Name: Pramod Mantya Raju

Spring-2023

DS853 Project on Apple Revenue Dataset

Introduction

In the present research, we use stepwise forward regression from the Applied Multivariate Analysis class to analyze the Apple revenue dataset. The dataset used in this investigation was compiled from three separate sets received from Kaggle. Date, Y=Close, X1=high, X2=open, X3=low, X4=profit or not, and X7=volume are taken from the Apple revenue dataset. Furthermore, X5=Close-Intel comes from the Intel Revenue dataset, while X6=Close-Samsung comes from the Samsung Revenue dataset. This dataset contains a number of variables, including the closing price (Y) and a number of predictors related to Apple's financial performance, market circumstances, and competitor stock prices. The primary goal of this analysis is to find the most significant predictors of the closing price(Y) and provide insights into their interrelationships.

BESTFIT MODEL

The choice of the closing price (Y) as the dependent variable in stock market analysis is critical for various reasons. The closing price shows a stock's last trading price for a given trading day and is seen as an important measure of a stock's performance. Investors and traders use it extensively to appraise the worth of their investments, calculate returns, and make trading decisions.

In our investigation, the low price (X3) appeared as a highly important predictor, which represents Apple's stock's lowest trading price throughout a trading day. The low price provides useful information about the market's lowest point of demand for the item. Market players frequently search for support levels in technical analysis, which are price levels where buyers

are expected to enter the market and provide support to the stock. The low price can assist identify key support levels, signaling probable stock price reversals or floors. Understanding the dynamics of the low price is thus critical for understanding Apple's stock price behavior.

Assessing a company's profitability is a critical component that investors evaluate in stock market analysis. The variable indicating whether Apple generated a profit or not (X4) offers information about the company's financial performance. Positive profits often reflect a strong financial situation and can benefit a company's stock price. Negative gains or losses, on the other hand, may cause stock values to fall if investors perceive higher risk or poor performance. We acquire significant insights into the financial aspects that may have influenced Apple's stock price by considering its profitability.

Another key indicator considered was the trade volume (X7). Volume measures the number of shares traded over a given time period and is an important indicator of market interest and activity. Higher trading volume frequently indicates higher liquidity and increased investor sentiment. High trading volume can have a substantial impact on stock prices in stock analysis, emphasizing its importance in understanding Apple's stock price dynamics.

High and open prices (X1=high, X2=open) did not have a good predictive capacity for the closing price. While these elements are important for understanding intraday price movements and market mood, they may not have a direct impact on the ultimate closing price, which represents the market's consensus at the conclusion of the trading day.

We also investigated the variables Close-Intel (X5) and Close-Samsung (X6), which compare Apple's closing price to Intel's and Samsung's closing prices, respectively. While examining stock prices across firms can provide insights into market dynamics, we discovered that these variables had no meaningful predictive potential for Apple's closing price alone. Each

company's stock price is driven by various circumstances, and there may not have been a strong correlation between Apple's closing price and the closing prices of Intel and Samsung.

APPENDIX

FORWARD SELECTION

STEP 1 – Correlation Matrix

Low is the highest correlated variable with Y=Close: 0.996346

	Y=Close	*high*	*open*	*low*	*Close- Intel*	*se - Samsu*	*volume*	*profit or not*
Y=Close	1							
high	0.995994	1						
open	0.994858	0.999255	1					
low	0.996346	0.998844	0.999017	1				
Close- Intel	-0.32714	-0.33089	-0.32388	-0.31654	1			
Close - Sam	0.704755	0.702982	0.706714	0.710325	0.149692	1		
volume	-0.54252	-0.52805	-0.54037	-0.55634	-0.13405	-0.55307	1	
profit or nc	0.018971	0.060483	0.060034	0.062159	0.059843	0.082543	-0.08232	1

Regression Model – Y: Close, X: Low

F Statistic Hypothesis Test (95% confidence)

F * = 75937.18

F table = 3.858148

	Ho: B(Low)=0 vs Ha: B(Low) =/= 0		
	F* =	75937.18	
	Ftable =	3.858148	
	F*	**>**	**Ftable**
	As F* is greater than Ftable Accept Ha		

STEP 2 – Correlation Matrix

Profit or not the indicator variable is the highest correlated variable with Y=Close: -0.503

	Y=Close	*high*	*open*	*low*	*Close- Intel*	*se - Samsu*	*volume*	*profit or not*	*T - Residual*
Y=Close	1								
high	0.995994	1							
open	0.994858	0.999255	1						
low	0.996346	0.998844	0.999017	1					
Close- Intel	-0.32714	-0.33089	-0.32388	-0.31654	1				
Close - Sam	0.704755	0.702982	0.706714	0.710325	0.149692	1			
volume	-0.54252	-0.52805	-0.54037	-0.55634	-0.13405	-0.55307	1		
profit or nc	0.018971	0.060483	0.060034	0.062159	0.059843	0.082543	-0.08232	1	
1ST - Resid	0.085408	0.009371	-0.00595	-1.4E-15	-0.13763	-0.03482	0.138064	-0.503	1

Regression Model – Y: Close , X: Low, Profit or not

F Statistic Hypothesis Test (95% confidence)

F * = 189.6414

F table = 3.858208

Ho: B(Profit or not)=0 given B(Low) vs Ha: B(Profit or not) =/= 0 given B(Low)

F* =	189.6414
Ftable =	3.858208

F* > Ftable

As F* is greater than Ftable Accept Ha

STEP 3 – Correlation Matrix

Volume is the highest correlated variable with Y=Close 0.131937

	Y=Close	*high*	*open*	*low*	*profit or not*	*Close- Intel*	*se - Samsu*	*volume*	*OW*PROFIT - Residual*	*ND - Residuals*
Y=Close	1									
high	0.995994	1								
open	0.994858	0.999255	1							
low	0.996346	0.998844	0.999017	1						
profit or nc	0.018971	0.060483	0.060034	0.062159	1					
Close- Intel	-0.32714	-0.33089	-0.32388	-0.31654	0.059843	1				
Close - Sam	0.704755	0.702982	0.706714	0.710325	0.082543	0.149692	1			
volume	-0.54252	-0.52805	-0.54037	-0.55634	-0.08232	-0.13405	-0.55307	1		
LOW*PROF	0.237769	0.279121	0.279442	0.281015	0.944001	-0.01102	0.235255	-0.20029	1	
1ST - Resid	0.085408	0.009371	-0.00595	-1.4E-15	-0.503	-0.13763	-0.03482	0.138064	-0.49433	1
2ND - Resic	0.073769	0.009911	-0.0081	1.75E-15	-7.9E-16	-0.11286	-0.01787	0.131937	-0.03065	0.8637172

Regression Model – Y: Close, X: Low, Profit or not, Volume

F Statistic Hypothesis Test (95% confidence)

F * = 14.42856721

F table = 3.858238

Ho: B(Volume)=0 given B(Low),B(Profit or not) vs Ha: B(Volume) =/= 0 given B(Low),B(Profit or not)		
F* =	14.42857	
Ftable =	3.858238	
F*	**>**	**Ftable**
As F* is greater than Ftable Accept Ha		

STEP 4 – Correlation Matrix

Volume is the Close - Intel correlated variable with Y=Close -0.05483

	Y=Close	high	open	low	profit or not	volume	Close- Intel	se - Samsu	OW*PROFI	olume * Prof	ST - Residual	2ND - Residuals	D - Residuals
Y=Close	1												
high	0.995994	1											
open	0.994858	0.999255	1										
low	0.996346	0.998844	0.999017	1									
profit or nc	0.018971	0.060483	0.060034	0.062159	1								
volume	-0.54252	-0.52805	-0.54037	-0.55634	-0.08232	1							
Close- Intel	-0.32714	-0.33089	-0.32388	-0.31654	0.059843	-0.13405	1						
Close - Sam	0.704755	0.702982	0.706714	0.710325	0.082543	-0.55307	0.149692	1					
LOW*PROF	0.237769	0.279121	0.279442	0.281015	0.944001	-0.20029	-0.01102	0.235255	1				
Volume * P	-0.20007	-0.15935	-0.16549	-0.16963	0.829824	0.284099	0.006813	-0.15698	0.675043	1			
1ST - Resid	0.085408	0.009371	-0.00595	-1.4E-15	-0.503	0.138064	-0.13763	-0.03482	-0.49433	-0.363702	1		
2ND - Resic	0.073769	0.009911	-0.0081	1.75E-15	-7.9E-16	0.131937	-0.11286	-0.01787	-0.03065	0.0702132	0.8637172	1	
3RD - Resic	0.07283	0.004685	-0.01118	3.28E-16	5.84E-16	1.92E-16	-0.05483	0.0122	-0.03113	0.0264546	0.8527237	0.987271868	1

Regression Model – Y: Close, X: Low, Profit or not, Volume, Close - Intel

F Statistic Hypothesis Test (95% confidence)

F * = 2.212293703

F table = 3.858268

Ho: B(Close- Intel)=0 given B(Low),B(Profit or not),B(Volume) vs Ha: B(Close- Intel) =/= 0 given B(Low),B(Profit or not), B(Volume)		
F* =	2.212294	
Ftable =	3.858268	
F*	**>**	**Ftable**
As F* is greater than Ftable Accept Ha		

As the number observations are 560, we will always end up with F* > Ftable. Hence, considering the P-Value with level 0.05 level of significance. So, Close Intel show

insignificance with Y=close and indicates to stop the regression. In conclusion only Low, profit or now and Volume are statistically significant with Y =Close

	Coefficients	andard Erro	t Stat	P-value	Lower 95%	Upper 95%	ower 95.0%	Upper 95.0%
Intercept	2.698899	1.631532	1.654211	0.09865	-0.50583	5.903632	-0.50583	5.903632
low	1.010412	0.004252	237.6335	0	1.00206	1.018764	1.00206	1.018764
profit or nc	-2.74013	0.20179	-13.5791	1.81E-36	-3.13649	-2.34376	-3.13649	-2.34376
volume	6.32E-09	2.16E-09	2.919356	0.00365	2.07E-09	1.06E-08	2.07E-09	1.06E-08
Close- Intel	-0.03288	0.022109	-1.48738	0.137483	-0.07631	0.010543	-0.07631	0.010543

Durbin Watson Test

To check the error terms are auto co-related or not

D.W(D) = 1.733

DL =1.74

DU = 1.8

D < DL

D= 1.733 < DL=1.74

Since, D is less than DL, null hypothesis is rejected and there is positive auto co-relation in error terms

D		DL	
1.733	<	1.74	Since D is lessthan DL,null Hypothesis is rejected and there is positive auto co relation
1.733	<	1.8	
		DU	

Homoscedasticity Test

S1 = STDEV (First 280 error values) = 2.402473

S2= STDEV (Remaining 280 error values) = 2.333349

S1/S2 = 2.402473/2.333349 = 1.029624

Therefore,

0.5 < S1/S2 < 2.0

0.5 < 1.029624 < 2.0

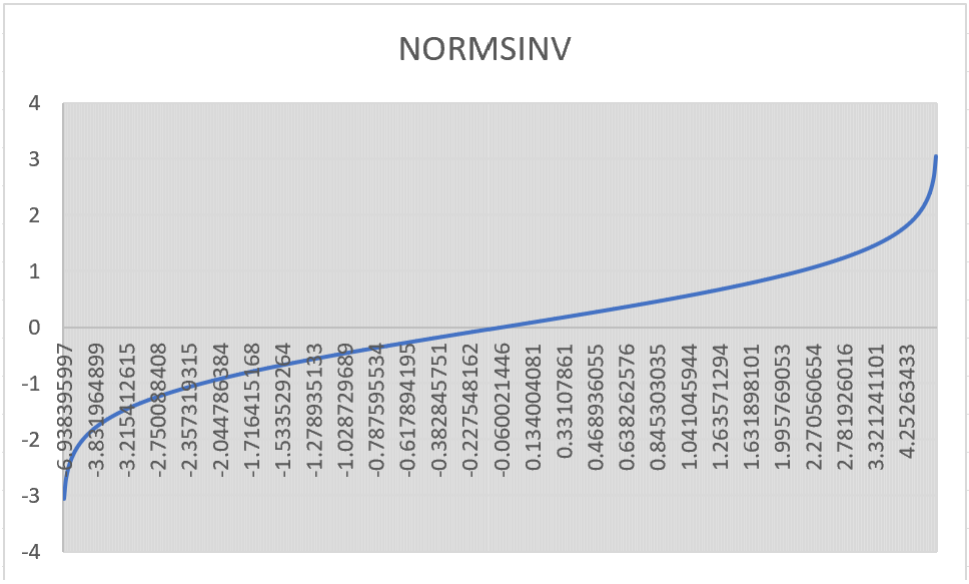
Homoscedasticity is holds good for the dataset because our S1/S2 values lies between the 0.5 and 2.0. Therefore, we assume it's a same spread

1.029624 Since homoscedasticity values is between 0.5 and 2.0 its on the same spread

Normality Test

Plot a graph for error versus norms inverse values and observe the graph

As the graph indicates Two tails in the graph being heavy suggests that the distribution deviates from a normal distribution.



Data Set Sample

date	Y=Close	X1 = high	X2 = open	X3 = low	X4 = profit	X5 = Close-	X6 = Close -	X7 = volume
03-01-2020	72.9666	73.7443	72.6747	72.4171	0	56.987	55200	135647456
06-01-2020	73.548	73.7394	72.8979	72.7384	0	56.2939	55500	146535512
07-01-2020	73.2021	73.5873	72.0736	71.8185	1	56.1347	55500	117288824
08-01-2020	74.3797	73.8179	73.5578	72.9789	0	55.198	55800	111510620
09-01-2020	75.9595	74.6863	72.9004	72.8994	0	55.2355	56800	132363784
10-01-2020	76.1313	76.1558	75.372	75.1181	1	55.5446	58600	167082252
13-01-2020	77.7578	76.7053	76.1975	75.621	0	55.2074	59500	140869088
14-01-2020	76.7078	77.7847	76.4526	76.3324	1	55.8162	60000	120114968
15-01-2020	76.379	77.9074	77.694	76.5827	1	55.6663	60000	162613828
16-01-2020	77.3358	77.3996	76.5042	75.9399	0	55.2074	59000	121923528
17-01-2020	78.192	77.4487	76.931	76.563	0	55.8818	60700	108829016
21-01-2020	77.6621	78.1944	77.5885	77.2769	1	55.8256	61300	137816468
22-01-2020	77.9393	78.2631	77.8142	77.5223	1	56.7154	61400	108940156
23-01-2020	78.3146	78.5011	78.1552	77.8436	0	58.7573	62300	101832460
28-01-2020	77.9368	78.3956	77.9933	77.4364	0	59.31	60800	104471972
29-01-2020	79.5683	78.111	76.6882	76.5876	1	63.0473	58800	162233944
30-01-2020	79.4529	80.4293	79.5952	78.8421	0	62.1294	59100	216599712
31-01-2020	75.9301	79.5069	78.6369	78.1969	1	62.2605	57200	126743232
03-02-2020	75.7216	79.161	78.7317	75.6308	0	59.8813	56400	199588384
04-02-2020	78.2214	76.9065	74.652	74.1417	0	60.3403	57200	173985604
05-02-2020	78.8593	78.4152	77.353	76.9419	1	61.3145	58900	136616536