.011337  79.8025 0.000000
shape   6.714222    0.856105   7.8428 0.000000

LogLikelihood : -4563.564

Information Criteria
------------------------------------

Akaike       2.7942
Bayes        2.8035
Shibata      2.7942
Hannan-Quinn 2.7976

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic  p-value
Lag[1]                      6.043 0.013964
Lag[2*(p+q)+(p+q)-1][2]     7.594 0.008288
Lag[4*(p+q)+(p+q)-1][5]     9.174 0.014959
d.o.f=0
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                      5.535 0.01864
Lag[2*(p+q)+(p+q)-1][5]     8.947 0.01707
Lag[4*(p+q)+(p+q)-1][9]    10.643 0.03642
d.o.f=2

Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[3]    0.3413 0.500 2.000  0.5591
ARCH Lag[5]    0.9163 1.440 1.667  0.7579
ARCH Lag[7]    2.2493 2.315 1.543  0.6646

Nyblom stability test
------------------------------------
Joint Statistic:  1.5101
Individual Statistics:
mu     0.3989
omega  0.1591
alpha1 0.1426
beta1  0.1516
shape  0.5106
```

```
Asymptotic Critical Values (10% 5% 1%)
Joint Statistic:         1.28 1.47 1.88
Individual Statistic:    0.35 0.47 0.75

Sign Bias Test
------------------------------------


Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20     62.67    1.451e-06
2    30     74.59    6.970e-06
3    40     88.61    9.995e-06
4    50    109.97    1.408e-06


Elapsed time : 0.9040341
```

	t-value <dbl>	prob <dbl>	sig <chr>
Sign Bias	2.118465	0.0342110173	**
Negative Sign Bias	1.570077	0.1164942133	
Positive Sign Bias	2.409398	0.0160338524	**
Joint Effect	19.573573	0.0002080278	***

4 rows

ACF of GARCH (1,1) residuals suggest that the t-distribution and normal distribution models both in the residuals across numerous lags show no significant autocorrelation, which is ideal and indicates that the models are sufficiently capturing the linear dependencies across all the series. However, the ACF of Squared Residuals in the squared residuals suggests that there is no significant autocorrelation, which shows that in modeling volatility clustering, the models are effective, observed in data of financial time series, and leaving no unexplained volatility in pattern.

The GARCH Model with t-Distribution indicates that the t-Distribution shift shows improvement in the peaks of the distribution and fitting the tails, which is very common in financial data due to extreme macro and micro foundational events or outliers, which are hard to predict and model with normal distribution. The Parameter Estimates in data distribution, the shape parameter inclusion significantly improves the model, allowing it to be adjusted for heavier tails. As for the shape parameter, the significant t-value confirms the importance in the model. The Log Likelihood for the t-distribution model is higher such as -4563.564, in comparison to normal distribution model, which indicates better overall fit. In terms of the Ljung-Box test, both models perform likewise showing no serial correlation in squared residuals and standardized residuals, hence validating model fit. In the ARCH LM Tests, there are no ARCH effects confirmed by non-significant results which confirm that conditional heteroscedasticity has been modeled successfully. The GARCH (1,1) model with t-distribution is recommended over the normal distribution model due to its better handling of extreme values and outliers, typical in financial markets, especially with the t-distribution, GARCH (1,1) model, sufficiently captures the distributional properties and volatility of data.

3. **<u>A government agency is asking your help to establish a model to predict State Final Demand (SFD) for every state in Australia. The proposed model must be able to produce quarterly forecast of SFD for all states in Australia and the forecast horizon should be at least 4 steps ahead (forecast for next year).</u>** **Please answer each of the following questions. Please label your solution to each part.**

   (a) Provide a concise description of SFD and its components. What does this measure and why is it important for managing state economy?

The State Final Demand (SFD) is a measure of demand in state economy which indicates the aggregate level of the total value of services and goods consumed within a territory or a state, excluding the international trade i.e., imports and exports, and involves Business Investment, Dwelling Investments, private consumption, government investments and Governments consumption(*Known components of expenditure - GSP(E)*, 2020). State Final Demand defined as the final use of services and goods within a specified period by businesses, households and government.

SFD =Government final consumption expenditure (GFCE) + Household final consumption expenditure (HFCE) + Gross fixed capital formation (GFCF)

It is more inclined towards the demand side of the economy which reflects the investment and consumption activity. Some of the main components of SFD include:

1. **Household Consumption**: One of the components is household consumption which involves households total spending done purchase goods and services like for example housing, food, leisure activities and healthcare.
2. **Government Consumption**: Another main component of the SFD is government consumption. Local government and state expenditures on services such as law enforcement, health, public and education. Across states the items for consumption are directly allocated to the location where consumption takes place like expenditure through the Pharmaceutical Benefits Scheme (PBS) or university expenditure by the commonwealth government for the benefit of its people. Based on indicators, remainder is allocated on defence and non-defence Commonwealth employment of government and estimated resident population.
3. **Private Investment**: Another component is the investment of Business in capital goods, which includes construction, infrastructure projects and machinery.
4. **Public Investment**: Public investment is another component of SFD which involves investments in infrastructure projects, and Government spendings on projects for people such as hospitals, schools and roads.

Within the state economy, SFD shows how much demand exists and acts as an indicator to monitor the increase in economic activity if SFD is growing which potentially leads to economic expansion and job creation. SFD helps the government and policy makers to make informed decisions on investment strategies, fiscal policy and budget allocations. SFD can help a country monitor the economic performance of various states and find areas that might need economic support or require drastic policy change.

(b) How many SFD data series does the Australian Bureau of Statistics provide? What are their differences? Which one would you use for forecasting purposes and why?

Australian Bureau of Statistics (ABS) offers original, seasonally adjusted and trend data series of State Final Demand (SFD), designed each for different kinds of analysis.

1. Original Data Series:

This is unadjusted raw data collected directly from many economic sectors. And reflects the economic activity of the economy which includes any irregularities or seasonal patterns that happen within the year. The raw data contains seasonal variations like agricultural cycles or holiday spending. The data is in raw form and can contain many outliers, and provides no adjustments for outliers, trends or seasonal patterns.

2. Seasonally Adjusted Data Series:

This data series is modified to remove any predictable seasonal impacts, like for example seasonal demand for tractors increase during the crop harvesting season and cyclical variations in construction during different seasons or increased spending during the holidays. This data series helps the policy makers to observe economic activity and underlying trends, making it easy to Corrects seasonal patterns. Seasonal adjusted data series offers a clearer picture in the economy for actual demand without misrepresentation from recurring events of season.

3. Trend Data Series:

The last data series is the trend data which is further transformed by removing both the unnecessary outliers like short term noise such as natural disasters and seasonal fluctuations and focus only on structural changes and long-term movements in economy. This data series helps the policymakers to understand the underlying trend of the data, which makes it more useful for understanding long-term shifts of economy.

Data Series	Features	Pros	Cons
Original Series	Raw, unadjusted data	Reflects actual economic activity	Volatile due to seasonal effects
Seasonally Adjusted Series	Adjusted for recurring seasonal patterns	Removes seasonal noise, clearer for short-term analysis	May hide seasonal trends
Trend Series	Smoothed to show long-term patterns	Highlights structural changes, stable	Lagging, may miss real-time changes

The Seasonally Adjusted Series is the most suitable for forecasting purposes, as the regular seasonal patterns are adjusted in this data series, which helps policy makers understand and monitor economic activity in a clear view, making it easy to model trends of future without distortion created by predictable events of season.
Furthermore, the seasonally adjusted data responds more quickly to policy changes or sudden demand shock, which is vital for making short to medium-term, accurate forecasts and understanding long term goals.

(c) What is your modelling strategy/approach and what metric would you use to measure forecast performance? Will the proposed model help explain the WA State Final Demand relative to other states? If so, how?

The VAR model would be used for analysis. The seasonal SFD data and its components would be taken for all the states of Australia. The dataset is from December 1985 to June 2024 with quarterly frequency. Var models enable us to model interdependencies among the other states in Australia and State Final Demand (SFD) of Western Australia (WA), along with the micro foundation and macroeconomic factors to capture the dynamic relationships among these variables over time. Three kinds of deterministic components would be explored to study the relationship between all variables.

- **Trend condition**: to capture any underlying trends Includes a linear time trend and long-term trends in the data
- **Constant condition**: fixed intercept is incorporated in the series to account for a non-zero mean.
- **None Condition**: neither a trend nor a constant is Assumed, letting the model depend only on the values of lagged.

We would use Root to mean squared error, AIC and BIC metrics to Measure Forecast Performance

1. **Root Mean Squared Error (RMSE):**
   It is suitable when error distribution is predicted to be distributed normally.
2. **Information Criteria (AIC, BIC):**
   BIC and AIC utilize models that fit through various specifications such as trend, none or constant and select the best model in terms of fitness.

The proposed VAR model will help to describe WA's SFD for numerous reasons. Firstly, the historical relationships would be utilized by VAR model among states to forecast WA's SFD and incorporates the past economic activities impact in various states and give a more detailed forecast by factoring both external (SFD from other states) and internal (WA's own SFD) economic dynamics. Furthermore, VAR model will explain and help the policy makers understand how the SFD changes in different states impacts the economic conditions of western Australia along with finding which states have the greatest influence on economic activity of WA and making decisions, especially planning economic policies which include resource sharing, economic development programs and interstate trade.

(d) Estimate the proposed model/s and assess its performance. Interpret the model findings.

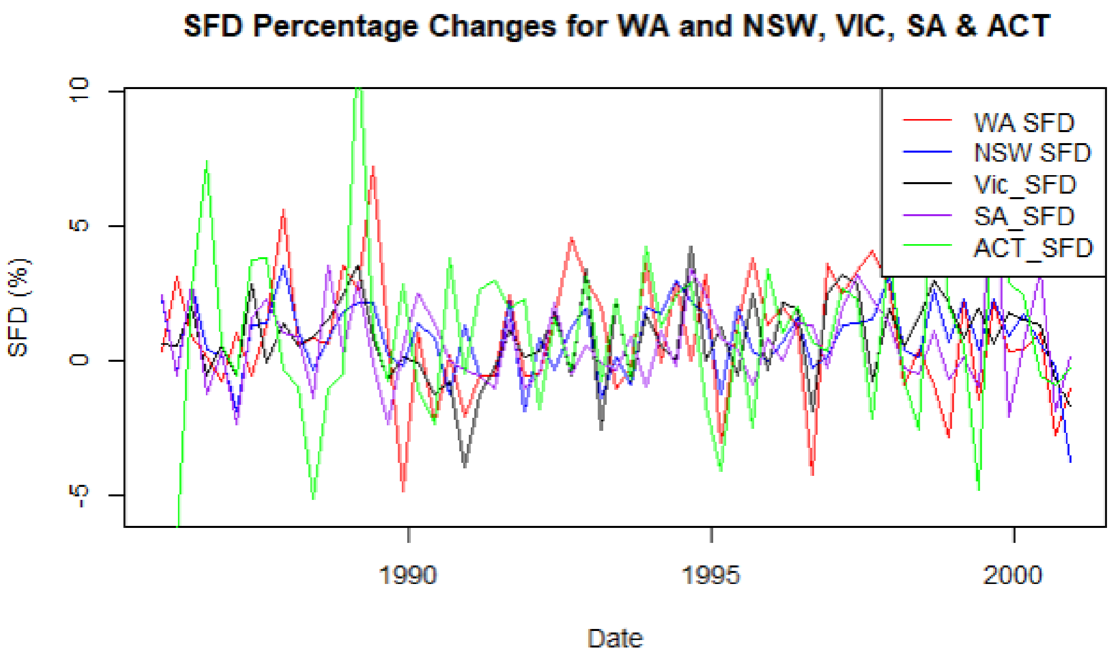
The Results of VAR model with constant, trend and none conditions are listed in the table below:

Summary Statistics:

```
     WA_SFD            NSW_SFD           Vic_SFD         queensland_SFD        SA_SFD            Tas_SFD
 Min.   :-5.6000   Min.   :-8.7000   Min.   :-8.3000   Min.   :-6.2000   Min.   :-6.3000   Min.   :-9.1000
 1st Qu.:-0.5000   1st Qu.: 0.1000   1st Qu.: 0.0500   1st Qu.: 0.1000   1st Qu.:-0.1000   1st Qu.:-0.5000
 Median : 1.0000   Median : 0.8000   Median : 0.8000   Median : 0.9000   Median : 0.6000   Median : 0.6000
 Mean   : 0.9665   Mean   : 0.7452   Mean   : 0.8129   Mean   : 0.9465   Mean   : 0.6903   Mean   : 0.6155
 3rd Qu.: 2.2500   3rd Qu.: 1.5000   3rd Qu.: 1.6000   3rd Qu.: 1.7000   3rd Qu.: 1.4000   3rd Qu.: 1.7500
 Max.   : 9.5000   Max.   : 7.7000   Max.   : 7.8000   Max.   : 7.0000   Max.   : 7.6000   Max.   : 8.6000
     NT_SFD             ACT_SFD
 Min.   :-12.5000   Min.   :-6.8000
 1st Qu.: -1.5500   1st Qu.:-0.5500
 Median :  0.8000   Median : 1.0000
 Mean   :  0.8581   Mean   : 0.9394
 3rd Qu.:  3.2500   3rd Qu.: 2.6000
 Max.   : 19.6000   Max.   :12.1000
```

The dataset includes the seasonally adjusted percentage change data for the SFDs of different states. The median of all the selected variables of the dataset has median range from 0.6 to 1.

Stationary Test:



The plot suggests that all the variables are stationary in their raw form. This could further be investigated using the ADF test.

```
        Augmented Dickey-Fuller Test

data:  SFD_ts[, "WA_SFD"]
Dickey-Fuller = -3.5519, Lag order = 5, p-value = 0.04
alternative hypothesis: stationary

New South Wales:

        Augmented Dickey-Fuller Test

data:  SFD_ts[, "NSW_SFD"]
Dickey-Fuller = -4.8627, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary

Victoria:

        Augmented Dickey-Fuller Test

data:  SFD_ts[, "Vic_SFD"]
Dickey-Fuller = -4.9138, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary

Queensland:

        Augmented Dickey-Fuller Test

data:  SFD_ts[, "queensland_SFD"]
Dickey-Fuller = -4.1341, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary

South Australia:

        Augmented Dickey-Fuller Test

data:  SFD_ts[, "SA_SFD"]
Dickey-Fuller = -5.2556, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary

Tasmania:

        Augmented Dickey-Fuller Test

data:  SFD_ts[, "Tas_SFD"]
Dickey-Fuller = -5.6928, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary

Northern Territory:

        Augmented Dickey-Fuller Test

data:  SFD_ts[, "NT_SFD"]
Dickey-Fuller = -5.3162, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary

Australian Capital Territory:

        Augmented Dickey-Fuller Test

data:  SFD_ts[, "ACT_SFD"]
Dickey-Fuller = -7.1233, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

The ADF test suggests that p-value of all the variables are 0.01 which is less than the threshold value of 0.05 which suggests that we reject the null of hypothesis, suggesting that all of the variables are stationary.

Optima Lag Selection:

```
AIC(n)  HQ(n)  SC(n) FPE(n)
     1      1      1      1
```

The optimal lag has came out to be 1 which would be used in the VAR model testing.

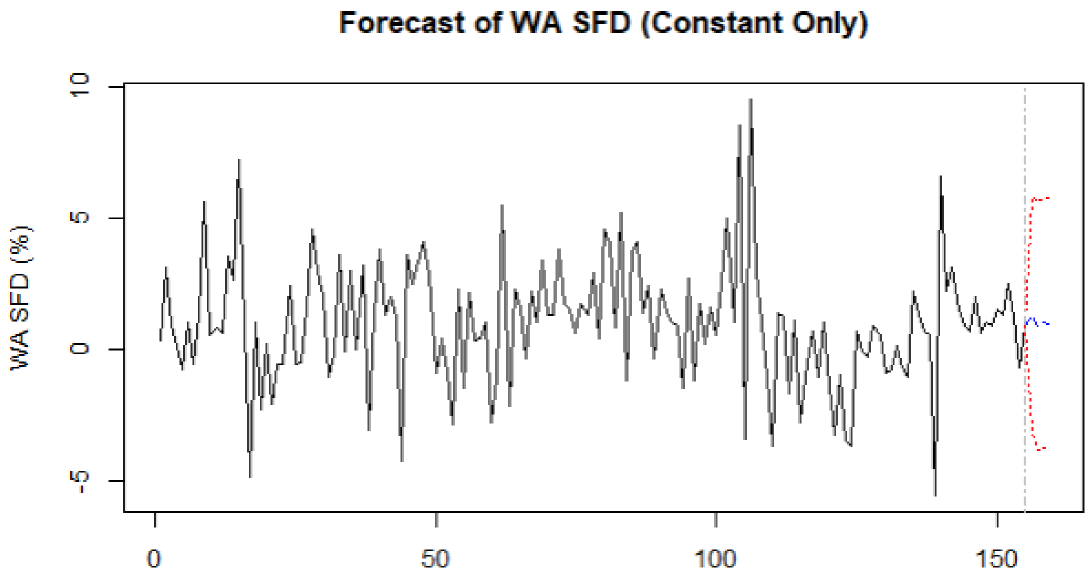
VAR MODEL RESULTS:
The Results of VAR model with constant, trend and None conditions are listed in the table below:

Model	Adjusted R-squared (WA_SFD)	Significant Variables ($p < 0.05$)	Portmanteau Test p-value (Autocorrelation)	ARCH Test p-value (Heteroscedasticity)
Constant	0.053	NSW_SFD.l1, Queensland_SFD.l1, SA_SFD.l1, Constant	0.2103	1
Trend	0.1505	NSW_SFD.l1, Queensland_SFD.l1, SA_SFD.l1, Trend	0.187	1
None	0.1238	NSW_SFD.l1, Queensland_SFD.l1, SA_SFD.l1	0.09753	1

For VAR Model with Constant condition, the significant variables for WA's SFD are NSW_SFD.l1 having the p-value of 0.020, Queensland_SFD.l1 having p- value of 0.011, SA_SFD.l1 having p-value of 0.011. The Adjusted R-squared for WA has come out to 0.053 which is very low. The Portmanteau Test for autocorrelation has a p-value of 0.2103. The high p value suggests that the absence of autocorrelation. For the ARCH test for heteroscedasticity, the p-values have come out to be 1 suggesting no heteroscedasticity

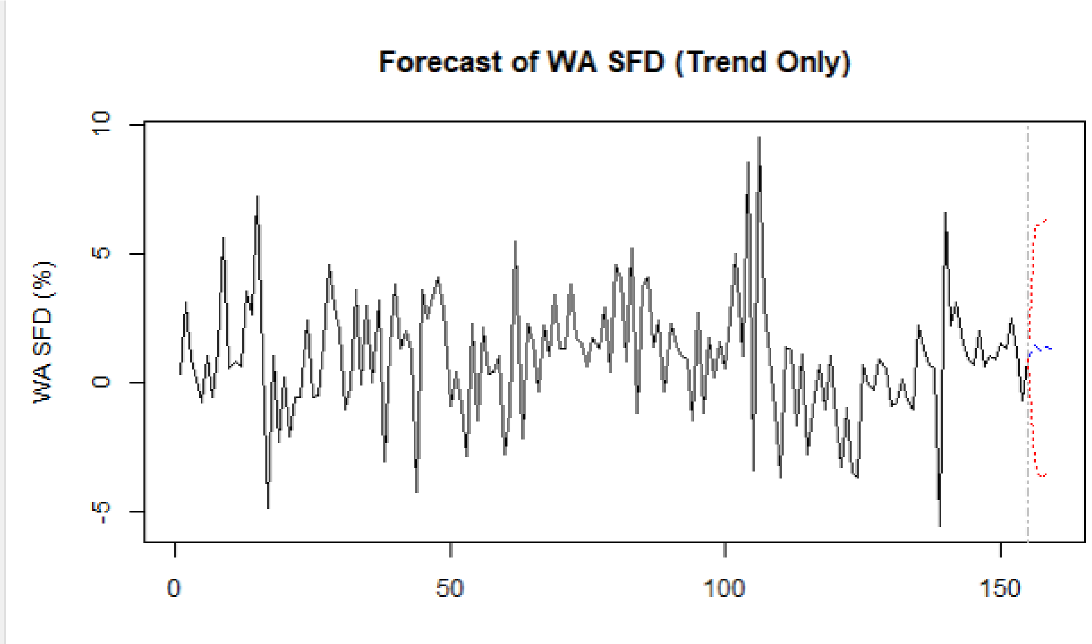
The VAR model with Trend condition suggests the significant variables for WA's SFD are NSW_SFD.l1 having p-value of 0.025, Queensland_SFD.l1 having p-value of 0.003, SA_SFD.l1 having p-value of 0.008. The adjusted R-squared for WA with trend condition has come out to be 0.1505 which is higher than constant model suggesting that 15% of the variability in WASFD can be explained by the model. The portmanteau test suggests a p-value of 0.187 suggesting there is no evidence of autocorrelation. The ARCH test has a p-value of 1 suggesting no evidence of heteroscedasticity.
The VAR model with none condition has significant variables for WA's SFD such as NSW_SFD.l1 having p-value of 0.025, Queensland_SFD.l1 having p-value of 0.001 and SA_SFD.l1 having p- value of 0.003. The adjusted R-squared for WA with none condition has come out to be 0.1238 which is higher than constant condition, however it is lower than trend condition. Portmanteau test suggests a p-value of 0.09753 indicating that there might be some autocorrelation that may exist which is close to the threshold. VAR with Trend is a better model with higher adjusted R-squared value, which shows a better fit for WA SFD with no heteroscedasticity or autocorrelation.

```
VAR Estimation Results:
=========================
Endogenous variables: WA_SFD, NSW_SFD, Vic_SFD, queensland_SFD, SA_SFD, Tas_SFD, NT_SFD, ACT_SFD
Deterministic variables: const
Sample size: 154
Log Likelihood: -2447.26
Roots of the characteristic polynomial:
0.4063 0.4063 0.3233 0.3233 0.2302 0.2302 0.1016 0.03323
Call:
VAR(y = SFD_ts, p = 1, type = "const")


Estimation results for equation WA_SFD:
=======================================
WA_SFD = WA_SFD.l1 + NSW_SFD.l1 + Vic_SFD.l1 + queensland_SFD.l1 + SA_SFD.l1 + Tas_SFD.l1 + NT_SFD.l1 + ACT_SFD.l1 + const

                   Estimate Std. Error t value Pr(>|t|)
WA_SFD.l1         -0.0368230  0.0868056  -0.424 0.672048
NSW_SFD.l1        -0.3745341  0.1592914  -2.351 0.020056 *
Vic_SFD.l1        -0.0002736  0.1517987  -0.002 0.998565
queensland_SFD.l1  0.3634905  0.1412364   2.574 0.011067 *
SA_SFD.l1          0.3931982  0.1535926   2.560 0.011490 *
Tas_SFD.l1        -0.0652885  0.0986687  -0.662 0.509217
NT_SFD.l1         -0.0162215  0.0448460  -0.362 0.718091
ACT_SFD.l1        -0.0690541  0.0740991  -0.932 0.352930
const              0.7911464  0.2309270   3.426 0.000797 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


Residual standard error: 2.313 on 145 degrees of freedom
Multiple R-Squared: 0.1025,     Adjusted R-squared: 0.05296
F-statistic:  2.07 on 8 and 145 DF,  p-value: 0.04242
```



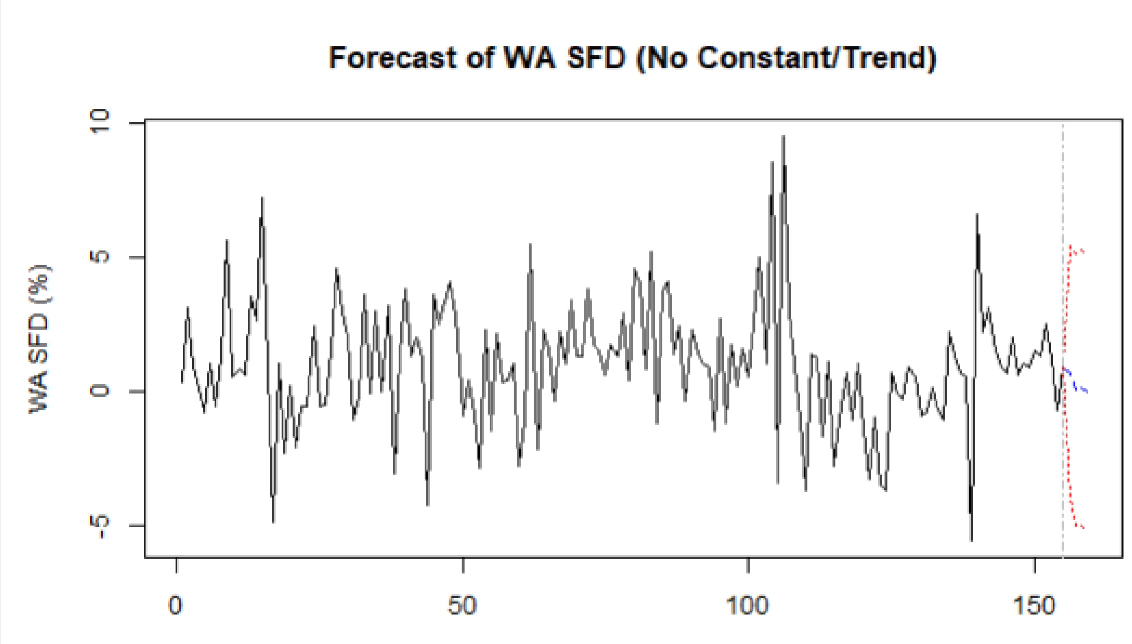
- In the forecast period using the constant condition, the forecast shows some variation at the end of actual data points with a little bit of a downward trend. The confidence interval seems wide, which shows higher forecast uncertainty, without the stabilizing impact of trend or constant.

```
VAR Estimation Results:
=========================
Endogenous variables: WA_SFD, NSW_SFD, Vic_SFD, queensland_SFD, SA_SFD, Tas_SFD, NT_SFD, ACT_SFD
Deterministic variables: trend
Sample size: 154
Log Likelihood: -2457.342
Roots of the characteristic polynomial:
0.423 0.423 0.3159 0.3159 0.2222 0.2222 0.2144 0.03753
Call:
VAR(y = SFD_ts, p = 1, type = "trend")

Estimation results for equation WA_SFD:
=======================================
WA_SFD = WA_SFD.l1 + NSW_SFD.l1 + Vic_SFD.l1 + queensland_SFD.l1 + SA_SFD.l1 + Tas_SFD.l1 + NT_SFD.l1 + ACT_SFD.l1 + trend

                  Estimate Std. Error t value Pr(>|t|)
WA_SFD.l1        -0.015075   0.088093  -0.171  0.86436
NSW_SFD.l1       -0.367486   0.162537  -2.261  0.02525 *
Vic_SFD.l1        0.017591   0.155200   0.113  0.90991
queensland_SFD.l1 0.427029   0.141748   3.013  0.00306 **
SA_SFD.l1         0.418792   0.156431   2.677  0.00828 **
Tas_SFD.l1       -0.072986   0.100645  -0.725  0.46951
NT_SFD.l1        -0.014770   0.045753  -0.323  0.74730
ACT_SFD.l1       -0.057580   0.075493  -0.763  0.44687
trend             0.005734   0.002424   2.365  0.01934 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.36 on 145 degrees of freedom
Multiple R-Squared: 0.2002,     Adjusted R-squared: 0.1505
F-statistic: 4.032 on 9 and 145 DF,  p-value: 0.0001259
```



- From historical data using the trend condition, the model shows a smooth transition in to forecast period with a continuing slight upward trend. compared to the trend or constant model, The confidence intervals are narrow. Showing more reliance and less uncertainty on an established data trend.

```
VAR Estimation Results:
=========================
Endogenous variables: WA_SFD, NSW_SFD, Vic_SFD, queensland_SFD, SA_SFD, Tas_SFD, NT_SFD, ACT_SFD
Deterministic variables: none
Sample size: 154
Log Likelihood: -2471.548
Roots of the characteristic polynomial:
0.422 0.422 0.3849 0.3193 0.3193 0.2037 0.2037 0.03324
Call:
VAR(y = SFD_ts, p = 1, type = "none")

Estimation results for equation WA_SFD:
=======================================
WA_SFD = WA_SFD.l1 + NSW_SFD.l1 + Vic_SFD.l1 + queensland_SFD.l1 + SA_SFD.l1 + Tas_SFD.l1 + NT_SFD.l1 + ACT_SFD.l1

                   Estimate Std. Error t value Pr(>|t|)
WA_SFD.l1          0.002834   0.089138   0.032  0.97468
NSW_SFD.l1        -0.374791   0.165045  -2.271  0.02462 *
Vic_SFD.l1         0.077398   0.155517   0.498  0.61946
queensland_SFD.l1  0.468074   0.142879   3.276  0.00132 **
SA_SFD.l1          0.463149   0.157728   2.936  0.00386 **
Tas_SFD.l1        -0.071610   0.102215  -0.701  0.48468
NT_SFD.l1         -0.015769   0.046466  -0.339  0.73482
ACT_SFD.l1        -0.036998   0.076161  -0.486  0.62785
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.397 on 146 degrees of freedom
Multiple R-Squared: 0.1693,     Adjusted R-squared: 0.1238
F-statistic: 3.719 on 8 and 146 DF,  p-value: 0.0005509
```



The forecast using the none condition shows that towards the end of actual data series, the forecast starts leveling off, with a little bit upward tendency in forecast period. This is similar to the trend model confidence intervals but shows a little more fluctuation which shows slight uncertainty in the forecast which is more than trend model

```
        Portmanteau Test (asymptotic)

data:  Residuals of VAR object var_model_const
Chi-squared = 603.09, df = 576, p-value = 0.2103


        Portmanteau Test (asymptotic)

data:  Residuals of VAR object var_model_trend
Chi-squared = 606.02, df = 576, p-value = 0.187


        Portmanteau Test (asymptotic)

data:  Residuals of VAR object var_model_none
Chi-squared = 620.41, df = 576, p-value = 0.09753
```

```
        ARCH (multivariate)

data:  Residuals of VAR object var_model_const
Chi-squared = 5364, df = 6480, p-value = 1


        ARCH (multivariate)

data:  Residuals of VAR object var_model_trend
Chi-squared = 5364, df = 6480, p-value = 1


        ARCH (multivariate)

data:  Residuals of VAR object var_model_none
Chi-squared = 5364, df = 6480, p-value = 1
```

The model with trend condition is seen to give the most stable forecast, with the narrower confidence intervals and least change in values of future, showing it captures the underlying data pattern effectively. Compared to the other models, The trend model provides more confidence in its forecasts, as shown by narrower confidence intervals. It indicates that trendy component incorporation might be important for capturing in WA's SFD the long-term movements. Model constant only indicates consistent levels off with upward tendency, showing model is picking some kind of average Impacts but might not capture fully dynamic changes with time like model trend.

Reference

Known components of expenditure - GSP(E). (2020). https://www.abs.gov.au/statistics/detailed-methodology-information/concepts-sources-methods/australian-system-national-accounts-concepts-sources-and-methods/2020-21/chapter-21-state-accounts/known-components-expenditure-gspe