STOCK PREDICTION USING FINANCIAL TIME SERIES ANALYSIS

Atheesh Krishnan
Rajeev Surve
Sagar Marathe
Snigdha Jadhav

April 2024

Abstract

The stock market's inherent complexity and dynamism are influenced by numerous factors including economic conditions, international events, and investor sentiment. In this context, this project undertook a detailed stock prediction analysis of two companies, Vale and Target, through both intersectoral and intertemporal lenses. Our focus was on cross-sector analysis, and for Target, the analysis spanned two distinct time frames to capture temporal shifts in market dynamics.

The study utilised data from over 20 variables believed to affect the stocks' prices, with multivariate regression methods identifying key predictors for each scenario. The study verified Gauss-Markov assumptions by testing for autocorrelation and heteroskedasticity. Additionally, tests were conducted on residuals and squared residuals to check model fit, while ACF and PACF plots were analysed to assess the need for ARIMA and GARCH models. DCF modelling, performed for the period 2018 to 2023, provided values that were integrated into another predictive model to evaluate stock pricing. Based on these comprehensive analyses, recommendations were made for future investment strategies. This methodology effectively identifies key variables influencing stock performance and enhances financial forecasting capabilities.

OUTLINE

1. Objective

2. Data Acquisition

A. Sources of data for stock prices and other influencing variables.

3. Methodology

3.1 Intersectoral Analysis

A. Description of the sectors: mining (Vale) and retail (Target).
B. Time frame for analysis.
C. Procedure for data analysis and variable selection.

3.2 Intertemporal Analysis

A. Detailed time frames for Target: 2012 - 2018 and 2018 - 2024.
B. Comparison and contrast between the two periods.

3.3 Regression Analysis

A. Execution of regular multilinear regression.
B. Calculation and interpretation of AIC/BIC scores for model optimization.
C. Identification of the most significant variables affecting stock prices.

3.4 Validation of Model Assumptions

A. Checking Gauss-Markov Assumptions for model validity.
B. Autocorrelation tests on residuals to consider ARIMA modelling.
C. Analysis of squared residuals to check for volatility clustering (GARCH).

4. Prediction Model Development

A. Integration of Discounted Cash Flow (DCF) model outputs for financial variables.
B. Application of financial time series models using DCF-derived values.

5. Results

A. Presentation of model predictions for Vale and Target stock prices.
B. Analysis of stock underpricing or overpricing based on the model predictions.

6. Recommendations

A. Investment recommendations based on predicted stock prices.
B. Strategy for going long or short on Vale and Target stocks for the upcoming year.

7. Conclusion

A. Summary of key findings and their implications for investors.
B. Reflection on the effectiveness of multivariable regression and time series analysis in stock price prediction.

OBJECTIVE

The primary objective of this project is to conduct a comprehensive and detailed analysis of stock performance for Vale and Target, employing both intersectoral and intertemporal approaches to uncover underlying market dynamics and the factors influencing stock prices. This involves:

- Gathering and analyzing data from over 20 potential influencing variables to construct and refine multivariate regression models.

- Testing for adherence to Gauss-Markov assumptions, including checks for autocorrelation and heteroskedasticity, and applying the Ljung-Box test on the squared residuals to ensure the robustness and reliability of the regression models.

- Examining Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots to evaluate the need for integrating ARIMA and GARCH models, which would further enhance predictive accuracy.

- Utilizing Discounted Cash Flow (DCF) modeling to assess the valuation of the stocks from 2018 to 2023, identifying potential overpricing or underpricing scenarios.

- Ultimately, the project aims to leverage the insights gained from these analyses to provide evidence-based investment recommendations, enabling investors to make informed decisions in a dynamic and often unpredictable market environment.

DATA ACQUISITION

Data was acquired from various sources for the variables we believe would affect the stock price of Vale and Target. The sources for data are as follows:-

- Wharton Data Research Services (WRDS)
- Federal Reserve Economic Data (FRED)
- CapitalIQ
- Yahoo Finance

INTERSECTORAL ANALYSIS

I. Mining Sector

We begin our exploration with an intersectoral analysis, focusing on the divergent dynamics of two distinct sectors. We examine *Vale*, a key player in the mining sector, and Target, a major retailer, both analysed over the concurrent timeframe of 2018 to 2023. This comparative study allows us to uncover the unique economic and market forces shaping each industry. Our analysis starts with *Vale*, delving into the intricacies of the mining sector and its influence on stock performance.

Regression Analysis

We have considered 27 variables for the multivariate regression for the share price of Vale.

Following are the variables we believe are important for the regression:-

x1:Volume
x2: Rio Tinto price
x3: BHP Price
x4: VIX
x5: Iron ore price
x6: Copper price
x7: Enterprise Value
x8: ResearchAndDevelopment
x9: TotalRevenue
x10: CostOfRevenue
x11: TaxProvision
x12: DepreciationAmortizationDepletion
x13: ChangeInWorkingCapital
x14: Free Cash Flow

x15: Tangible Book Value
x16: Total Debt
x17: Ordinary Shares Number
x18: Capital Expenditures
x19: Total Assets
x20: Cash And Cash Equivalents
x21: Inventory
x22: FinancialAssets
x23: Working Capital
x24: Repurchase Of Capital Stock
x25: Cash Dividends Paid
x26: Long Term Debt Payments
x27: CashFlowFromContinuingFinancingActivities

Therefore the model we have used is as follows:-

$$y = \beta_0 + \beta_1 x1 + \beta_2 x2 + \dots\dots\dots + \beta_{27} x27$$

Based on this we have conducted AIC and BIC tests to find the most significant variables out of these 27 variables that affect the price.

AIC:

```
> summary(lm_AIC)

Call:
lm(formula = y ~ x1 + x2 + x4 + x5 + x9 + x11 + x12 + x13 + x14 +
    x15 + x16 + x17 + x18 + x19 + x21 + x22 + x24, data = vale)

Residuals:
     Min       1Q   Median       3Q      Max
-1.03583 -0.26441  0.00184  0.27520  0.93584

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.672e+00  1.013e+01   0.165  0.86977
x1           1.046e-09  6.268e-10   1.669  0.10246
x2           1.829e-01  2.924e-02   6.253 1.72e-07 ***
x4          -4.812e-02  1.854e-02  -2.596  0.01295 *
x5           4.052e-02  7.417e-03   5.463 2.34e-06 ***
x9           1.189e-09  4.189e-10   2.839  0.00695 **
x11         -1.978e-09  5.647e-10  -3.502  0.00111 **
x12         -5.988e-09  2.300e-09  -2.604  0.01269 *
x13         -2.647e-10  1.350e-10  -1.961  0.05659 .
x14         -5.245e-10  2.414e-10  -2.173  0.03546 *
x15          7.215e-10  3.880e-10   1.859  0.06998 .
x16          5.384e-10  3.318e-10   1.622  0.11221
x17         -1.601e-08  7.257e-09  -2.206  0.03289 *
x18          7.930e-09  1.380e-09   5.749 9.13e-07 ***
x19          4.219e-10  1.274e-10   3.312  0.00191 **
x21          3.409e-09  5.848e-10   5.830 6.98e-07 ***
x22         -2.889e-08  9.867e-09  -2.928  0.00549 **
x24          6.701e-10  2.967e-10   2.259  0.02916 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5427 on 42 degrees of freedom
Multiple R-squared:  0.9778,	Adjusted R-squared:  0.9688
F-statistic: 108.6 on 17 and 42 DF,  p-value: < 2.2e-16
```

BIC:

```
> summary(lm_BIC)

Call:
lm(formula = y ~ x2 + x4 + x5 + x9 + x11 + x12 + x13 + x17 + 
    x18 + x19 + x21 + x22, data = vale)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.08979 -0.41728 -0.00474  0.28160  1.13857 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.569e+01  4.920e+00  -3.189  0.00254 ** 
x2           1.812e-01  2.698e-02   6.718 2.19e-08 ***
x4          -3.982e-02  1.556e-02  -2.559  0.01376 *  
x5           3.541e-02  5.925e-03   5.977 2.92e-07 ***
x9           3.094e-10  7.206e-11   4.294 8.71e-05 ***
x11         -9.855e-10  2.074e-10  -4.752 1.94e-05 ***
x12         -4.437e-09  1.766e-09  -2.513  0.01545 *  
x13         -4.373e-10  9.820e-11  -4.453 5.21e-05 ***
x17         -2.586e-09  9.637e-10  -2.683  0.01003 *  
x18          5.871e-09  9.261e-10   6.340 8.23e-08 ***
x19          1.900e-10  3.438e-11   5.528 1.38e-06 ***
x21          4.206e-09  4.488e-10   9.371 2.50e-12 ***
x22         -1.126e-08  1.463e-09  -7.695 7.28e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5718 on 47 degrees of freedom
Multiple R-squared:  0.9724,	Adjusted R-squared:  0.9653 
F-statistic: 137.8 on 12 and 47 DF,  p-value: < 2.2e-16
```

Since, the F-stat value for BIC is greatest the BIC model parameters are finalised.

```
> model_2 <- lm(y ~ x2 + x4 + x5 +x9 + x11 + x12 + x13 + x17 + x18 + x19 + x21 +  x22, data =vale)
> summary(model_2)

Call:
lm(formula = y ~ x2 + x4 + x5 + x9 + x11 + x12 + x13 + x17 + 
    x18 + x19 + x21 + x22, data = vale)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.08979 -0.41728 -0.00474  0.28160  1.13857 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.569e+01  4.920e+00  -3.189  0.00254 ** 
x2           1.812e-01  2.698e-02   6.718 2.19e-08 ***
x4          -3.982e-02  1.556e-02  -2.559  0.01376 *  
x5           3.541e-02  5.925e-03   5.977 2.92e-07 ***
x9           3.094e-10  7.206e-11   4.294 8.71e-05 ***
x11         -9.855e-10  2.074e-10  -4.752 1.94e-05 ***
x12         -4.437e-09  1.766e-09  -2.513  0.01545 *  
x13         -4.373e-10  9.820e-11  -4.453 5.21e-05 ***
x17         -2.586e-09  9.637e-10  -2.683  0.01003 *  
x18          5.871e-09  9.261e-10   6.340 8.23e-08 ***
x19          1.900e-10  3.438e-11   5.528 1.38e-06 ***
x21          4.206e-09  4.488e-10   9.371 2.50e-12 ***
x22         -1.126e-08  1.463e-09  -7.695 7.28e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5718 on 47 degrees of freedom
Multiple R-squared:  0.9724,	Adjusted R-squared:  0.9653 
F-statistic: 137.8 on 12 and 47 DF,  p-value: < 2.2e-16
```

Gauss Markov Assumptions

A. Auto-correlation

For autocorrelation we have conducted the Durbin Watson test.

```
> durbinWatsonTest(model_2)
 lag Autocorrelation D-W Statistic p-value
   1      -0.07981911       2.15523   0.474
 Alternative hypothesis: rho != 0
```

The p-value is greater than the significance level, therefore, we do not reject the null hypothesis for Durbin Watson. There is enough evidence that there is no presence of autocorrelation.

B. Conditional Heteroskedasticity

We have conducted Breusch-Pagan test to check for conditional heteroskedasticity.

```
> bptest(model_2)

        studentized Breusch-Pagan test

data:  model_2
BP = 12.367, df = 12, p-value = 0.4166
```

The p-value is greater than the significance level, therefore, we do not reject the null hypothesis for Breusch-Pagan. There is enough evidence that there is no conditional heteroskedasticity.

C. Multicollinearity

For Multicollinearity we have conducted the VIF test.

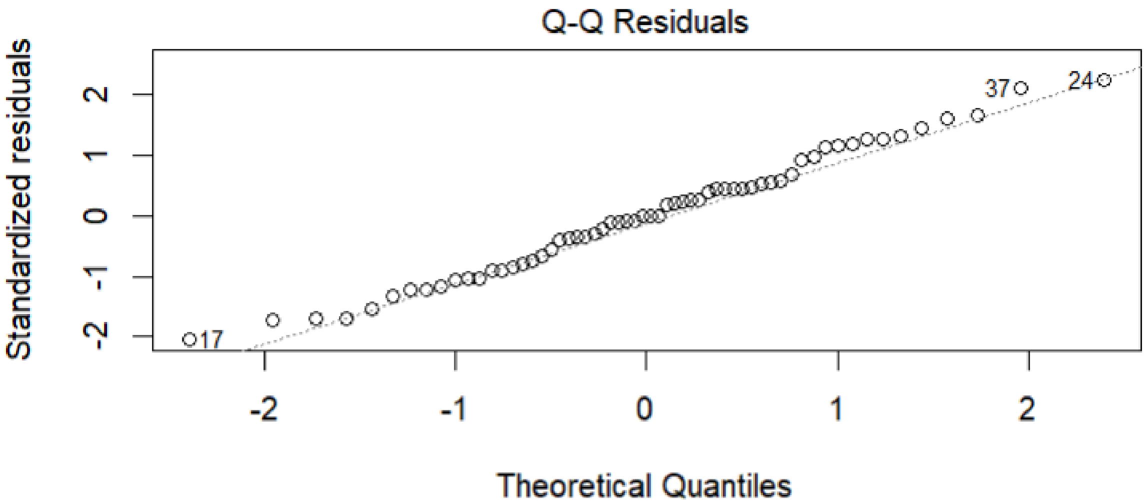
```
> vif(model_2)
       x2        x4        x5        x9       x11       x12       x13       x17       x18       x19       x21       x22 
17.771489  2.776474  5.800610  4.855783  7.268632  5.030030  4.490721 14.985466 13.370283  4.802091  5.986026  9.535180 
```

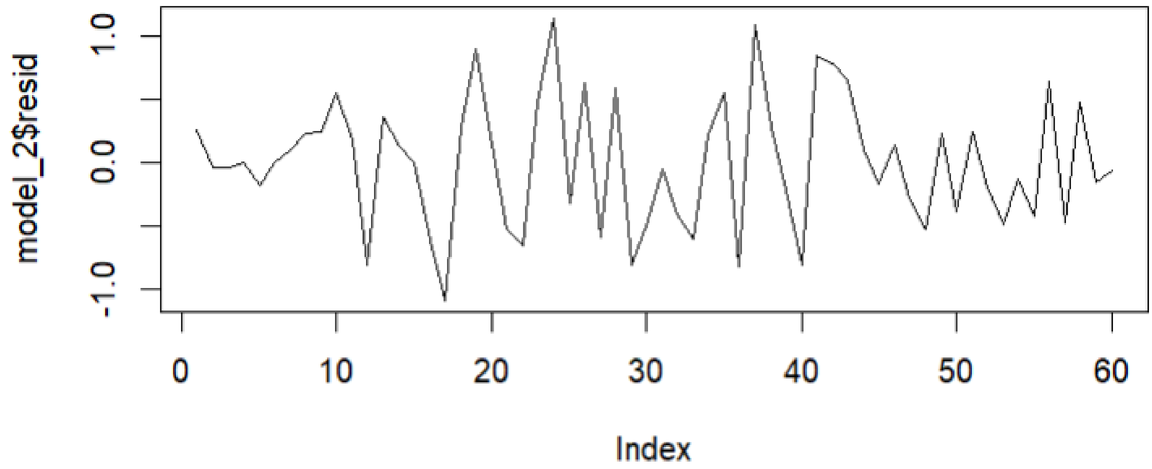
We can observe that a lot of variables have higher values which signify that these variables are collinear with the share price (y). Higher the value better will be the stock prediction.

Following is the share price output for testing the model fitness with respect to the model we developed using only the significant variables.

```
> fitted(model_2)
        1         2         3         4         5         6         7         8         9        10        11        12 
 7.999294  8.936782  8.634053  7.286973  7.791203  7.771532  7.681255  8.507888  7.731406  6.125253  5.459549  6.412708 
       13        14        15        16        17        18        19        20        21        22        23        24 
 6.278756  6.877687  7.904724  8.088321  8.276344  7.203769  9.422412 11.729164 11.959377 12.620668 11.835785 13.766572 
       25        26        27        28        29        30        31        32        33        34        35        36 
16.268597 16.277674 16.462313 13.816463 11.344067 11.122549 10.373612 12.115695 13.268767 15.211127 16.131620 14.918126 
       37        38        39        40        41        42        43        44        45        46        47        48 
13.979943 11.927526 11.447104 11.173743 10.851083 10.579077 13.822435 14.794964 16.611335 14.254999 14.181434 13.498212 
       49        50        51        52        53        54        55        56        57        58        59        60 
11.185116 12.472252 12.928667 12.057340 12.927214 12.856176 14.350890 14.539036 13.571512 12.360395 11.823924 12.330428 
```

D. Normality of residuals





The first plot gives us information that there aren't any significant outliers for residual plot. The latter two plots show us that the residual values have magnitudes closer to zero.

Residual Analysis

Ljung-Box test for residuals

```
> Box.test(model_2$resid,lag=10,type='Ljung')

        Box-Ljung test

data:  model_2$resid
X-squared = 8.4633, df = 10, p-value = 0.5837
```

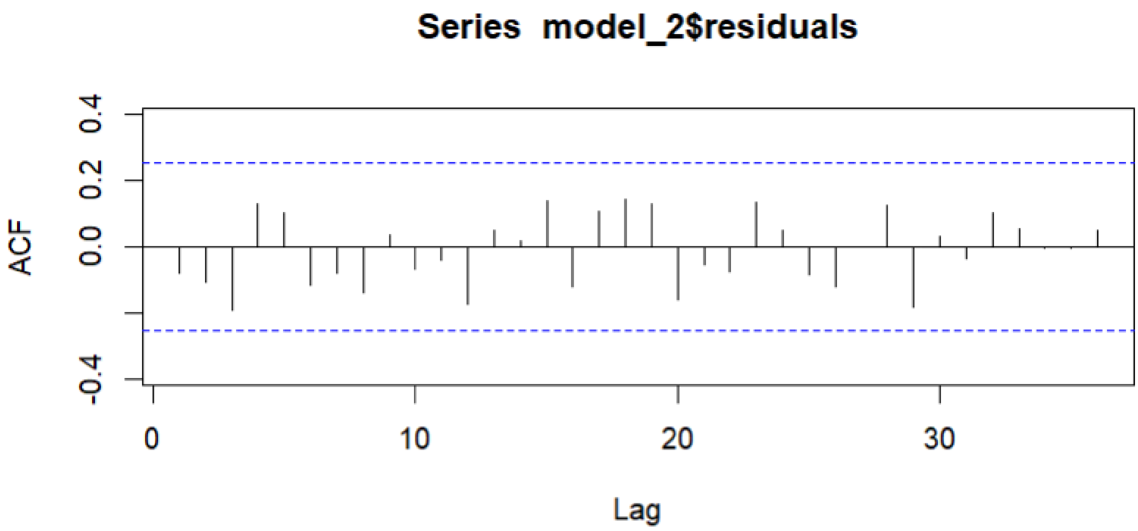
The p-value is greater than the significance level, therefore, we do not reject the null hypothesis for the Ljung-Box test. There is enough evidence that there is no autocorrelation in residuals.

Ljung-Box test for square of residuals

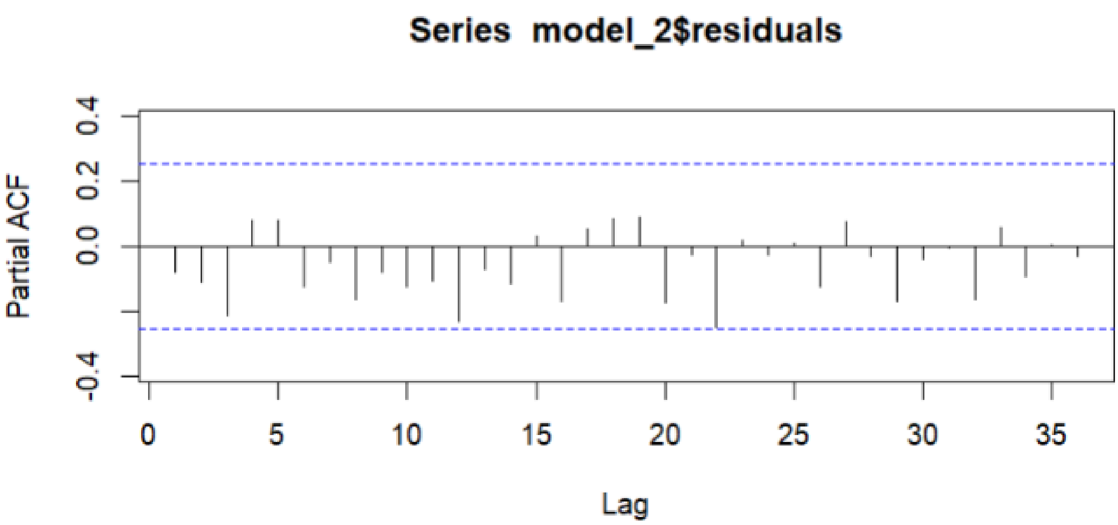
```
> Box.test(k^2,lag=10,type='Ljung')

        Box-Ljung test

data:  k^2
X-squared = 17.378, df = 10, p-value = 0.0664
```

The p-value is slightly greater than the significance level, therefore, we do not reject the null hypothesis for the Ljung-Box test. The volatility clustering is insignificant for significance level 5%.

ACF plot

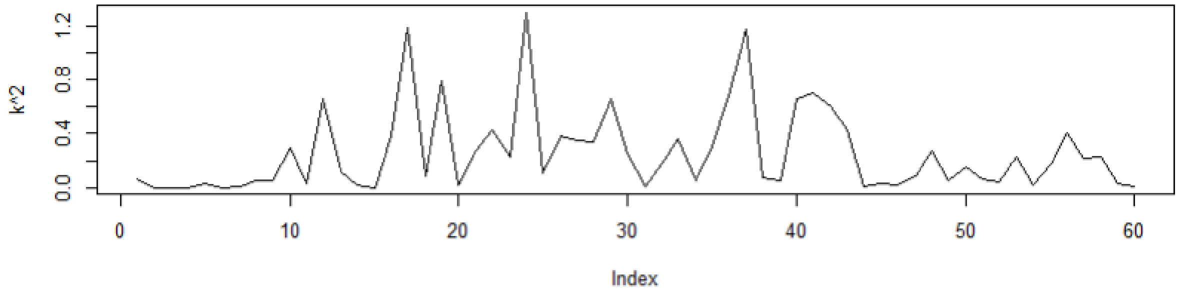


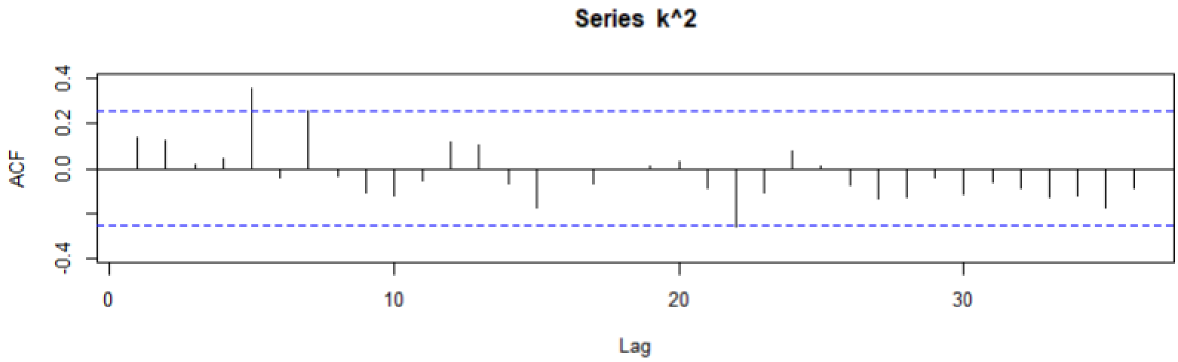
PACF plot

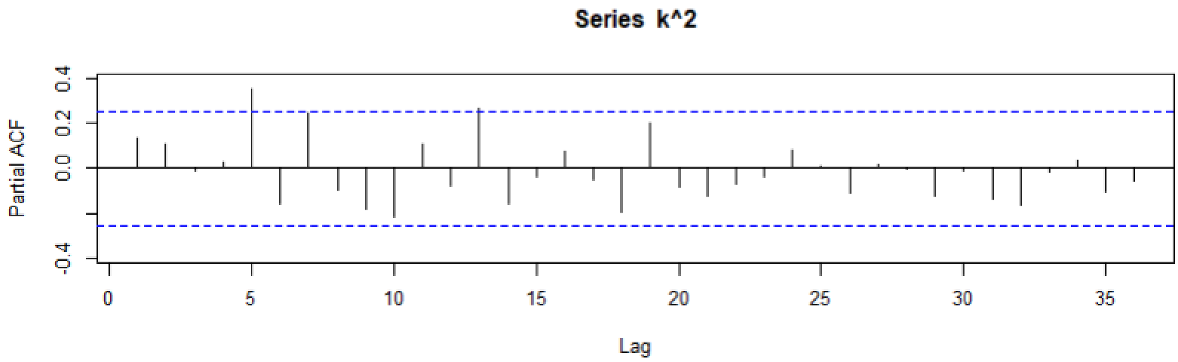


We can see from the plots that no line intersects the threshold line which signifies that there is no need for an ARIMA model.

Residual Analysis for volatility clustering setup







Stationarity

Augmented Dickey Fuller test is conducted to check for stationarity of the series.

```
> adf.test(model_2$residuals)

        Augmented Dickey-Fuller Test

data:  model_2$residuals
Dickey-Fuller = -4.0323, Lag order = 3, p-value = 0.01428
alternative hypothesis: stationary
```

The p-value is less than the significance level, therefore, we reject the null hypothesis for the ADF test. There is enough evidence that the series is stationary.

ARIMA Model testing

Command is passed to check for the best ARIMA model required.

```
> auto.arima(model_2$residuals, d = 0)
Series: model_2$residuals
ARIMA(0,0,0) with zero mean

sigma^2 = 0.2562:  log likelihood = -44.28
AIC=90.55   AICc=90.62   BIC=92.65
```

We can observe that there is no need to incorporate the ARIMA model which complements our decision to not consider ARIMA from the ACF and PACF plots.

Garch Model testing

Garch (1,1)

```
 garchFit(formula = ~garch(1, 1), data = model_2$residuals, cond.dist = "std",
    Trace = F)

Mean and Variance Equation:
 data ~ garch(1, 1)
<environment: 0x000001e5db563340>
 [data = model_2$residuals]

Conditional Distribution:
 std

Coefficient(s):
         mu        omega       alpha1        beta1        shape
-5.4549e-17   3.6410e-02   4.8096e-01   4.9151e-01   1.0000e+01

Std. Errors:
 based on Hessian

Error Analysis:
         Estimate  Std. Error  t value Pr(>|t|)
mu     -5.455e-17   6.559e-02    0.000    1.000
omega   3.641e-02   4.676e-02    0.779    0.436
alpha1  4.810e-01   4.494e-01    1.070    0.284
beta1   4.915e-01   3.494e-01    1.407    0.160
shape   1.000e+01   6.724e+00    1.487    0.137

> predict(model_3,5)
   meanForecast meanError standardDeviation
1 -5.454861e-17 0.3645176         0.3645176
2 -5.454861e-17 0.4069705         0.4069705
3 -5.454861e-17 0.4443818         0.4443818
4 -5.454861e-17 0.4779630         0.4779630
5 -5.454861e-17 0.5084971         0.5084971
```

Garch (5,5)

```
 garchFit(formula = ~garch(5, 5), data = model_2$residuals, cond.dist = "std",
    Trace = F)

Mean and Variance Equation:
 data ~ garch(5, 5)
<environment: 0x000001e5c83d8e70>
 [data = model_2$residuals]

Conditional Distribution:
 std

Coefficient(s):
         mu        omega       alpha1       alpha2       alpha3       alpha4       alpha5
-5.4549e-17   9.6874e-02   1.0000e-08   8.1874e-02   1.0000e-08   1.7281e-02   3.6897e-01
      beta1        beta2        beta3        beta4        beta5        shape
 1.0000e-08   2.3904e-01   1.0000e-08   1.0000e-08   1.0000e-08   1.0000e+01

Std. Errors:
 based on Hessian

Error Analysis:
         Estimate  Std. Error  t value Pr(>|t|)
mu     -5.455e-17   9.139e-02    0.000    1.000
omega   9.687e-02         NaN      NaN      NaN
alpha1  1.000e-08   5.268e-02    0.000    1.000
alpha2  8.187e-02   1.847e-01    0.443    0.658
alpha3  1.000e-08         NaN      NaN      NaN
alpha4  1.728e-02   9.313e-02    0.186    0.853
alpha5  3.690e-01   2.887e-01    1.278    0.201
beta1   1.000e-08         NaN      NaN      NaN
beta2   2.390e-01   6.891e-01    0.347    0.729
beta3   1.000e-08         NaN      NaN      NaN
beta4   1.000e-08         NaN      NaN      NaN
beta5   1.000e-08         NaN      NaN      NaN
shape   1.000e+01   6.534e+00    1.530    0.126

> predict(model_3,5)
  meanForecast meanError standardDeviation
1 -5.454861e-17 0.5416402         0.5416402
2 -5.454861e-17 0.4920325         0.4920325
3 -5.454861e-17 0.5233373         0.5233373
4 -5.454861e-17 0.4288173         0.4288173
5 -5.454861e-17 0.4372437         0.4372437
```

For both Garch models the forecasted values are very small and insignificant. Therefore the Garch modelling is neglected for Vale.

Discounted Cash Flow (DCF) Model

This financial model was constructed to predict the values for Vale for the next five years. The values of the variables we have for our regression model is taken from this financial model to predict for 2024.

INCOME STATEMENTS

		Historical	Projected						
		2023	2024	2025	2026	2027	2028	2029	2030
Revenue		41,613.0	43,485.6	45,442.4	47,487.3	49,624.3	51,857.4	54,191.0	56,629.5
Less: Total COGS		(27,063.0)	(26,221.8)	(27,401.8)	(28,634.9)	(29,923.4)	(31,270.0)	(32,677.1)	(34,147.6)
Gross Profit		14,550.0	17,263.8	18,040.6	18,852.5	19,700.8	20,587.4	21,513.8	22,481.9
Less: Total SG&A		(2,004.6)	(2,087.3)	(2,181.2)	(2,279.4)	(2,382.0)	(2,489.2)	(2,601.2)	(2,718.2)
EBIT		12,545.4	15,176.5	15,859.4	16,573.1	17,318.9	18,098.2	18,912.6	19,763.7
Interest & Other Expense / (Income):	*Rate*								
Total Interest Expense	**5.8%**		1,392.6	1,165.0	1,255.2	2,251.2	1,553.4	1,553.4	1,553.4
Less: Interest Income	**4.0%**		**(94.0)**	**(91.0)**	**(95.1)**	**(99.4)**	**(103.9)**	**(108.6)**	**(113.5)**
Financing Costs Amortization	**7.0 y**		**17.1**	**17.1**	**17.1**	**17.1**	**17.1**	**17.1**	**17.1**
Pretax Income			13,860.7	14,768.4	15,395.9	15,150.0	16,631.6	17,450.7	18,306.7
Less: Income Taxes	*34.00%*		(4,712.7)	(5,021.2)	(5,234.6)	(5,151.0)	(5,654.8)	(5,933.2)	(6,224.3)
Net Income			9,148.1	9,747.1	10,161.3	9,999.0	10,976.9	11,517.5	12,082.4
Shares Outstanding			4,300	4,300	4,300	4,300	4,300	4,300	4,300
Earnings per Share (EPS)			$2.1	$2.3	$2.4	$2.3	$2.6	$2.7	$2.8
EBITDA Reconciliation:									
EBIT			15,176.5	15,859.4	16,573.1	17,318.9	18,098.2	18,912.6	19,763.7
Plus: Depreciation & Amortisation			3,174.4	3,317.3	3,466.6	3,622.6	3,785.6	3,955.9	4,134.0
EBITDA			18,350.9	19,176.7	20,039.7	20,941.4	21,883.8	22,868.6	23,897.7

BALANCE SHEETS

	Historical	Projected						
	Dec-23	Dec-24	Dec-25	Dec-26	Dec-27	Dec-28	Dec-29	Dec-30
ASSETS:								
Required Cash	$2,349.2	$2,955.4	$3,088.4	$3,227.4	$3,372.6	$3,524.4	$3,683.0	$3,848.7
Excess Cash	$1,195.8	$479.4	$501.0	$523.5	$547.1	$571.7	$597.4	$624.3
Total Cash	$3,545.0	3,434.8	3,589.4	3,750.9	3,919.7	4,096.1	4,280.4	4,473.0
Accounts Receivable	4,063.0	4,245.8	4,436.9	4,636.6	4,845.2	5,063.2	5,291.1	5,529.2
Inventory	4,536.0	4,395.0	4,592.8	4,799.5	5,015.4	5,241.1	5,477.0	5,723.4
Prepaid Expenses	429.0	415.7	434.4	453.9	474.3	495.7	518.0	541.3
Other	5,532.0	7,341.5	7,671.9	8,017.1	8,377.9	8,754.9	9,148.8	9,560.5
Current Assets	18,105.0	19,832.8	20,725.3	21,657.9	22,632.5	23,651.0	24,715.3	25,827.5
PP&E - Gross	78,884.8	85,384.8	92,201.2	99,324.3	106767.9088	114546.5142	122675.16	131169.5883
Less: Accum. Depn.	(32,024.4)	(35,198.8)	(38,516.1)	(41,982.7)	(45,605.3)	(49,390.9)	(53,346.8)	(57,480.8)
Net PP&E	46,860.4	50,186.0	53,685.0	57,341.5	61,162.6	65,155.6	69,328.3	73,688.8
Intangibles	11,261.8	12,318.0	11,793.6	15,576.2	17,217.2	19,014.6	20,899.9	22,877.3
Other Non-Current assets	14,850.3	20,199.4	29,631.0	35,157.9	42,413.3	50,599.6	59,205.4	68,250.4
Financing cost	119.7	102.6	85.5	68.4	51.3	34.2	17.1	0.0
Total Assets	$91,197	$102,639	$115,920	$129,802	$143,477	$158,455	$174,166	$190,644
LIABILITIES & EQUITY:								
Accounts Payable	4,101.0	3,973.5	4,152.3	4,339.2	4,534.5	4,738.5	4,951.7	5,174.6
Accrued Expenses	637.0	617.2	645.0	674.0	704.3	736.0	769.1	803.8
Other	9,451.6	9,157.8	9,569.9	10,000.6	10,450.6	10,920.9	11,412.3	11,925.9
Current Liabilities	14,189.6	13,748.5	14,367.2	15,013.8	15,689.4	16,395.4	17,133.2	17,904.2
Capital Lease Obligations	1,161.0	1,427.8	1,712.7	2,009.6	2,301.8	2,622.6	2,959.1	3,312.2
Long Term Loan	10,807.5	13,297.0	15,949.9	18,715.5	21,436.5	24,423.6	27,557.8	30,845.5
Total Debt	11,968.0	14,724.8	17,662.6	20,725.1	23,738.3	27,046.2	30,516.9	34,157.7
Other Liabilities	25,374.2	25,351.9	25,330.0	25,341.2	25,328.3	25,315.6	25,300.6	25,284.4
Total Liabilities	51,531.8	53,825.3	57,359.8	61,080.0	64,756.0	68,757.2	72,950.7	77,346.3
Common Equity	39,665.4	48,813.5	58,560.6	68,721.9	78,720.9	89,697.8	101215.28	113297.6823
Liabilities & Equity	$91,197	$102,639	$115,920	$129,802	$143,477	$158,455	$174,166	$190,644
Check	0.000	0.000	0.00	0.000	0.000	0.000	0.000	0.000

Relative Valuation Calculation	Share Price
P/FCF	14.29
P/B	15.84
EV/EBITDA	14.96
Relative yeild	15.80

DCF Forecast	20.94
DDM (Historical analysis) forecast	18.01
DDM gordon Growth Model	19.51

RESULTS

Vale Stock Prediction

We have therefore developed from the significant variables keeping in mind the effects of ARCH and GARCH. Using the predicted values for those variables from the DCF model we will now predict the price for the following year.

Following is the our model and the values we need to use for the prediction:-

$$\mathbf{y \sim x2 + x4 + x5 + x9 + x11 + x12 + x13 + x17 + x18 + x19 + x21 + x22}$$

y = Share price

x2 = Rio Tinto price = 76

x4 = VIX = 21.67

x5 = Iron ore price = 120.53

x9 = Total revenue = 10871400000

x11 = Tax provision = 1178317500

x12 = depreciation = 742740000.00; 804583333.33(average)

x13 = change in working capital = -123140000

x17 = shares outstanding = 4303392869

x18 = capital expenditure = -1625160000

x19 = total assets = 102639700000

x21 = inventory= 5426685000; 4660350000.00(average)

x22 = financial assets = 205291666

```
> x <-predict(model_2, newdata=new)
> x
       1
16.57961
```

RECOMMENDATIONS

This Forecast is based on 2024 annual statement forecasts. Based on our chosen model and the new data obtained from the values of 2024 Financial model, we get a stock price prediction of *16.57961* for 2024. Current stock price is *11.60*. Hence, we conclude that the stock is underpriced and provide a recommendation of holding a *long position* for Vale stock.

II. Retail Sector

Now We continue our exploration with an intersectoral analysis, focusing on Target, a major retailer, with our analysis time frame being 2018 to 2023.

Regression Analysis

We have considered 23 variables for the multivariate regression for the share price of Target.

Following are the variables we believe are important for the regression:-

x1: CPI
x2: VIX
x3: EnterpriseValue
x5: WMT_Stock_price
x5: EnterprisesValueRevenueRatio
x6: EnterprisesValueEBITDARatio
x7: TotalRevenue
x8: CostOfRevenue
x9: GrossProfit
x10: OperatingExpense
x11: OperatingIncome
x12: EBIT
x13: EBITDA
x14: TotalAssets
x15: CashAndCashEquivalents
x16: Inventory
x17: LongTermDebt
x18: WorkingCapital
x19: OrdinarySharesNumber
x20: OperatingCashFlow
x21: ChangeInWorkingCapital
x22: CapitalExpenditure
x23: FreeCashFlow

Therefore the model we have used is as follows:-

$$y = \beta_0 + \beta_1 x1 + \beta_2 x2 + \ldots\ldots\ldots + \beta_{23} x23$$

Based on this we have conducted AIC and BIC tests to find the most significant variables out of these 23 variables that affect the price.

AIC:

```
> summary(lm_AIC)

Call:
lm(formula = y ~ x2 + x5 + x8 + x14 + x17 + x18 + x19 + x22,
    data = target)

Residuals:
    Min      1Q  Median      3Q     Max 
-36.715  -8.040   2.224   6.739  47.860 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  5.556e+02  2.505e+02   2.218  0.03014 *  
x2          -4.694e-01  2.271e-01  -2.067  0.04284 *  
x5           2.664e+01  4.599e+00   5.791 2.40e-07 ***
x8           4.029e-09  9.025e-10   4.465 3.39e-05 ***
x14          1.661e-09  9.501e-10   1.748  0.08530 .  
x17         -1.183e-08  2.845e-09  -4.157 9.92e-05 ***
x18          1.652e-08  3.011e-09   5.486 7.77e-07 ***
x19         -9.982e-07  3.512e-07  -2.843  0.00602 ** 
x22         -1.996e-08  1.045e-08  -1.909  0.06077 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.47 on 63 degrees of freedom
Multiple R-squared:  0.947,	Adjusted R-squared:  0.9402 
F-statistic: 140.6 on 8 and 63 DF,  p-value: < 2.2e-16
```

BIC:

```
> summary(lm_BIC)

Call:
lm(formula = y ~ x2 + x5 + x8 + x17 + x18 + x19 + x22, data = target)

Residuals:
    Min      1Q  Median      3Q     Max 
-37.336  -7.635   1.963   7.432  46.909 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  8.483e+02  1.893e+02   4.481 3.14e-05 ***
x2          -5.240e-01  2.285e-01  -2.293  0.02514 *  
x5           2.557e+01  4.631e+00   5.521 6.56e-07 ***
x8           4.323e-09  9.008e-10   4.799 9.93e-06 ***
x17         -1.349e-08  2.723e-09  -4.954 5.60e-06 ***
x18          1.936e-08  2.575e-09   7.517 2.29e-10 ***
x19         -1.394e-06  2.729e-07  -5.106 3.18e-06 ***
x22         -2.717e-08  9.757e-09  -2.785  0.00704 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.69 on 64 degrees of freedom
Multiple R-squared:  0.9444,	Adjusted R-squared:  0.9383 
F-statistic: 155.3 on 7 and 64 DF,  p-value: < 2.2e-16
```

Since, the F-stat value for BIC is greatest the BIC model parameters are finalised.

```
> model_BIC <- lm(y ~ x2 +  x5 + x8  + x17 + x18 + x19 + x22, data = target)
> summary(model_BIC)

Call:
lm(formula = y ~ x2 + x5 + x8 + x17 + x18 + x19 + x22, data = target)

Residuals:
    Min      1Q  Median      3Q     Max 
-37.336  -7.635   1.963   7.432  46.909 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  8.483e+02  1.893e+02   4.481 3.14e-05 ***
x2          -5.240e-01  2.285e-01  -2.293  0.02514 *  
x5           2.557e+01  4.631e+00   5.521 6.56e-07 ***
x8           4.323e-09  9.008e-10   4.799 9.93e-06 ***
x17         -1.349e-08  2.723e-09  -4.954 5.60e-06 ***
x18          1.936e-08  2.575e-09   7.517 2.29e-10 ***
x19         -1.394e-06  2.729e-07  -5.106 3.18e-06 ***
x22         -2.717e-08  9.757e-09  -2.785  0.00704 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.69 on 64 degrees of freedom
Multiple R-squared:  0.9444,	Adjusted R-squared:  0.9383 
F-statistic: 155.3 on 7 and 64 DF,  p-value: < 2.2e-16
```

Gauss Markov Assumptions

A. Auto-correlation

For autocorrelation we have conducted the Durbin Watson test.

```
> durbinWatsonTest(model_2)
 lag Autocorrelation D-W Statistic p-value
   1     0.007884658      1.937491   0.346
 Alternative hypothesis: rho != 0
```

The p-value is greater than the significance level, therefore, we do not reject the null hypothesis for Durbin Watson. There is enough evidence that there is no presence of autocorrelation.

B. Conditional Heteroskedasticity

We have conducted Breusch-Pagan test to check for conditional heteroskedasticity.

```
> bptest(model_2)

	studentized Breusch-Pagan test

data:  model_2
BP = 24.745, df = 7, p-value = 0.000842
```

The p-value is less than the significance level, therefore, we reject the null hypothesis for Breusch-Pagan. There is enough evidence that there is conditional heteroskedasticity.

C. Multicollinearity

For Multicollinearity we have conducted the VIF test.

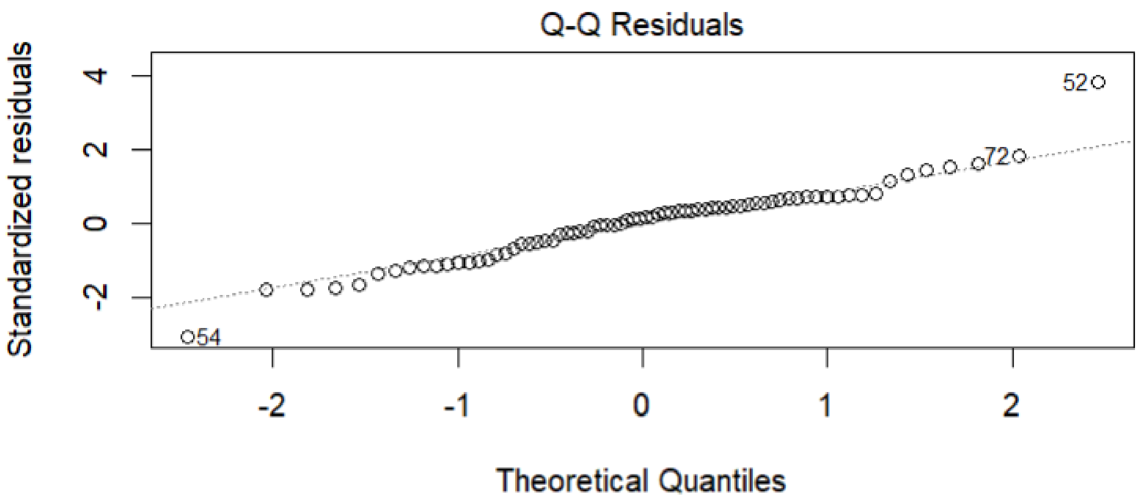
```
> vif(model_BIC)
       x2        x5        x8       x17       x18       x19       x22 
 1.196694  5.106358  3.339344  9.803488  6.741036 18.373373  4.374210 
```

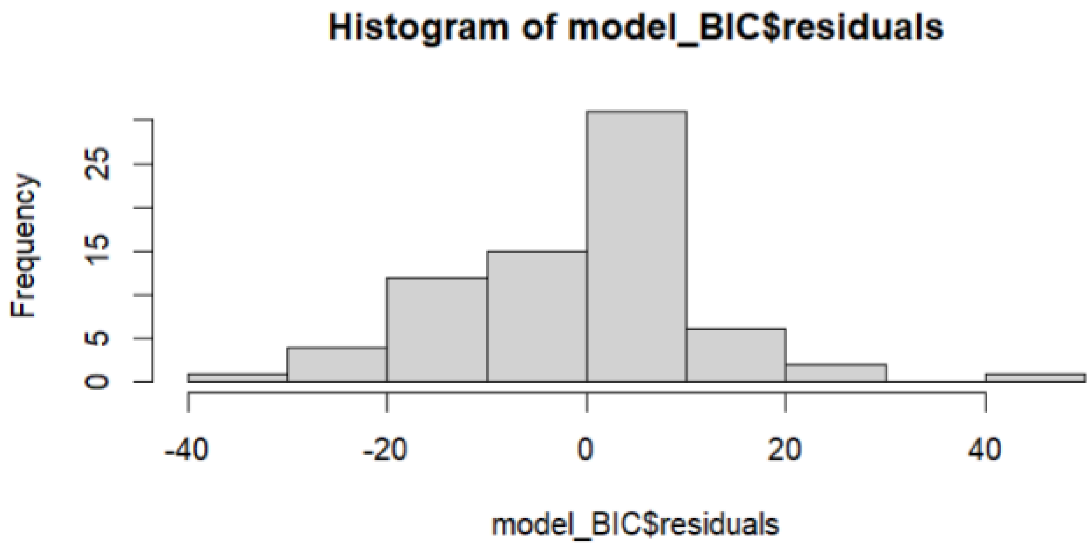
We can observe that a lot of variables have higher values which signify that these variables are collinear with the share price (y). Higher the value better will be the stock prediction.

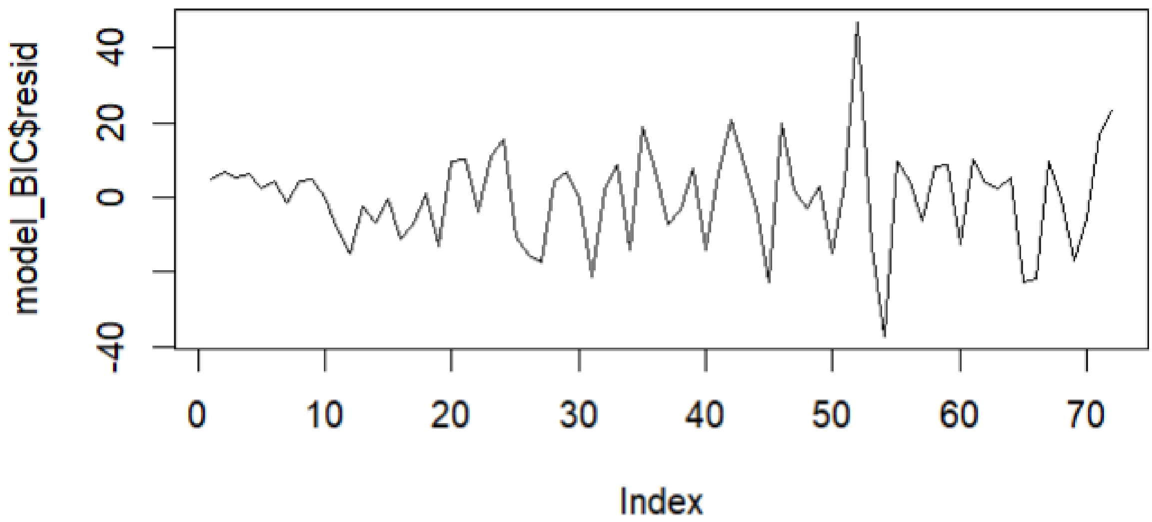
Following is the share price output for testing the model fitness with respect to the model we developed using only the significant variables.

```
> fitted(model_BIC)
        1         2         3         4         5         6         7         8         9        10        11 
 70.61673  68.67793  63.96194  66.39303  70.65314  71.71686  81.93682  83.22062  83.21538  83.51532  79.66917 
       12        13        14        15        16        17        18        19        20        21        22 
 81.18876  75.55499  79.26491  80.61159  88.35183  87.61823  85.49079  99.55446  97.55802  96.62006 110.89142 
       23        24        25        26        27        28        29        30        31        32        33 
114.17166 112.80402 121.53740 118.65541 110.55962 105.20388 115.61574 120.31077 147.28817 149.56233 148.59817 
       34        35        36        37        38        39        40        41        42        43        44 
166.08911 160.62380 169.19643 188.44647 186.73299 190.34334 221.23720 220.72368 220.93852 252.01154 249.92603 
       45        46        47        48        49        50        51        52        53        54        55 
251.68142 239.64076 242.09832 234.39029 217.37568 214.56704 208.61441 181.74135 175.08132 178.56591 153.91101 
       56        57        58        59        60        61        62        63        64        65        66 
155.93365 154.50837 156.11537 158.36333 161.49161 162.22927 164.86498 163.44495 152.40303 153.69731 153.92263 
       67        68        69        70        71        72 
127.06306 126.87442 127.31458 116.58678 116.97454 119.19629 
```

D. Normality of Residuals







The first plot gives us information that there are a few outliers for residual plots. The latter two plots show us that the residual values have magnitudes a bit far from zero due to these outliers.

Residual Analysis

Ljung-Box test for residuals

```
> Box.test(model_BIC$resid,lag=10,type='Ljung')

        Box-Ljung test

data:  model_BIC$resid
X-squared = 21.623, df = 10, p-value = 0.01715
```

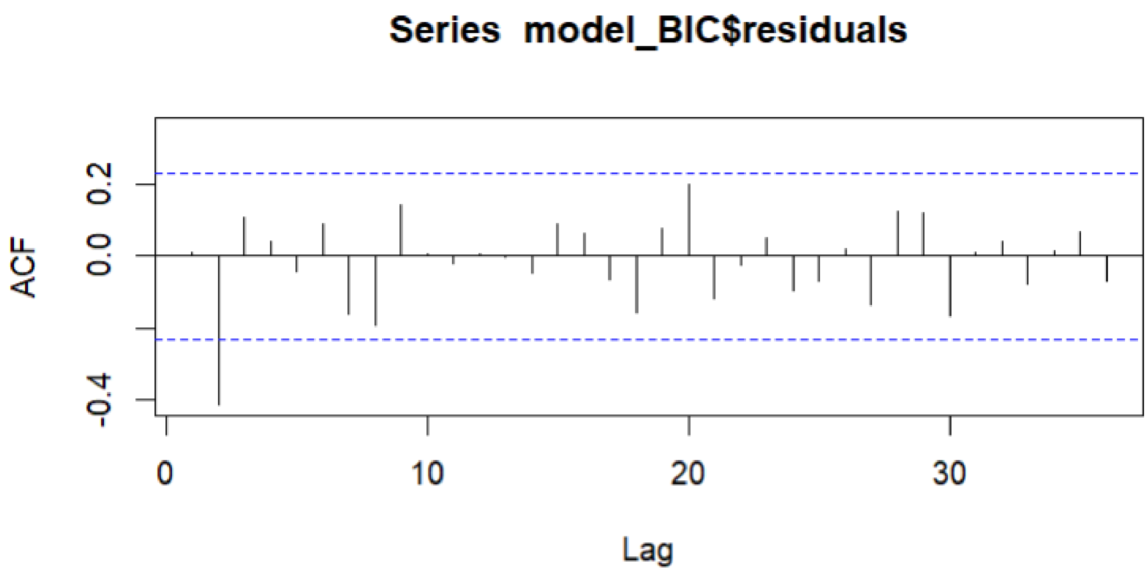
The p-value is less than the significance level, therefore, we reject the null hypothesis for the Ljung-Box test. There is enough evidence that there is autocorrelation in residuals.

Ljung-Box test for square of residuals

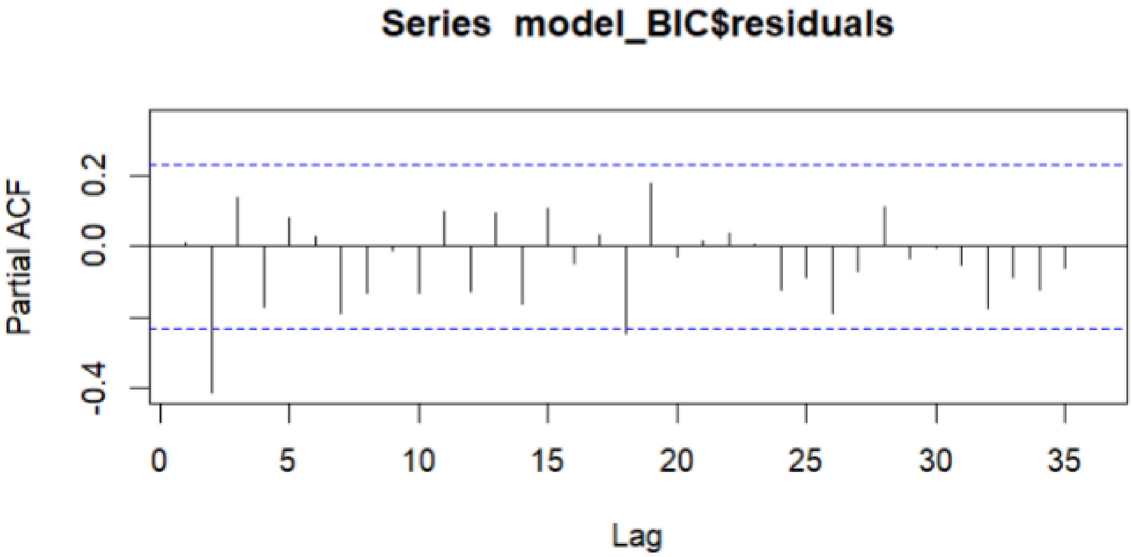
```
> Box.test(k^2,lag=10,type='Ljung')

        Box-Ljung test

data:  k^2
X-squared = 10.231, df = 10, p-value = 0.4205
```

The p-value is greater than the significance level, therefore, we do not reject the null hypothesis for the Ljung-Box test. The volatility clustering is insignificant for significance level 5%. The Garch model is therefore not needed.

ACF plot

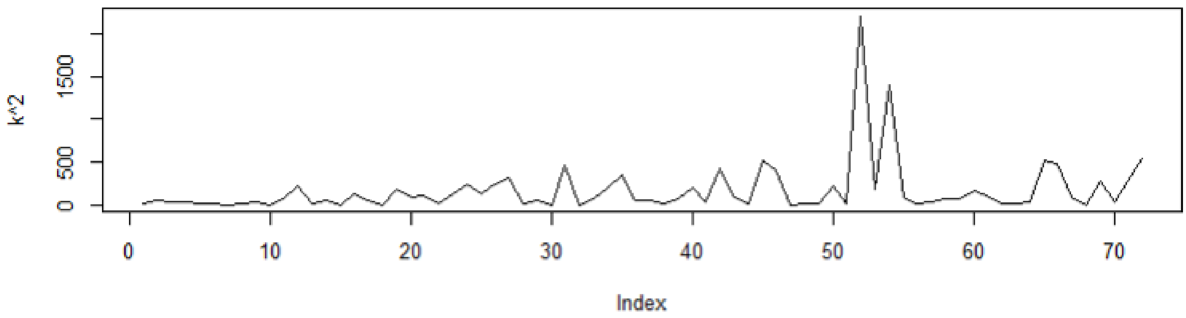


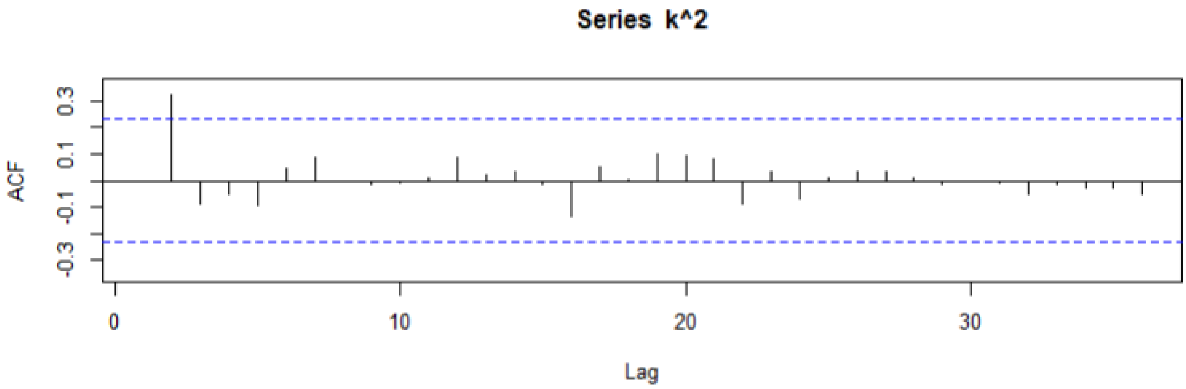
PACF plot

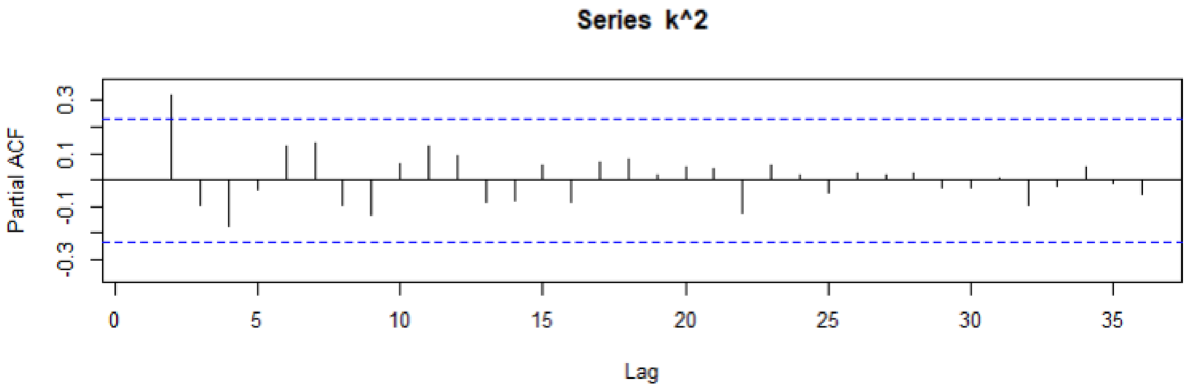


We can see from the plots that there is an intersection of the threshold line which signifies that there is a need for an ARIMA model. Since, the intersection happens at 2nd lag we analysed the ARIMA model to be (2,0,2).

Residual Analysis for volatility clustering setup







Stationarity

Augmented Dickey Fuller test is conducted to check for stationarity of the series.

```
> adf.test(model_BIC$residuals)

	Augmented Dickey-Fuller Test

data:  model_BIC$residuals
Dickey-Fuller = -3.9184, Lag order = 4, p-value = 0.01835
alternative hypothesis: stationary
```

The p-value is less than the significance level, therefore, we reject the null hypothesis for the ADF test. There is enough evidence that the series is stationary.

ARIMA Model

ARIMA(2,0,2) model is required.

```
> m2 <- arima(model_BIC$residuals, order=c(2,0,2))
> m2

Call:
arima(x = model_BIC$residuals, order = c(2, 0, 2))

Coefficients:
         ar1     ar2     ma1      ma2  intercept
      0.0151  0.2589  0.0888  -0.7921    -0.3396
s.e.  0.3433  0.3246  0.2820   0.2864     0.5789

sigma^2 estimated as 126.1:  log likelihood = -277,  aic = 566.01
```

Based on the estimates and their respective standard error values we can say that ar1 and ma1 are insignificant therefore we will remove those values and re-run ARIMA.

```
> c1 <- c(0,NA,0,NA,NA)
> m1 <- arima(model_BIC$residuals, order=c(2,0,2), fixed = c1)
Warning message:
In arima(model_BIC$residuals, order = c(2, 0, 2), fixed = c1) :
  some AR parameters were fixed: setting transform.pars = FALSE
> m1

Call:
arima(x = model_BIC$residuals, order = c(2, 0, 2), fixed = c1)

Coefficients:
      ar1     ar2  ma1      ma2  intercept
        0  0.4060    0  -0.9063    -0.2998
s.e.    0  0.2046    0   0.1570     0.2972

sigma^2 estimated as 127:  log likelihood = -277.53,  aic = 563.05
```

We can now take this model with significant variables and further predict the value.

```
>  predict(m1,2)
$pred
Time Series:
Start = 73
End = 74
Frequency = 1
[1] -2.183617 -5.576020

$se
Time Series:
Start = 73
End = 74
Frequency = 1
[1] 11.26796 11.26796
```

Therefore, we have our prediction as -2.1836 which we will take into consideration when we will predict our final stock price for 2024 using the DCF model.

Garch Model testing
For Garch models the forecasted values are very small and insignificant. Therefore the Garch modelling is neglected for Target as well.

Discounted Cash Flow Model
This financial model was constructed to predict the values for Target for the next five years. The values of the variables we have for our regression model is taken from this financial model to predict for 2024.

INCOME STATEMENTS

	Historical			Projected				
	2020	**2021**	**2022**	**2023**	**2024**	**2025**	**2026**	**2027**
Revenue	93,561.0	106,005.0	109,120.0	107,412.0	116,669.2	123,462.9	128,344.2	133,744.3
Less: Total COGS	(68,407.0)	(77,307.0)	(84,614.0)	(80,151.0)	(86,978.3)	(92,486.4)	(96,779.2)	(100,136.9)
Gross Profit	25,154.0	28,698.0	24,506.0	27,261.0	29,690.9	30,976.5	31,565.0	33,607.4
Less: Total SG&A	(18,631.0)	(19,370.0)	(20,610.0)	(21,462.0)	(22,499.5)	(23,587.1)	(24,727.3)	(25,922.6)
EBIT	6,523.0	9,328.0	3,896.0	5,799.0	7,191.4	7,389.4	6,837.7	7,684.8
Interest & Other Expense / (Income):								
Total Interest Expense	977.0	421.0	478.0	502.0	505.1	507.7	509.7	511.2
Less: Interest Income					0.0	0.0	0.0	0.0
Pretax Income	5,546.0	8,907.0	3,418.0	5,297.0	6,686.3	6,881.7	6,328.1	7,173.5
Less: Income Taxes	(1,178.0)	(1,961.0)	(638.0)	(1,159.0)	(1,463.0)	(1,505.7)	(1,384.6)	(1,569.6)
Net Income	4,368.0	6,946.0	2,780.0	4,138.0	5,223.3	5,375.9	4,943.5	5,603.9
Shares Outstanding	500.6	488.1	462.1	461.5	461.5	461.5	461.5	461.5
Earnings per Share (Basic EPS)	$8.73	$14.23	$6.02	$8.97	$11.3	$11.6	$10.7	$12.1
EBITDA Reconciliation:								
EBIT	6,523.0	9,328.0	3,896.0	5,799.0	7,191.4	7,389.4	6,837.7	7,684.8
Plus: Depreciation	2,485.0	2,642.0	2,700.0	2,801.0	2,983.9	3,127.3	3,264.0	3,424.3
Plus: Amortization	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
EBITDA	9,008.0	11,970.0	6,596.0	8,600.0	10,175.4	10,516.7	10,101.7	11,109.1

BALANCE SHEETS

	Historical				Projected			
	2020.00	2021	2022	2023	2024	2025	2026	2027
ASSETS:								
Required Cash	8,511.0	5,911.0	$2,229.0	3,805.0	2,291.3	2,436.4	2,549.5	2,637.9
Excess Cash	0.0	0.0	0	0.0	4,538.1	7,947.7	10,812.3	14,204.5
Total Cash	$8,511.0	$5,911.0	$2,229.0	3,805.0	6,829.3	10,384.1	13,361.8	16,842.5
Accounts Receivable	1,135.0	1,353.0	1,695.0	1,404.0	1,812.3	1,917.8	1,993.6	2,077.5
Inventory	10,653.0	13,902.0	13,499.0	11,886.0	13,876.2	14,754.9	15,439.8	15,975.5
Other Current Assets	286.0	237.0	235.0	202.0	219.2	233.1	243.9	252.4
Prepaid Expenses	171.0	170.0	188.0	201.0	193.3	205.5	215.0	222.5
Current Assets	20,756.0	21,573.0	17,846.0	17,498.0	22,930.3	27,495.4	31,254.2	35,370.3
PP&E - Gross	49,384.0	51,874.0	56,800.0	60,871.0	65,881.4	71,183.5	76,695.3	82,439.0
Less: Accum. Depn.	(20,278.0)	(21,137.0)	(22,631.0)	(24,413.0)	(27,396.9)	(30,524.2)	(33,788.3)	(37,212.6)
Net PP&E	29,106.0	30,737.0	34,169.0	36,458.0	38,484.5	40,659.3	42,907.1	45,226.4
Intangibles	668.0	656.0	645.0	639.0	633.1	627.2	621.3	615.6
Defined Pension Benefit	450.0	470.0	515.0	540.0	586.0	623.1	652.0	674.7
Other Non Current Assets	268.0	375.0	160.0	221.0	239.8	255.0	266.8	276.1
Cap. Financing Costs								
Total Assets	$51,248.0	$53,811.0	$53,335.0	$55,356.0	$62,873.6	######	$75,701.4	$82,163.0
LIABILITIES & EQUITY:								
Accounts Payable	14,303.0	16,944.0	14,756.0	13,433.0	15,168.3	16,128.9	16,877.5	17,463.1
Current Accrued Expenses	1,756.0	1,697.0	1,413.0	1,657.0	1,798.1	1,912.0	2,000.8	2,070.2
Pension and Other Post Retireme	169.0	169.0	173.0	192.0	208.4	221.5	231.8	239.9
Current Debt And Capital Lease I	1,355.0	425.0	426.0	1,445.0	1,568.1	1,667.4	1,744.8	1,805.3
Other Current Liabilities	2,542.0	2,512.0	2,732.0	2,577.0	2,796.5	2,973.6	3,111.6	3,219.6
Current Liabilities	20,125.0	21,747.0	19,500.0	19,304.0	21,539.4	22,903.5	23,966.5	24,798.0
Long Term Debt And Capital Lea	13,754.0	16,042.0	18,647.0	18,201.0	18,201.0	18,201.0	18,201.0	18,201.0
Non Current Deferred Liabilities	2,048.0	2,617.0	3,195.0	3,475.0	3,475.0	3,475.0	3,475.0	3,475.0
Trade and Other Payables Non C	436.0	139.0	168.0	272.0	272.0	272.0	272.0	272.0
Employee Benefits	398.0	395.0	424.0	491.0	532.8	566.6	592.9	613.4
Other Non Current Liabilities	47.0	44.0	169.0	181.0	196.4	208.9	218.6	226.1
Total Debt	16,683.0	19,237.0	22,603.0	22,620.0	22,677.2	22,723.4	22,759.4	22,787.6
Total Liabilities	36,808.0	40,984.0	42,103.0	41,924.0	44,216.7	45,626.9	46,725.9	47,585.6
Common Equity	14,440.0	12,827.0	11,232.0	13,432.0	18,655.3	24,031.2	28,974.7	34,578.7
Liabilities & Equity	$51,248.0	$53,811.0	$53,335.0	$55,356.0	$62,872.0	$69,658.1	$75,700.6	$82,164.2
Check	0.000	0.000	0.000	0.000	1.637	1.864	0.797	1.255

RESULTS

Target Stock Prediction (2028-2023)

We have therefore developed from the significant variables keeping in mind the effects of ARCH and GARCH. Using the predicted values for those variables from the DCF model we will now predict the price for the following year.

Following is the our model and the values we need to use for the prediction:-

$$y \sim x2 + x5 + x8 + x17 + x18 + x19 + x22$$

y=Share price

x2 = Volatility Index) (VIX) = 20.18

x5 =EV Revenue Ratio= 2.821

x8 = Cost of Revenue= 21744581584

x17 = Long term debt= 14883000000

x18 = Working capital= -3147211865
x19 = Shares outstanding= 461651176
x22 = Capital Expenditure= -1127000000

```
> x <-predict(model_BIC, newdata=new)
> x
       1
129.2792
>
> # Adjustment with ARIMA prediction values for next year for 2024
> price <- x  - 2.1836
> price
       1
127.0956
```

RECOMMENDATIONS

This Forecast is based on 2024 annual statement forecasts. Based on our chosen model and the new data obtained from the values of 2024 Financial model, we get a stock price prediction of *127.0956* for 2024. Current stock price is *164.44*. Hence, we conclude that the stock is overpriced and provide a recommendation of going in a *short position* for Target stock.

INTERTEMPORAL ANALYSIS

We now start with an intertemporal analysis, focusing on the same stock but different time frame and then inspecting the significance of the same variables over those periods. We examine *Target*, with this time a timeframe of 2012 to 2018.

Regression Analysis

We have considered the same 23 variables for the multivariate regression for the share price of Target as we did for the time frame 2018 to 2023.

The model we have used is as follows:-

$$y = \beta_0 + \beta_1 x1 + \beta_2 x2 + \ldots\ldots\ldots \beta_{23} x23$$

Based on this we have conducted AIC and BIC tests to find the most significant variables out of these 23 variables that affect the price.

AIC:

```
> summary(lm_AIC)

Call:
lm(formula = y ~ x1 + x4 + x5 + x6 + x7 + x14 + x15 + x16 + x17 +
    x18 + x19 + x20 + x21 + x22, data = target_old)

Residuals:
    Min      1Q  Median      3Q     Max
-8.8210 -2.2031 -0.2146  1.8887  7.0466

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -9.599e+01  2.119e+01  -4.530 3.06e-05 ***
x1           2.410e+02  1.866e+02   1.292  0.20169
x4           1.041e+00  2.331e-01   4.468 3.79e-05 ***
x5           1.546e+01  2.758e+00   5.606 6.29e-07 ***
x6          -2.779e-01  2.065e-01  -1.346  0.18366
x7           3.802e-09  4.873e-10   7.802 1.47e-10 ***
x14         -3.960e-09  6.930e-10  -5.715 4.20e-07 ***
x15          1.128e-09  7.444e-10   1.516  0.13503
x16          6.607e-09  1.021e-09   6.472 2.40e-08 ***
x17          3.158e-09  1.145e-09   2.758  0.00780 **
x18          1.515e-09  8.946e-10   1.693  0.09593 .
x19          1.650e-07  5.470e-08   3.017  0.00381 **
x20         -7.592e-09  3.011e-09  -2.521  0.01452 *
x21          5.425e-09  2.792e-09   1.943  0.05697 .
x22         -1.079e-08  4.220e-09  -2.557  0.01325 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.403 on 57 degrees of freedom
Multiple R-squared:  0.8688,    Adjusted R-squared:  0.8366
F-statistic: 26.96 on 14 and 57 DF,  p-value: < 2.2e-16
```

BIC:

```
> summary(lm_BIC)

Call:
lm(formula = y ~ x4 + x5 + x7 + x14 + x16 + x17 + x18 + x19 +
    x20 + x21 + x22, data = target_old)

Residuals:
    Min      1Q  Median      3Q     Max
-7.6888 -2.0205 -0.1613  2.1513  8.8430

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -6.273e+01  1.548e+01  -4.053 0.000148 ***
x4           1.135e+00  2.299e-01   4.938 6.63e-06 ***
x5           1.564e+01  2.464e+00   6.349 3.19e-08 ***
x7           4.194e-09  4.735e-10   8.857 1.72e-12 ***
x14         -3.426e-09  6.749e-10  -5.077 3.99e-06 ***
x16          4.971e-09  6.734e-10   7.383 5.56e-10 ***
x17          2.022e-09  1.018e-09   1.985 0.051740 .
x18          2.368e-09  7.386e-10   3.206 0.002158 **
x19          9.538e-08  4.643e-08   2.054 0.044311 *
x20         -7.247e-09  2.957e-09  -2.451 0.017171 *
x21          5.514e-09  2.651e-09   2.080 0.041802 *
x22         -8.191e-09  4.201e-09  -1.950 0.055913 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.496 on 60 degrees of freedom
Multiple R-squared:  0.8543,    Adjusted R-squared:  0.8276
F-statistic: 31.98 on 11 and 60 DF,  p-value: < 2.2e-16
```

Since, the F-stat value for BIC is greatest the BIC model parameters are finalised.

```
> model_new <- lm(y ~ x4 + x5 + x7+ x14  + x16 + x17 + x18 + x19 + x20+x21+ x22, data =target_old)
> summary(model_new)

Call:
lm(formula = y ~ x4 + x5 + x7 + x14 + x16 + x17 + x18 + x19 +
    x20 + x21 + x22, data = target_old)

Residuals:
    Min      1Q  Median      3Q     Max
-7.6888 -2.0205 -0.1613  2.1513  8.8430

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -6.273e+01  1.548e+01  -4.053 0.000148 ***
x4           1.135e+00  2.299e-01   4.938 6.63e-06 ***
x5           1.564e+01  2.464e+00   6.349 3.19e-08 ***
x7           4.194e-09  4.735e-10   8.857 1.72e-12 ***
x14         -3.426e-09  6.749e-10  -5.077 3.99e-06 ***
x16          4.971e-09  6.734e-10   7.383 5.56e-10 ***
x17          2.022e-09  1.018e-09   1.985 0.051740 .
x18          2.368e-09  7.386e-10   3.206 0.002158 **
x19          9.538e-08  4.643e-08   2.054 0.044311 *
x20         -7.247e-09  2.957e-09  -2.451 0.017171 *
x21          5.514e-09  2.651e-09   2.080 0.041802 *
x22         -8.191e-09  4.201e-09  -1.950 0.055913 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.496 on 60 degrees of freedom
Multiple R-squared:  0.8543,    Adjusted R-squared:  0.8276
F-statistic: 31.98 on 11 and 60 DF,  p-value: < 2.2e-16
```

Gauss Markov Assumptions

A. Auto-correlation

For autocorrelation we have conducted the Durbin Watson test.

```
> durbinWatsonTest(model_new)
 lag Autocorrelation D-W Statistic p-value
   1      0.06081186      1.849698   0.056
 Alternative hypothesis: rho != 0
```

The p-value is greater than the significance level, therefore, we do not reject the null hypothesis for Durbin Watson. There is enough evidence that there is no presence of autocorrelation.

B. Conditional Heteroskedasticity

We have conducted Breusch-Pagan test to check for conditional heteroskedasticity.

```
> bptest(model_new)

	studentized Breusch-Pagan test

data:  model_new
BP = 15.223, df = 11, p-value = 0.1725
```

The p-value is greater than the significance level, therefore, we do not reject the null hypothesis for Breusch-Pagan. There is enough evidence that there is no conditional heteroskedasticity.

C. Multicollinearity

For Multicollinearity we have conducted the VIF test.

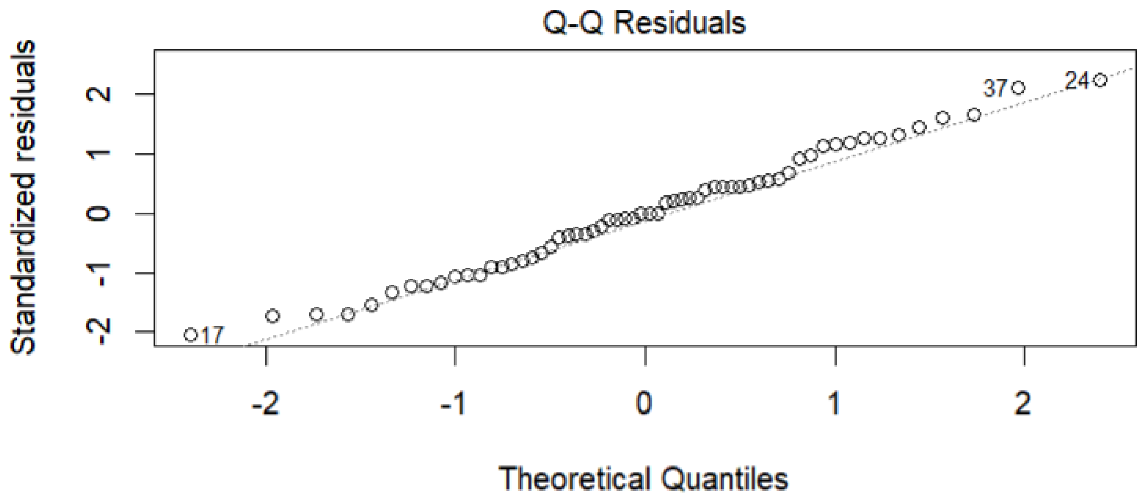
```
> vif(model_new) # Fir stock prediction higher the better
       x4        x5        x7       x14       x16       x17       x18       x19       x20       x21
 1.838996  7.078206  5.435683 36.936587  2.492752  9.589692  7.176471 18.642051 26.249774 18.477006
      x22
 5.723036
```

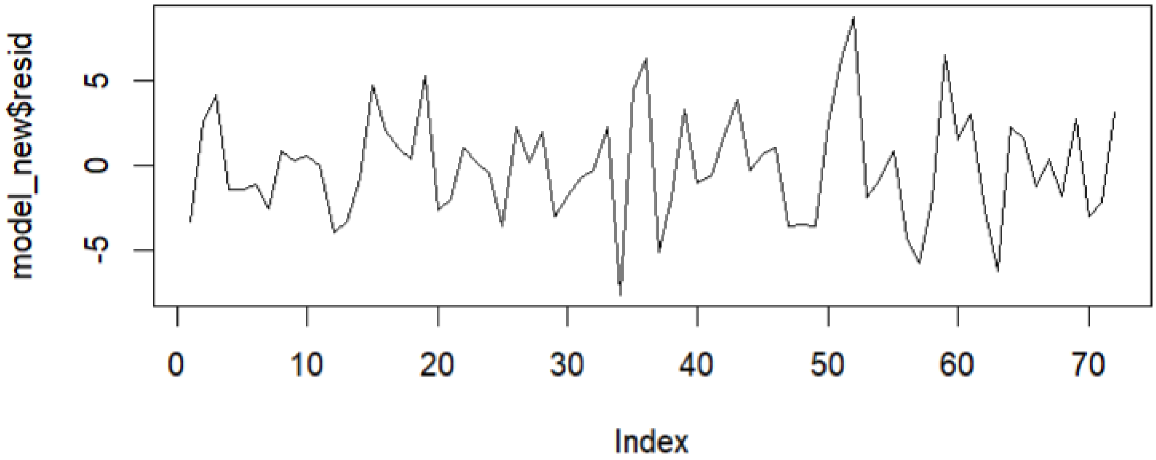
We can observe that a lot of variables have higher values which signify that these variables are collinear with the share price (y). Higher the value better will be the stock prediction.

Following is the share price output for testing the model fitness with respect to the model we developed using only the significant variables.

```
> fitted(model_new)
       1        2        3        4        5        6        7        8        9       10       11 
54.12883 54.12883 54.12883 59.35860 59.35860 59.35860 63.18346 63.18346 63.18346 63.16352 63.16352 
      12       13       14       15       16       17       18       19       20       21       22 
63.16352 63.70299 63.70299 63.70299 68.44475 68.44475 68.44475 65.94539 65.94539 65.94539 63.73039 
      23       24       25       26       27       28       29       30       31       32       33 
63.73039 63.73039 60.28044 60.28044 60.28044 59.74798 59.74798 59.74798 60.38232 60.38232 60.38232 
      34       35       36       37       38       39       40       41       42       43       44 
69.50881 69.50881 69.50881 78.73266 78.73266 78.73266 79.84508 79.84508 79.84508 77.99910 77.99910 
      45       46       47       48       49       50       51       52       53       54       55 
77.99910 76.09141 76.09141 76.09141 76.02116 76.02116 76.02116 70.65695 70.65695 70.65695 74.45733 
      56       57       58       59       60       61       62       63       64       65       66 
74.45733 74.45733 70.67125 70.67125 70.67125 61.41281 61.41281 61.41281 53.54913 53.54913 53.54913 
      67       68       69       70       71       72 
56.27976 56.27976 56.27976 62.08589 62.08589 62.08589 
```

D. Normality of residuals





The first plot gives us information that there aren't any significant outliers for residual plot. The latter two plots show us that the residual values have magnitudes not that far off to zero.

Residual Analysis

Ljung-Box test for residuals

```
> Box.test(model_new$resid,lag=10,type='Ljung')

        Box-Ljung test

data:  model_new$resid
X-squared = 18.955, df = 10, p-value = 0.04084
```

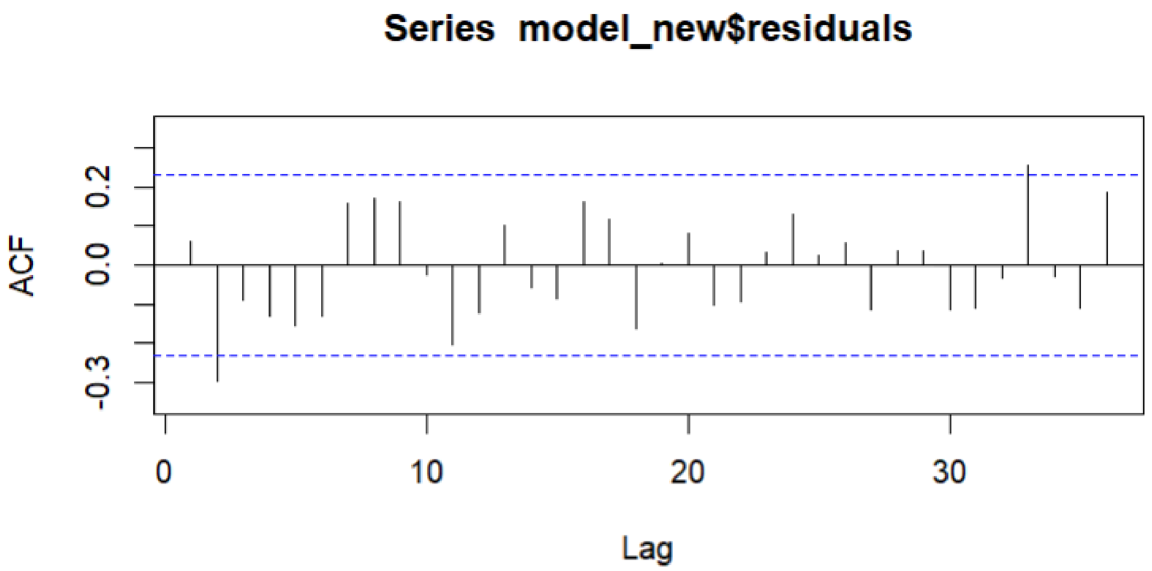
The p-value is less than the significance level, therefore, we reject the null hypothesis for the Ljung-Box test. There is enough evidence that there is autocorrelation in residuals.

Ljung-Box test for square of residuals

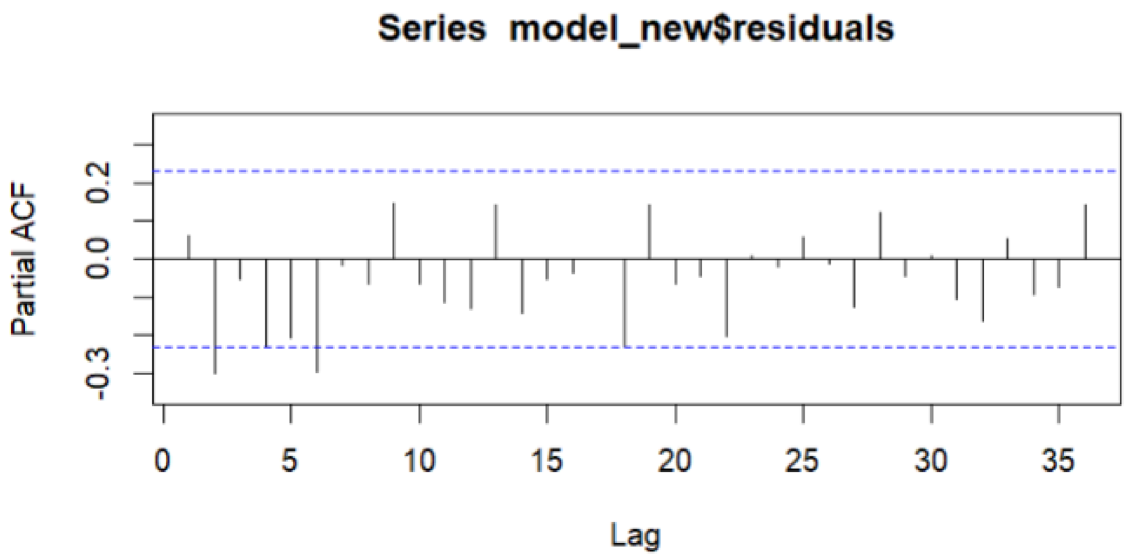
```
> Box.test(k^2,lag=10,type='Ljung')

        Box-Ljung test

data:  k^2
X-squared = 6.3372, df = 10, p-value = 0.7862
```

The p-value is greater than the significance level, therefore, we do not reject the null hypothesis for the Ljung-Box test. The volatility clustering is insignificant for significance level 5%.

ACF plot

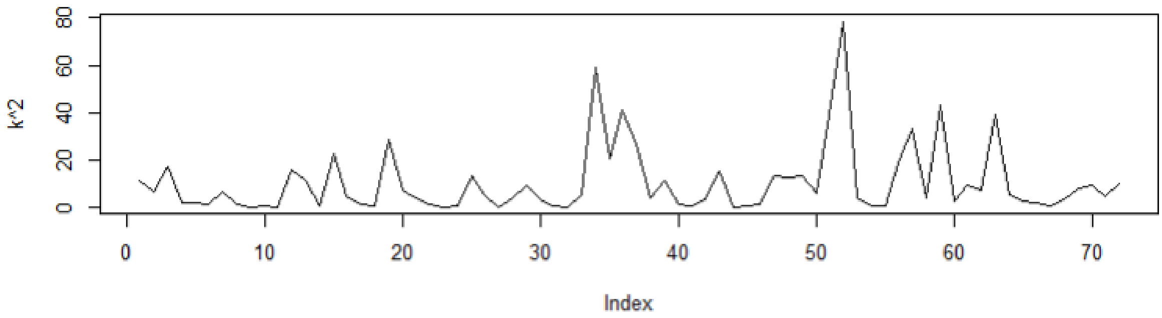


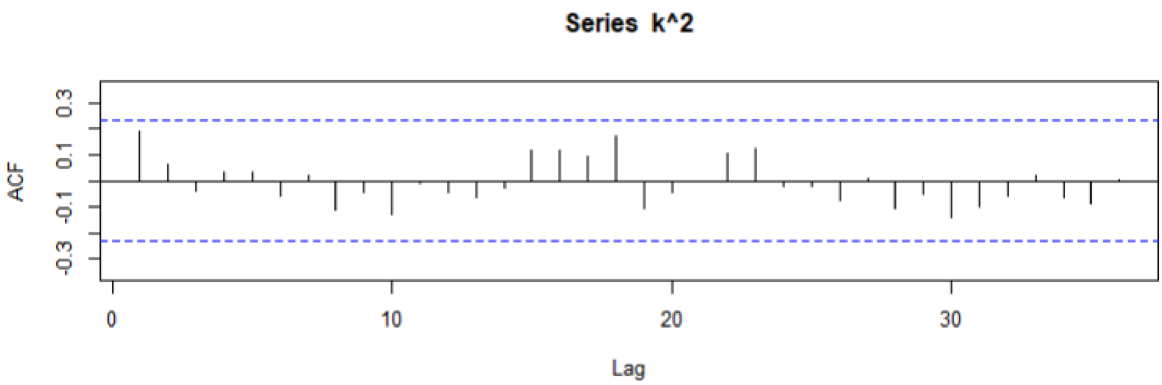
PACF plot

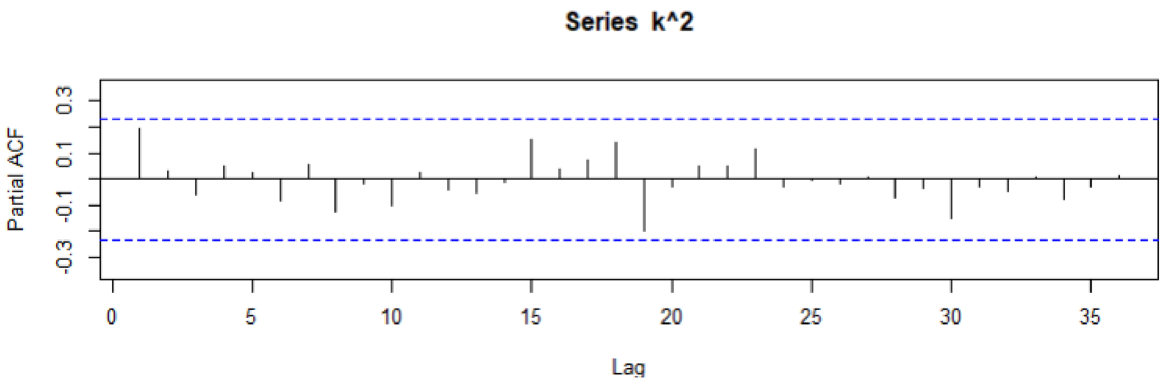


We can see from the plots that there is an intersection of the threshold line which signifies that there is a need for an ARIMA model. Since, the intersection happens at 2nd lag we analysed the ARIMA model to be (2,0,2).

Residual Analysis for volatility clustering setup







Stationarity

Augmented Dickey Fuller test is conducted to check for stationarity of the series.

```
> adf.test(model_new$residuals)

        Augmented Dickey-Fuller Test

data:  model_new$residuals
Dickey-Fuller = -5.7427, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

The p-value is less than the significance level, therefore, we reject the null hypothesis for the ADF test. There is enough evidence that the series is stationary.

ARIMA Model testing

Command is passed to check for the best ARIMA model required.

```
> auto.arima(model_new$residuals, d = 0)
Series: model_new$residuals
ARIMA(2,0,2) with zero mean

Coefficients:
          ar1     ar2      ma1      ma2
      -0.0210  0.3648  -0.0432  -0.9184
s.e.   0.1374  0.1364   0.0926   0.0924

sigma^2 = 7.833:  log likelihood = -175.53
AIC=361.05   AICc=361.96   BIC=372.44
```

We can observe that there is a need to incorporate the ARIMA model which complements our decision to consider ARIMA (2,0,2) from the ACF and PACF plots.

ARIMA Model

ARIMA(2,0,2) model is required.

```
> m1 <- arima(k, order=c(2,0,2))
> m1

Call:
arima(x = k, order = c(2, 0, 2))

Coefficients:
          ar1     ar2      ma1      ma2  intercept
      -0.0392  0.3450  -0.0638  -0.9362     0.0446
s.e.   0.1301  0.1273   0.0712   0.0697     0.0392

sigma^2 estimated as 7.164:  log likelihood = -174.95,  aic = 361.9
```

Based on the estimates and their respective standard error values we can say that ar1 and ma1 are insignificant therefore we will remove those values and re-run ARIMA.

```
> c1 = c(0,NA,0,NA,NA)
> m2 <- arima(model_new$residuals, order=c(2,0,2), fixed = c1)
Warning message:
In arima(model_new$residuals, order = c(2, 0, 2), fixed = c1) :
  some AR parameters were fixed: setting transform.pars = FALSE
> m2

Call:
arima(x = model_new$residuals, order = c(2, 0, 2), fixed = c1)

Coefficients:
      ar1     ar2  ma1      ma2  intercept
        0  0.3664    0  -0.9151     0.0385
s.e.    0  0.1332    0   0.0744     0.0631

sigma^2 estimated as 7.459:  log likelihood = -175.64,  aic = 359.29
```

We can now take this model with significant variables and further predict the value.

```
> predict(m2,2)
$pred
Time Series:
Start = 73
End = 74
Frequency = 1
[1] -0.4907606  1.0487321

$se
Time Series:
Start = 73
End = 74
Frequency = 1
[1] 2.731318 2.731318
```

Therefore, we have our prediction as *-0.491* which we will take into consideration when we will predict our final stock price for 2024 using the DCF model values.

Garch Model testing
For the Garch model the forecasted values are very small and insignificant. Therefore the Garch modelling is neglected for Vale.

RESULTS
Target Stock Prediction (2012-2018)
We have therefore developed from the significant variables keeping in mind the effects of ARCH and GARCH. Using the predicted values for those variables from the DCF model we will now predict the price for the year 2019 and verify if our model is working properly.
Following is the our model and the values we need to use for the prediction:-

$$\mathbf{y \sim x4 + x5 + x7 + x14 + x16 + x17 + x18 + x19 + x20 + x21 + x22}$$

y=Share price
x4= Walmart stock price = 39.086666
x5= EV Revenue Ratio=3.568
x7= Total Revenue= 18665000000
x14= Total Assets = 43741000000
x16= Inventory= 11396000000
x17= Long term debt= 10513000000
x18= Working capital= -2803000000
x19= shares outstanding= 506677740
x20= Operating cash flow= 1347000000
x21= Change in working capital= -138000000
x22= Capital expenditure= -1009000000

```
> x <-predict(model_new, newdata=new)
> x # 83.2222
       1
83.22219
>
> price <- x - 0.491
> price
       1
82.73119
```

RECOMMENDATIONS

This Forecast is based on the 2019 annual statement. Based on our chosen model and the new data obtained from the financial statements of 2019, we get a stock price prediction of *82.73119* for 2019. Stock price at the end of 2019 was *around 66*. Hence, we conclude that it is underpriced and provide a recommendation of holding a *long position* for Target stock for 2019. If we look at the 2019 returns the Target stock reaches *over 80* within 6 months thereby confirming our buy call for the stock and verifying our developed model.

CONCLUSIONS

Intersectoral Analysis

Model equations and variables for both stocks (VALE and TARGET) for time period 2018-2023 are as follows:-

VALE

$$y \sim x2 + x4 + x5 + x9 + x11 + x12 + x13 + x17 + x18 + x19 + x21 + x22$$

y = Share price
x2 = Rio Tinto price
x4 = VIX
x5 = Iron ore price
x9 = Total revenue
x11 = Tax provision
x12 = depreciation
x13 = change in working capital
x17 = Shares outstanding
x18 = Capital expenditure
x19 = Total assets
x21 = Inventory
x22 = Financial assets

TARGET

$$y \sim x2 + x5 + x8 + x17 + x18 + x19 + x22$$

y = Share price
x2 = Volatility Index (VIX)
x5 = EV Revenue Ratio
x8 = Cost of Revenue
x17 = Long term debt
x18 = Working capital
x19 = Shares outstanding
x22 = Capital Expenditure

The common variables that we can observe for the stocks are *Shares outstanding, Capex,* and *VIX.* Speaking of Vale, the mining industry is cyclical in nature. The revenue depends heavily on the underlying commodity price (Iron ore). Most mining companies also show the same behaviour hence we see a strong correlation between

Rio Tinto and VALE, where Rio Tinto is a competitive mining company for VALE. We see significant variables related to assets in VALE, this makes sense since it requires a lot of preliminary setup to operate a mine. Target (TGT) is a heavyweight in the retail space. Model for Target was supposedly affected by COVID related shocks. VALE and TARGET both have been conducting share buybacks, hence we see correlation between *stock price* and *shares outstanding*. Both are also capex intensive industries, they are mature and are expanding. Hence we see capex as a significant variable in both models.

Intertemporal Analysis

Target model equations and variables for time frames 2012-2018 and 2018-2023 are as follows:-

Target (2012 - 2018)

$$y \sim x4 + x5 + x7 + x14 + x16 + x17 + x18 + x19 + x20 + x21 + x22$$

y = Share price
x4= Walmart stock price
x5= EV Revenue Ratio
x7= Total Revenue
x14= Total Assets
x16= Inventory
x17= Long term debt
x18= Working capital
x19= Shares outstanding
x20= Operating cash flow
x21= Change in working capital
x22= Capital expenditure

Target (2018 - 2023)

$$y \sim x2 + x5 + x8 + x17 + x18 + x19 + x22$$

y = Share price
x2 = Volatility Index (VIX)
x5 =EV Revenue Ratio
x8 = Cost of Revenue
x17 = Long term debt
x18 = Working capital
x19 = Shares outstanding
x22 = Capital Expenditure

One of the primary discrepancies in the 2012-20/23 TGT analysis is the impact of COVID on supply chain disruptions and ries of outliers because of it. The residual plot for normality showed outliers which could cause difference in the equation. Case in point is that TGT posted a stunning drop in profit circa mid 2022 and due to this, some unanticipated variables are present in the 2018-2023 model i.e. VIX, and cost of revenue. The common variables that we can observe for both timeframes are *EV revenue ratio*, *long term debt, working capital, shares outstanding,* and *capex*. TGT's mismanagement of its supply chain could explain the deviation and the drop of Walmart as a significant explanatory variable compared between 2012-2018 and 2018-2023.

APPENDIX

Code

VALE (2018-2023)

```
library(quantmod)
library("readxl")
library(memisc)
library(dplyr)
library(lmtest)
library(sjPlot)
library(sgof)
library(ggplot2)
library(hexbin)
library(lmtest)
library(GGally)
library(rugarch)
library(forecast)
library(fGarch)
library(fUnitRoots)
library(tseries)
library(car)

vale <- read_excel("vale2.xlsx")
head(vale)

# Modifying column names for simplicity
colnames(vale) <- c("y", "x1", "x2", "x3","x4","x5","x6", "x7", "x8",
                                      "x9",  "x10",  "x11",  "x12",  "x13",
"x14","x15","x16","x17",
                             "x18","x19","x20",  "x21","x22",  "x23",  "x24",
"x25","x26","x27")
model_1 <- lm(y~ .-y, data = vale)

summary(model_1)

lm(formula = y ~., data =vale)
lm_AIC <- step(model_1, direction = 'backward', k =2)
lm_BIC <- step(model_1, direction = 'backward', k = log(nrow(vale)))

summary(lm_AIC)
summary(lm_BIC)

model_2 <- lm(y ~ x2 + x4 + x5 +x9 + x11 + x12 + x13 + x17 + x18 + x19 +
x21 +  x22, data =vale)
summary(model_2)
```

```
#model_2 <- lm(y ~ x2 +  x5 +x9 + x11  + x13 + x18 + x19 + x21 + x22, data
=vale)

lm_BIC_2 <- step(model_2, direction = 'backward', k = log(nrow(vale)))
summary(lm_BIC_2)
# Test for Gauss-Markov assumption

# Auto-correlation
durbinWatsonTest(model_2) #   No auto-correlation

#Conditional heteroskedasticity
bptest(model_2) # no conditional heteroskedasticity

# Normality of residuals
par(mfcol = c(2,1))
plot(model_2,2)
hist(model_2$residuals)

#Multicollinearity
vif(model_2) # For stock prediction higher the better

# Testing model fitness
fitted(model_2)

# Residuals analysis
k = residuals(model_2)

plot(model_2$resid,type='l')
Box.test(model_2$resid,lag=10,type='Ljung')   #   No   autocorrelation   in
residuals
# No need for ARIMA
Box.test(k^2,lag=10,type='Ljung')
# 6.64% volatility clustering insiginificant for 5%
#but we will test ARCH/GARCH for 10% just for purpose of studies
# Residual analysis for ARIMA model setup
par(mfcol = c(2,1))
Acf(model_2$residuals,lag = 36)
Pacf(model_2$residuals,lag = 36)

# Residual analysis for volatility. clustering setup
par(mfcol = c(3,1))
plot(k^2,type='l')
Acf(k^2,lag = 36)
Pacf(k^2,lag = 36)
```

```
# Test for stationarity
adf.test(model_2$residuals)

# Check best model
auto.arima(model_2$residuals, d = 0)
# arima(0,0,0) so further confirmation of our analysis

#Check volaitility using GARCH
#  After  running  various  models  ;  no  need  for  GARCH  also  ;  Tried  to
implement for case study purpose
model_3  =  garchFit(formula  =  ~  garch(1,1),  data  =  model_2$residuals,
cond.dist = "std", Trace=F)
summary(model_3)
predict(model_3,5)

model_3  =  garchFit(formula  =  ~  garch(5,5),  data  =  model_2$residuals,
cond.dist = "std", Trace=F)
summary(model_3)
predict(model_3,5)

# norm better than std other model tried but best fit is arch(1,0)

# Final model
# y ~ x2 + x4 + x5 +x9 + x11 + x12 + x13 + x17 + x18 + x19 + x21 +  x22
# arch (1,0) for price prediction interval
#y = share rpice
#x2 = Rio tinto price = 75
#x4 = VIX =  21.67
#x5 = Iron ore price = 120.53
#x9 = total revenue = 10174400000.00
#x11 = Tax provision = 110730000
#x12 = depreciation = 742740000.00. ; 804583333.33(average)
#x13 = change in working capital = 11284000.00
#x17 = shares outstanding = 4243392869.00
#x18 = capital expenditure = -1526160000.00
#x19 = total assets = 100668700000.00
#x21 = inventory= 5426685000 ; 4660350000.00 (average)
#x22 = financial assets = 205291666.67

# Values from 2024 financial model
new   <-   data.frame(x2   =c(76),x4   =   c(21.67),   x5   =   c(120.53),   x9=
c(10871400000), x11= c(1178317500), x12= c(742740000) ,
                          x13  =  c(-123140000),  x17  =c(4303392869)  ,x18  =
c(-1625160000) ,x19 = c(102639700000) ,
                  x21 = c(4660350000.00) ,  x22= c(205291666) )
# stock price :  (16.57961)
```

```
# Price prediction code
x <-predict(model_2, newdata=new)
x
```

TARGET (2018-2023)

```
library(quantmod)
library("readxl")
library(memisc)
library(dplyr)
library(lmtest)
library(sjPlot)
library(sgof)
library(ggplot2)
library(hexbin)
library(lmtest)
library(GGally)
library(rugarch)
library(forecast)
library(fGarch)
library(fUnitRoots)
library(tseries)
library(car)

target <- read_excel("target2018_23.xlsx")
head(target)

# Modifying column names for simplicity
colnames(target) <- c("y", "x1", "x2", "x3","x4","x5","x6", "x7", "x8",
                                      "x9",  "x10",  "x11",  "x12",  "x13",
"x14","x15","x16","x17",
                    "x18","x19","x20", "x21","x22", "x23")
model_1 <- lm(y~ .-y, data = target)

summary(model_1)

lm(formula = y ~., data = target)
lm_AIC <- step(model_1, direction = 'backward', k =2)
lm_BIC <- step(model_1, direction = 'backward', k = log(nrow(target)))

summary(lm_AIC)
summary(lm_BIC)

model_2 <- lm(y ~ x2 +  x5 + x8  + x17 + x18 + x19 + x22, data = target)
summary(model_2)
model_BIC <- lm(y ~ x2 +  x5 + x8  + x17 + x18 + x19 + x22, data = target)
```

```
summary(model_BIC)
#model_AIC <- lm(y ~ x2 +  x5 + x8 + x14  + x17 + x18 + x19 + x22, data =
target)
#summary(model_BIC)

lm_BIC_2 <- step(model_BIC, direction = 'backward', k = log(nrow(target)))
summary(lm_BIC_2)
# Test for Gass-Markov assumption

# Auto-correlation
durbinWatsonTest(model_2)

#Conditional heteroskedasticity
bptest(model_2)
# p-value is lower than 5% significance , enough evidence for conditional
heteroskedasticity

# Normality of residuals
par(mfcol = c(2,1))
plot(model_BIC,2)
hist(model_BIC$residuals)

#Multicollinearity
vif(model_BIC)

# Testing model fitness
fitted(model_BIC)
# Normality of residuals
par(mfcol = c(2,1))
plot(model_BIC,2)
hist(model_BIC$resid)
# Residuals analysis
k = residuals(model_BIC)
k
plot(model_BIC$resid,type='l')
Box.test(model_BIC$resid,lag=10,type='Ljung')
# low value; ARIMA model will be needed as signs of auto-correlation in
time series

# volatility clustering analysis
Box.test(k^2,lag=10,type='Ljung')
# Above 5% significance. (42%) No volatility clustering; ARCH/GARCH model
not necessary to be built

# Residual analysis for ARIMA model setup
par(mfcol = c(2,1))
Acf(model_BIC$residuals,lag = 36)
```

```
Pacf(model_BIC$residuals,lag = 36)
# As per ACF- PACF ARIMA(2,0,2) but let's check stationarity before

# Test for stationarity
adf.test(model_BIC$residuals)
# low p-value we reject null hypothesis( no presence of stationarity)

# Residual analysis for volatility. clustering setup
par(mfcol = c(3,1))
plot(k^2,type='l')
Acf(k^2,lag = 36)
Pacf(k^2,lag = 36)

#ARIMA model building
# Check best model
m1 <- auto.arima(model_BIC$residuals, d = 0)
m1
m2 <- arima(model_BIC$residuals, order=c(2,0,2))
m2
c1 <- c(0,NA,0,NA,NA)
m1 <- arima(model_BIC$residuals, order=c(2,0,2), fixed = c1)
m1
#m2 is better because both ar2 amd ma2 are significant compared to m1(ar1,ma1) not significant
 predict(m1,2) # -2.183617

#Check volatility using GARCH
# Not required as volatility clustering is not significant
# we did try running models; but were not signifcant ans also plotting residuals didn't show any significant need for ARCH/GARCH


# Values from 2024 financial model

# x2 = 20.18 ( Volatility Index) ( VIX)
# x5 = 2.67 ( Walmart price)
# x8 = 23121610733 (Cost of Revenue)
# x17 = 14883000000 (Long term debt)
# x18 = -3355754157.90 (Working capital)
# x19 = 461651176 (Shares outstanding)
# x22 = -1127000000 (Capital Expenditure)

new <- data.frame(x2 =c(20.18),x5 = c(2.818) ,x8 = c(21744581584), x17 = c(14883000000)
                        ,x18 = c(-3147211865) ,x19 = c(461651176) ,  x22= c(-1127000000) )
```

```
x <-predict(model_BIC, newdata=new)
x

# Adjustment with ARIMA prediction values for next year for 2024
price <- x  - 2.1836
price
```

TARGET (2012-2018)

```
library(quantmod)
library("readxl")
library(memisc)
library(dplyr)
library(lmtest)
library(sjPlot)
library(sgof)
library(ggplot2)
library(hexbin)
library(lmtest)
library(GGally)
library(rugarch)
library(forecast)
library(fGarch)
library(fUnitRoots)
library(tseries)
library(car)

target_old <- read_excel("target_2012_18.xlsx")
head(target_old)

# Modifying column names for simplicity
colnames(target_old)  <-  c("y",  "x1",  "x2",  "x3","x4","x5","x6",  "x7",
"x8",
                                      "x9",  "x10",  "x11",  "x12",  "x13",
"x14","x15","x16","x17",
                    "x18","x19","x20", "x21","x22", "x23")
model_1 <- lm(y~ .-y, data = target_old)

summary(model_1)

lm(formula = y ~., data =target_old)
lm_AIC <- step(model_1, direction = 'backward', k =2)
lm_BIC <- step(model_1, direction = 'backward', k = log(nrow(target_old)))

summary(lm_AIC)
summary(lm_BIC)
```

```
model_new <- lm(y ~ x4 + x5 + x7+ x14  + x16 + x17 + x18 + x19 + x20+x21+
x22, data =target_old)
summary(model_new)

#model_n <- lm(y ~  x5 + x7+ x14  + x17 + x18 + x19 + x20+x21+ x22, data =
target_old)
summary(model_new)


lm_BIC_2    <-    step(model_new,    direction    =    'backward',    k    =
log(nrow(target_old)))
summary(lm_BIC_2)
# Test for Gass-Markov assumption

# Auto-correlation
durbinWatsonTest(model_new)
# d-w stat between 1.5 -2.5 sp no auto-correlation
#Conditional heteroskedasticity
bptest(model_new)
# p-value is lower than 5% significance

# Normality of residuals
par(mfcol = c(2,1))
plot(model_2,2)
hist(model_2$residuals)

#Multicolliearity
vif(model_new) # Fir stock prediction higher the better

# Testing model fitness
fitted(model_new)

# Residuals analysis
k = residuals(model_new)
plot(model_new$resid,type='l')
Box.test(model_new$resid,lag=10,type='Ljung')
#  low  value;  ARIMA  model  will  be  needed  as  signs  of  auto-correlation  in
time seires

Box.test(k^2,lag=10,type='Ljung')
# Above 5% significance. (79%) No volatility clustering; ARCH/GARCH model
not necessary to be buil
par(mfcol = c(3,1))

# Test for stationarity
adf.test(model_new$residuals)
# low p-value we reject noll hypothesis( no presence of stationarity)
```

```
# Residual analysis for ARIMA model setup
par(mfcol = c(2,1))
Acf(model_new$residuals,lag = 36)
Pacf(model_new$residuals,lag = 36)


par(mfcol = c(3,1))
plot(k^2,type='l')
Acf(k^2,lag = 36)
Pacf(k^2,lag = 36)
# low value; ARIMA model will be needed as signs of auto-correlation in
time seire
par(mfcol = c(2,1))
Acf(model_n$residuals,lag = 36)
Pacf(model_n$residuals,lag = 36)


# Check best model
auto.arima(model_new$residuals, d = 0)


m1 <- arima(k, order=c(2,0,2))
m1

c1 = c(0,NA,0,NA,NA)
m2 <- arima(model_new$residuals, order=c(2,0,2), fixed = c1)
m2
predict(m2,2) # -0.491


#Check volaitility using GARCH # No need for Garch as Box test value is
high
#model_3 = garchFit(formula = ~ arma(2,2)+garch(1,1), data = k, cond.dist
= "std")
#summary(model_3)
#predict(model_3,5)


model_3 = garchFit(formula = ~arma(0,2) + garch(1,1), data =
model_2$residuals, cond.dist = "std")
summary(model_3)
predict(model_3,5)


# Values from 2019 financial statements
# X4: 39.086666 ( wmt stock)
```

```
# X5: 3.568 (EV/REVENUE)
# X7: 18665000000 (TOTAL REVENUE)
# X14:43741000000 (TOTAL ASSETS)
# X16:11396000000 (inventory)
# X17:10513000000 ( long term)
# X18:-2803000000 (Working capital)
# X19: 506677740 (shares outstanding)
# X20: 1347000000 (op. cash flow)
# X21:-138000000 (change in working capita)
# X22: -1009000000 (capital expenditure)
# Values from 2019 financial statement to create dataframes
new <- data.frame(x4 =c(39.086),x5 = c(3.568) ,x7 = c(18665000000), x14 = c(43741000000)
                      ,x16 = c(11396000000) ,x17 = c(10513000000) ,  x18= c(-2803000000), x19 =c(506677740)
                             ,x20=  c(1347000000),  x21  =c(-138000000),x22= c(-1009000000))

x <-predict(model_new, newdata=new)
x # 83.2222

price <- x - 0.491
price
# As per the model stock price predicted was  $ 82.73 which is a buy call
# because at end of 2018 stock price was  $ 66 ( which suggested a 25%) upmove
# as per the 2019 stock updates targets of 86 were reached in June-july of 2019 and which further moved ahead.
# So our moved showed insights as per the move seen in 2019
```