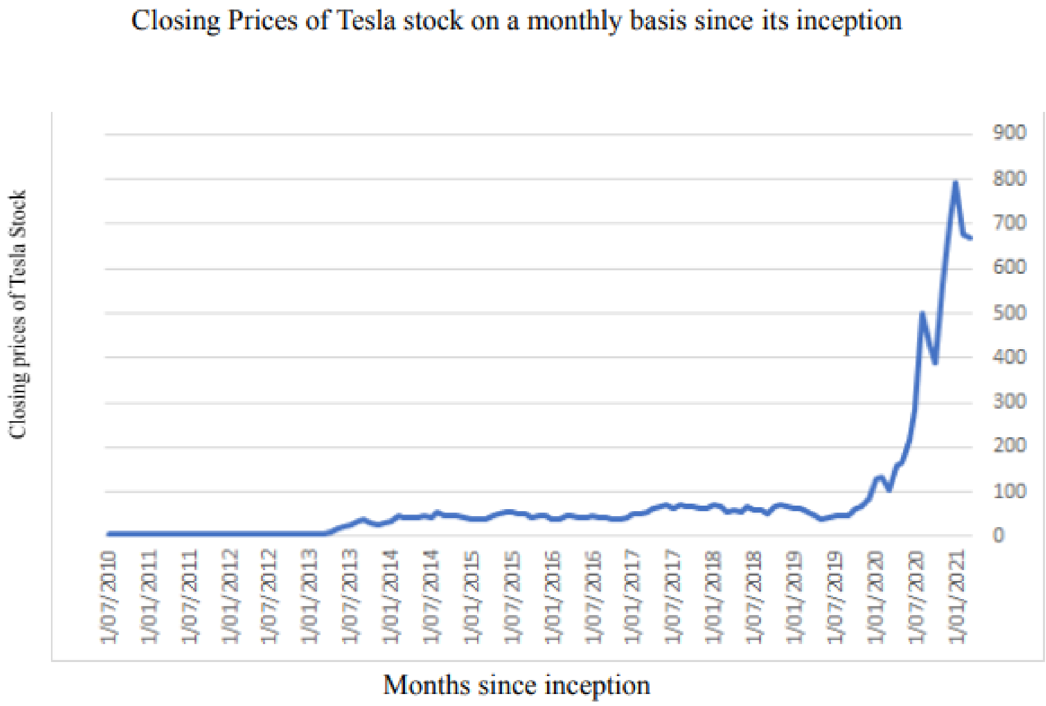
Tesla Stock Analysis



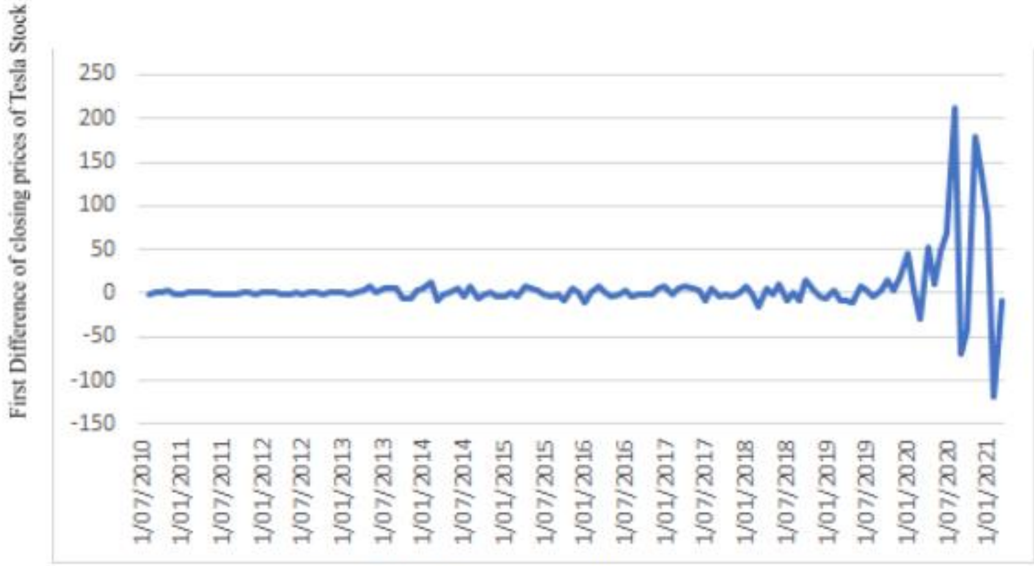
(Figure 1: Plot of Tesla's closing prices over the period of 07/2010 to 03/2021)

Observations

In Figure 1, between July 2010 and March 2013, the closing price ranged between approx. $3.90 and $7.60. Between March and August 2013 the stock price underwent a substantial re-rate, where it then ranged between $30 and $70 between August 2013 and November 2019. Moreover, the stock price then rose steeply from Mid-2019 until Early 2021, such that the closing price increased from $45 to $793 between August 2019 and January 2021. As such, linear regression Modelling would be deemed inappropriate as it would not effectively take into account the steep rise from 2019 onwards, as it represents less than 20% of the data. Throughout the entire data set, an analysis of descriptive statistics shows that the mean stock price was $79.14, whilst the median was noticeably lower at $45.58. In addition to this, the skewness of the data is 3.56, indicating a strong negative skewness, largely due to the steep rise in the share price after Mid-2019. Lastly, relative to the mean, the sample variance and standard deviation of 20279.49 and 142.41 respectively are very high, as they are heavily influenced by the substantially higher data points from 2020 onwards.

Mean	79.14446904
Standard Error	12.58703765
Median	45.581999
Mode	#N/A
Standard Deviation	142.4060749
Sample Variance	20279.49016
Kurtosis	12.51310108
Skewness	3.562293062
Range	789.634029
Minimum	3.896
Maximum	793.530029
Sum	10130.49204
Count	128

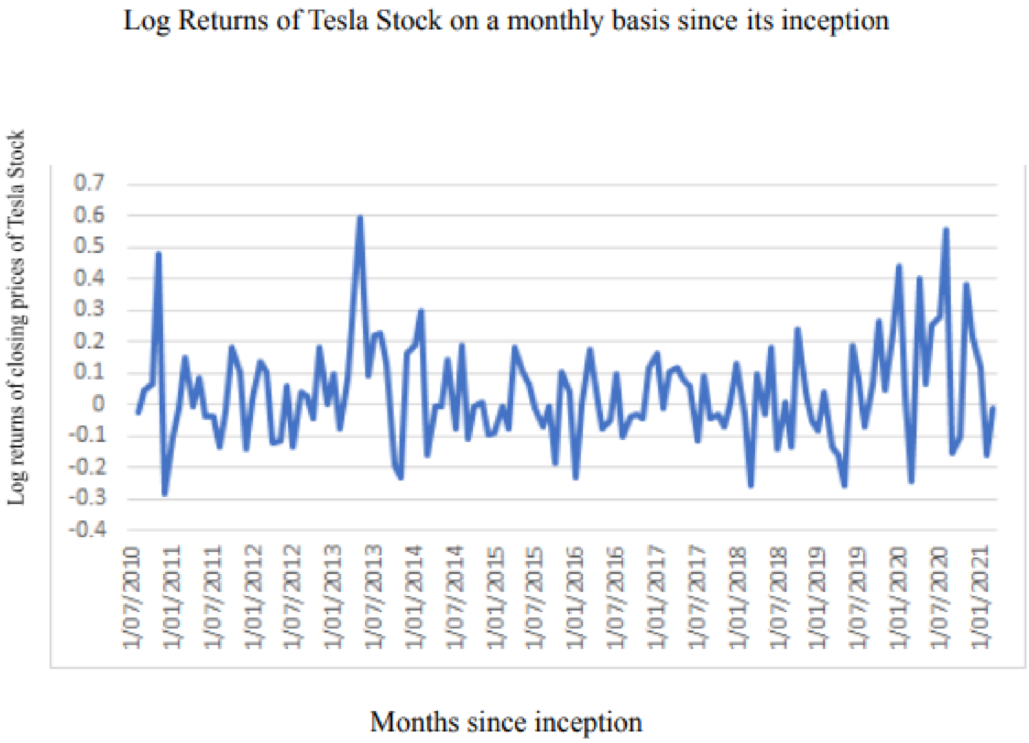
First Difference of Closing Prices of Tesla stock on a monthly basis since its inception



(Figure 2: Plot of Tesla's returns from first differences of closing prices over the period of 07/2010 to 03/2021)

Observations

By analysing Figure 2 of first differences, it can be observed that returns were generally stable until 2013 where there was more significant variability. Subsequently, it is evident that after November 2019 returns rapidly increase in volatility. In addition to this, the greatest return was $212.17 (max) in August 2020 whilst the lowest return was -$118.03 (min) in February 2021. The fundamental problem with fitting an OLS model to this data is accounting for the great variability in returns in the more recent time periods, even though the mean seems to remain the same, yet the variance increases over time (like a random walk model).



(Figure 3: First differences of the logarithm of the close price)

Observations

When analysing Figure 3, it can be seen that the graph is similar to a stationary time series graph. From just this visual, we can say that the mean and variance seem to be a finite constant. Figure 3 seems to be mean reverting where it returns to the mean many, many times. In comparison, the first difference of close price returns looks like a random walk as the variance grows with time (constant mean but not variance). If log_return is truly a stationary time series, we will be able to use the log_returns as a dependent/independent variable in a dynamic model. Given that we have a dynamically complete model, by relying on the law of large numbers and the central limit theorem, we will have consistent and asymptotically normal estimators which are required for meaningful inference. In comparison to Figure 2, by first using the natural log to scale returns, we account for the skewness caused by larger values. The benefit derived from using the first differences of logarithm of the close price is that it appears that the variance and mean will be constant and thus it will be easier and more reliable to model than using the first differences of close prices.

Initially, an OLS regression is run on ‘log_return’ with only a constant. By doing so we can conduct a Breush-Pagan test to check for the presence of any serial correlation in the model or look for any significant partial correlation in the correlogram. The presence of serial correlation in the OLS regression means that the unbiased estimators are not BLUE. This is because serial correlation indicates the cov(u|x) no longer has a diagonal matrix of constant variance of residuals. However, it is important to note that this still means that the OLS estimators remain unbiased despite being inconsistent.

Our regression on constant only is:

$log\_return\ t$ = 0. 0400

Equation: UNTITLED Workfile: TESLA DATA COPY::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LOG_RETURN
Method: Least Squares
Date: 05/19/21 Time: 15:54
Sample (adjusted): 1/08/2010 1/03/2021
Included observations: 128 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.040007	0.014322	2.793377	0.0060

R-squared	0.000000	Mean dependent var	0.040007
Adjusted R-squared	0.000000	S.D. dependent var	0.162036
S.E. of regression	0.162036	Akaike info criterion	-0.794216
Sum squared resid	3.334464	Schwarz criterion	-0.771935
Log likelihood	51.82983	Hannan-Quinn criter.	-0.785163
Durbin-Watson stat	1.797778		

or

Equation: UNTITLED Workfile: TESLA DATA COPY::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Correlogram of Residuals

Date: 05/19/21 Time: 15:55
Sample (adjusted): 1/08/2010 1/03/2021
Included observations: 128 after adjustments

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.100	0.100	1.3133	0.252
		2	-0.059	-0.070	1.7718	0.412
		3	-0.003	0.010	1.7732	0.621
		4	0.249	0.248	10.077	0.039
		5	-0.013	-0.071	10.101	0.072
		6	0.035	0.079	10.269	0.114
		7	0.055	0.047	10.692	0.153
		8	0.044	-0.029	10.961	0.204
		9	-0.044	-0.017	11.227	0.260
		10	0.062	0.049	11.765	0.301
		11	-0.035	-0.082	11.935	0.369
		12	-0.020	-0.005	11.995	0.446
		13	0.004	0.020	11.997	0.528
		14	0.007	-0.043	12.005	0.606
		15	-0.053	-0.014	12.414	0.647

(Figure 4: regression output and its correlogram)

Inspecting the Partial Correlation section of the correlogram in Figure 4, it is evident that only lags of order 4 are significant. Thus, we start by adding lags of 1, 2, 3 and 4 to the regression and every time run the Breusch-Godfrey test to see whether serial correlation has been resolved or not. Eventually, only when lags order up to 4 is added the regression becomes free of serial correlation and the steps taken for the BP test is included below:

Dependent Variable: LOGFD
Method: Least Squares
Date: 05/26/21 Time: 13:56
Sample (adjusted): 2010M12 2021M03
Included observations: 124 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.025169	0.015521	1.621548	0.1075
LOGFD(-1)	0.102042	0.085909	1.187781	0.2373
LOGFD(-2)	-0.057000	0.086904	-0.655904	0.5132
LOGFD(-3)	-0.007790	0.087039	-0.089496	0.9288
LOGFD(-4)	0.253005	0.086830	2.913798	0.0043

R-squared	0.080987	Mean dependent var	0.036685
Adjusted R-squared	0.050096	S.D. dependent var	0.159619
S.E. of regression	0.155569	Akaike info criterion	-0.843966
Sum squared resid	2.880008	Schwarz criterion	-0.730245
Log likelihood	57.32591	Hannan-Quinn criter.	-0.797770
F-statistic	2.621689	Durbin-Watson stat	1.813511
Prob(F-statistic)	0.038235		

(Figure 5: Regression of the AR(4) model of log_return)

Note: logfd is equivalent to log_return as per notation

$$log\_return_t = \beta_0 + \beta_1 log\_return_{t-1} + \beta_2 log\_return_{t-2} + \beta_3 log\_return_{t-3} + \beta_4 log\_return_{t-4}$$

Our AR(4) model is:

$$log\_return_t = \underset{(0.016)}{0.0252} + \underset{(0.086)}{0.1020} log\_return_{t-1} - \underset{(0.087)}{0.0570} log\_return_{t-2} - \underset{(0.087)}{0.008} log\_return_{t-3} + \underset{(0.087)}{0.2530} log\_return_{t-4}$$

Get error terms from original regression

- Run the auxiliary regression, including explanatory variables on top of lags of u:

$$\widehat{u}_t = \beta_0 + \beta_1 log\_return_{t-1} + \beta_2 log\_return_{t-2} + \beta_3 log\_return_{t-3} + \beta_4 log\_return_{t-4} + \rho_1 \widehat{u}_{t-1} + \rho_2 \widehat{u}$$
$$+ \rho_3 \widehat{u}_{t-3} + \rho_4 \widehat{u}_{t-4} + e_t$$

Getting our auxiliary regression:

1. Make series of residuals
2. Estimate $\widehat{u}_t$ equation

Dependent Variable: UHAT
Method: Least Squares
Date: 05/26/21 Time: 15:11
Sample (adjusted): 1/04/2011 1/03/2021
Included observations: 120 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000354	0.027817	0.012729	0.9899
LOG_RETURN(-1)	-0.190163	0.391431	-0.485813	0.6281
LOG_RETURN(-2)	0.292166	0.372541	0.784253	0.4346
LOG_RETURN(-3)	0.106369	0.370035	0.287456	0.7743
LOG_RETURN(-4)	-0.120138	0.386377	-0.310935	0.7564
UHAT(-1)	0.257929	0.403234	0.639650	0.5237
UHAT(-2)	-0.278586	0.379088	-0.734884	0.4640
UHAT(-3)	-0.158139	0.372069	-0.425025	0.6716
UHAT(-4)	0.129683	0.389364	0.333063	0.7397

R-squared	0.017702	Mean dependent var	0.004034
Adjusted R-squared	-0.053094	S.D. dependent var	0.151976
S.E. of regression	0.155959	Akaike info criterion	-0.806412
Sum squared resid	2.699868	Schwarz criterion	-0.597350
Log likelihood	57.38470	Hannan-Quinn criter.	-0.721511
F-statistic	0.250040	Durbin-Watson stat	1.995048
Prob(F-statistic)	0.979900		

(Figure 6: Auxiliary regression which is estimating the errors)

$$\widehat{u}_t = \underset{(0.0278)}{0.0004} - \underset{(0.3914)}{0.1902} log\_return_{t-1} + \underset{(0.3725)}{0.2922} log\_return_{t-2} + \underset{(0.3700)}{0.1064} log\_return_{t-3} - \underset{(0.3864)}{0.1201} log$$

$$\underset{(0.4032)}{0.2579}\widehat{u}_{t-1} - \underset{(0.3791)}{0.2786}\widehat{u}_{t-2} - \underset{(0.3721)}{0.1581}\widehat{u}_{t-3} + \underset{(0.3894)}{0.1297}\widehat{u}_{t-4}$$

Breush-Godfrey test:

$$H_0: \rho_1 = \rho_2 = \rho_3 = \rho_4 = 0$$
$$H_1: at\ least\ one\ of\ \rho_i \neq 0,\ for\ 1 \leq i \leq 4$$

Let $R_{\widehat{u}}^2$ denote the R^2 obtained from the auxiliary regression:

$R_{\widehat{u}}^2 = 0.0177$

n = 124

We know that $(n - p)R_{\widehat{u}}^2 \sim \chi_p^2$ under H_0

$BG_{calc} = 120 \times 0.0177 = 2.124$

$BG_{crit} = 9.49\ (\chi_4^2)$

$BG_{calc} < BG_{crit}$: 2. 124 < 9. 49 $\Rightarrow$ $We\ do\ not\ reject\ H_0$

Conclusion: There is no serial correlation

After inspecting the correlogram with the appropriate lags being accounted for, it is clear that the serial correlation has been removed from the model to a great extent:

Date: 05/26/21 Time: 15:15
Sample (adjusted): 1/12/2010 1/03/2021
Q-statistic probabilities adjusted for 4 dynamic regressors

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.070	0.070	0.6285	0.428
		2	-0.003	-0.008	0.6300	0.730
		3	-0.005	-0.005	0.6338	0.889
		4	0.010	0.011	0.6480	0.958
		5	-0.068	-0.070	1.2577	0.939
		6	0.033	0.044	1.4060	0.965
		7	0.079	0.074	2.2384	0.945
		8	0.010	-0.001	2.2528	0.972
		9	-0.024	-0.022	2.3292	0.985
		10	0.086	0.086	3.3513	0.972

(Figure 7: Correlogram of the AR(4) model)

So far the serial correlation issue has been resolved, however, we do not know whether this model is free from heteroskedasticity or not. Heteroskedasticity undermines the ‘homoskedasticity’ assumption which states that the covariance matrix of residuals must be:

$$Var\ (\ u \mid X) = \sigma^2 I_n$$

This is problematic, since the standard errors of the estimators cannot be used for t-tests and f-tests to make inferences as they are based upon the variance of residuals and thus the equation below is no longer true:

$$Var\ (\hat{\beta}) = \sigma^2 (X'X)^{-1}$$

In order to see whether our model suffers from this issue, we need to conduct the Breusch-Paugan or White’s (more general BP) test. Here we used a Breusch-Pagan test.

We get the error terms from original regression.

H_0: Var (u_t | log_return_{t-1}, log_return_{t-2}, log_return_{t-3}, log_return_{t-4}) = σ^2 for all t

H_1: Var (u_t | log_return_{t-1}, log_return_{t-2}, log_return_{t-3}, log_return_{t-4}) = $\alpha_0 + \alpha_1 log\_return_{t-1} + \alpha_2 log\_return_{t-2} + \alpha_3 log\_return_{t-3} + \alpha_4 log\_return_{t-4}$

Equation: UNTITLED Workfile: WORKING FILE::Untitled\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Heteroskedasticity Test: Breusch-Pagan-Godfrey
Null hypothesis: Homoskedasticity

F-statistic	3.070564	Prob. F(4,119)	0.0190
Obs*R-squared	11.60096	Prob. Chi-Square(4)	0.0206
Scaled explained SS	14.05438	Prob. Chi-Square(4)	0.0071

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 05/24/21 Time: 18:25
Sample: 1/12/2010 1/03/2021
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.018017	0.003653	4.932326	0.0000
LOG_RETURN(-1)	0.050340	0.020218	2.489828	0.0142
LOG_RETURN(-2)	0.027536	0.020452	1.346326	0.1808
LOG_RETURN(-3)	0.034239	0.020484	1.671483	0.0973
LOG_RETURN(-4)	0.013909	0.020435	0.680644	0.4974

R-squared	0.093556	Mean dependent var	0.023226
Adjusted R-squared	0.063087	S.D. dependent var	0.037825
S.E. of regression	0.036612	Akaike info criterion	-3.737372
Sum squared resid	0.159516	Schwarz criterion	-3.623650
Log likelihood	236.7170	Hannan-Quinn criter.	-3.691175
F-statistic	3.070564	Durbin-Watson stat	1.795652
Prob(F-statistic)	0.019016		

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: UHAT^2
Method: Least Squares
Date: 05/27/21 Time: 15:04
Sample (adjusted): 2010M12 2021M03
Included observations: 124 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.018017	0.003653	4.932326	0.0000
LOGFD(-1)	0.050340	0.020218	2.489828	0.0142
LOGFD(-2)	0.027536	0.020452	1.346326	0.1808
LOGFD(-3)	0.034239	0.020484	1.671483	0.0973
LOGFD(-4)	0.013909	0.020435	0.680644	0.4974

R-squared	0.093556	Mean dependent var	0.023226
Adjusted R-squared	0.063087	S.D. dependent var	0.037825
S.E. of regression	0.036612	Akaike info criterion	-3.737372
Sum squared resid	0.159516	Schwarz criterion	-3.623650
Log likelihood	236.7170	Hannan-Quinn criter.	-3.691175
F-statistic	3.070564	Durbin-Watson stat	1.795652
Prob(F-statistic)	0.019016		

(Figure 8: Breusch-Pagan test on AR(4) model and estimation of squared errors.)

$$\widehat{u}_t^2 = \underset{(0.004)}{0.0180} + \underset{(0.020)}{0.050} log\_return_{t-1} + \underset{(0.020)}{0.028} log\_return_{t-2} + \underset{(0.020)}{0.034} log\_return_{t-3} + \underset{(0.020)}{0.014} log\_return_{t-4}$$

Let $R_{\widehat{u}^2}^2$ denote the R^2 obtained from the auxiliary regression:

$R_{\widehat{u}^2}^2 = 0.093556$

n = 124

LM = $n \times R_{\widehat{u}^2}^2 \sim \chi_q^2$, under the null hypothesis.

$BP_{calc} = 124 \times R_{\widehat{u}^2}^2 = 11.60$

$BP_{crit} = \chi_4^2 = 9.49$

Decision rule: We reject H_0 at the 5% significance level if $BPcalc > BPcrit$

$11.60 > 9.49 \Rightarrow We\ reject\ the\ H_0$

Hence, we can reject the null hypothesis that the results are homoskedastic, proving that we have heteroskedasticity in our model. Even though there is heteroskedasticity, our OLS estimators are unbiased but they are not efficient and not BLUE.

In order to fix this issue, we can use 'robust standard errors' or White's standard errors. This is a method of estimating coefficients' standard errors in presence of heteroskedasticity. It is important to note that this method of calculating the correct standard error values is only valid in large enough samples. To get the correct standard error values, we will use the HAC / White coefficient covariance matrix.. The regression with the adjusted (HAC) standard errors are shown below:

Dependent Variable: LOG_RETURN
Method: Least Squares
Date: 05/24/21 Time: 20:41
Sample (adjusted): 1/12/2010 1/03/2021
Included observations: 124 after adjustments
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 5.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.025169	0.012804	1.965690	0.0517
LOG_RETURN(-1)	0.102042	0.088817	1.148898	0.2529
LOG_RETURN(-2)	-0.057000	0.076706	-0.743099	0.4589
LOG_RETURN(-3)	-0.007790	0.093677	-0.083154	0.9339
LOG_RETURN(-4)	0.253005	0.056518	4.476531	0.0000

R-squared	0.080987	Mean dependent var	0.036685
Adjusted R-squared	0.050096	S.D. dependent var	0.159619
S.E. of regression	0.155569	Akaike info criterion	-0.843966
Sum squared resid	2.880008	Schwarz criterion	-0.730245
Log likelihood	57.32591	Hannan-Quinn criter.	-0.797770
F-statistic	2.621689	Durbin-Watson stat	1.813511
Prob(F-statistic)	0.038235	Wald F-statistic	6.029515
Prob(Wald F-statistic)	0.000188		

(Figure 9: Regression model with adjusted (HAC) standard errors.)

The equation with adjusted standard errors.

$$log\_return_t = \underset{(0.013)}{0.0252} + \underset{(0.089)}{0.1020}\, log\_return_{t-1} - \underset{(0.077)}{0.0570}\, log\_return_{t-2} - \underset{(0.094)}{0.0078}\, log\_return_{t-3} + \underset{(0.057)}{0.2530}\, log\_return_{t-4}$$

Since our data is a monthly time series, it would be reasonable to check for monthly seasonality first. In order to do so, the first 11 columns of dummy variables – for each month except January – are added to the data set. Note that adding 12 dummy variables for 12 months is avoided to prevent perfect collinearity (dummy variable trap). After adding the monthly dummy variables, we compare this ‘unrestricted’ model to our initial/’restricted’ model using an F-test to see whether the new monthly independent variables are significant or not.

Dependent Variable: LOG_RETURN
Method: Least Squares
Date: 05/25/21 Time: 18:36
Sample (adjusted): 1/12/2010 1/03/2021
Included observations: 124 after adjustments
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 5.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.070380	0.049266	1.428566	0.1560
LOG_RETURN(-1)	0.131056	0.075161	1.743669	0.0841
LOG_RETURN(-2)	-0.080455	0.080033	-1.005264	0.3170
LOG_RETURN(-3)	0.006397	0.102233	0.062572	0.9502
LOG_RETURN(-4)	0.284830	0.068886	4.134775	0.0001
FEB	-0.061601	0.073550	-0.837539	0.4041
MARCH	-0.117080	0.071948	-1.627295	0.1066
APRIL	0.026898	0.064513	0.416938	0.6776
MAY	-0.054524	0.085242	-0.639639	0.5238
JUNE	0.017161	0.059365	0.289069	0.7731
JULY	-0.054619	0.071721	-0.761546	0.4480
AUG	-0.016111	0.072525	-0.222144	0.8246
SEP	-0.118059	0.060697	-1.945043	0.0544
OCT	-0.076766	0.065234	-1.176773	0.2419
NOV	-0.026399	0.073807	-0.357680	0.7213
DEC	-0.079154	0.052298	-1.513532	0.1331

R-squared	0.161983	Mean dependent var	0.036685
Adjusted R-squared	0.045592	S.D. dependent var	0.159619
S.E. of regression	0.155937	Akaike info criterion	-0.758809
Sum squared resid	2.626182	Schwarz criterion	-0.394902
Log likelihood	63.04616	Hannan-Quinn criter.	-0.610981
F-statistic	1.391713	Durbin-Watson stat	1.799369
Prob(F-statistic)	0.164284	Wald F-statistic	2.620404
Prob(Wald F-statistic)	0.002113		

(Figure 10: Regression of AR(4) model including monthly dummy variables.)

$$\begin{aligned}
log_return_t &= \underset{(0.049)}{0.0704} + \underset{(0.075)}{0.1311}\, log_return_{t-1} - \underset{(0.080)}{0.0805}\, log_return_{t-2} + \underset{(0.102)}{0.0064}\, log_return_{t-3} \\
&+ \underset{(0.069)}{0.2848}\, log_return_{t-4} - \underset{(0.074)}{0.0616}\, FEB - \underset{(0.072)}{0.1171}\, MARCH + \underset{(0.065)}{0.0269} APRIL - \underset{(0.085)}{0.0546}\, MAY \\
&+ \underset{(0.059)}{0.0172}\, JUNE - \underset{(0.072)}{0.0546}\, JULY - \underset{(0.073)}{0.0161}\, AUG - \underset{(0.061)}{0.1181}\, SEP - \underset{(0.065)}{0.0768}\, OCT - \underset{(0.074)}{0.0264}\, NOV \\
&- \underset{(0.052)}{0.0792}\, DEC
\end{aligned}$$

We can test for joint significance of all months' coefficients via an F test to see whether the monthly seasonality is significant or not.

H_0: $\beta_5 = \beta_6 = \beta_7 = \beta_8 = ... = \beta_{14} = \beta_{15} = 0$
H_1: at least one of the $\beta_i \neq 0, \quad for\ 5 \leq i \leq 15$

Test statistic

$F = \frac{(SSR_R - SSR_{UR})/q}{SSR_{UR}/(n-k-1)} \sim F_{q, n-k-1}$, under the null hypothesis

(Where q is the number of restrictions and k is the number of variables)

$$F_{calc} = \frac{2.88 - 2.63/11}{2.63/(124-15-1)} = 0.933$$

$$F_{crit} = F_{11, 108} \simeq 1.878$$

Decision rule: We reject H_0 if $F_{calc} > F_{crit}$

$F_{crit} > 0.933$

0.933 *is not greater than* $1.878 \Rightarrow$ *We do not reject the null at the* 5% *significance level* that there is no monthly seasonality.

The effect that each one of these tweets (figures 1-3 on the assignment sheet) would have on Tesla returns have been implemented into the model using dummy variables. Our assumption was that each of these actions by Elon Musk only affected the share returns of the following month. The only exception was Musk's tweet on 1st of May 2020 that we assumed that it affected the price on 1st of May and not the following month (1st of June), because the tweet was posted during stock market working hours in the USA.

Dependent Variable: LOG_RETURN
Method: Least Squares
Date: 05/26/21 Time: 12:36
Sample (adjusted): 1/12/2010 1/03/2021
Included observations: 124 after adjustments
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 5.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.023947	0.013298	1.800863	0.0743
LOG_RETURN(-1)	0.122303	0.091963	1.329908	0.1862
LOG_RETURN(-2)	-0.067936	0.076372	-0.889544	0.3756
LOG_RETURN(-3)	0.008275	0.092706	0.089263	0.9290
LOG_RETURN(-4)	0.272136	0.062808	4.332823	0.0000
PIC1	0.185503	0.021983	8.438656	0.0000
PIC2	-0.144045	0.053473	-2.693798	0.0081
PIC3	-0.112847	0.037844	-2.981920	0.0035

R-squared	0.101467	Mean dependent var	0.036685
Adjusted R-squared	0.047245	S.D. dependent var	0.159619
S.E. of regression	0.155802	Akaike info criterion	-0.818116
Sum squared resid	2.815828	Schwarz criterion	-0.636162
Log likelihood	58.72319	Hannan-Quinn criter.	-0.744202
F-statistic	1.871331	Durbin-Watson stat	1.784184
Prob(F-statistic)	0.080449		

(Figure 11: Regression of AR(4) model with Elon's online activity as dummy variables)

$$\underset{(0.0133)}{log\_return_t = 0.0239} + \underset{(0.0920)}{0.1223 log\_return_{t-1}} - \underset{(0.0764)}{0.0679 log\_return_{t-2}} + \underset{(0.0927)}{0.0083 log\_return_{t-3}}$$
$$+ \underset{(0.0628)}{0.2721 log\_return_{t-4}} + \underset{(0.0220)}{0.1855 PIC1} - \underset{(0.0535)}{0.1440 PIC2} - \underset{(0.0378)}{0.1128 PIC3}$$

The results show that the pic (Musk smoking in a podcast) had a positive effect on the return, while his tweets about TSLA and bitcoin being overpriced had a reducing effect on the price. All the effects were around 0.1 to 0.2 in terms of log-returns. These coefficients could also be interpreted as percentage change on Tesla returns (as an approximation). For instance, the first figure resulted in an 18.5% (exact: 20.3%) increase while the 2nd and the 3rd figures resulted in approximately 14.4% (exact: 15.4%) and 11.3% (exact: 12%) drop in returns.

Furthermore, in order to see whether the new added variables were significant or not, we conducted an F-test to compare the model with and without the figures 1-3.

H_0: $\beta_5 = \beta_6 = \beta_7 = 0$
H_1: at least one of the $\beta_i \neq 0$, $for\ 5 \leq i \leq 7$

$$F = \frac{(SSR_r - SSR_{ur})/3}{SSR_{ur}/(124-6-1)} \sim F_{3,116}\ under\ H_0$$

$$F = \frac{(2.880 - 2.816)/3}{2.816/(124-7-1)} \simeq 0.879$$

$F_{crit} \simeq 2.683$

=> $F_{crit} > 2.683$, hence we cannot reject the null hypothesis that all these actions (pic 1,2,3) had no effect on Tesla's log return. However, we kept these variables in the model as it was required for the rest of the questions.

To examine whether all these actions by Elon Musk had the same effect over TSLA returns, the following model was used as the restrictive model:

$$log\_return_t = \beta_0 + \beta_1 log\_return_{t-1} + \beta_2 log\_return_{t-2} + \beta_3 log\_return_{t-3} + \beta_4 log\_return_{t-4} + \beta_5 (PIC1 + PIC2 + PIC3)$$

Then, an F-test was conducted to see whether this model, that stated all the 3 actions had the same on TSLA returns, was more accurate than the initial model. NOTE: Musk_effect is the sum of pic1, pic2 and pic3.

Dependent Variable: LOG_RETURN
Method: Least Squares
Date: 05/26/21 Time: 12:48
Sample (adjusted): 1/12/2010 1/03/2021
Included observations: 124 after adjustments
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 5.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.025489	0.013346	1.909796	0.0586
LOG_RETURN(-1)	0.102045	0.088816	1.148950	0.2529
LOG_RETURN(-2)	-0.058092	0.076863	-0.755780	0.4513
LOG_RETURN(-3)	-0.008315	0.093221	-0.089198	0.9291
LOG_RETURN(-4)	0.258181	0.060330	4.279446	0.0000
MUSK_EFFECT	-0.018966	0.094218	-0.201302	0.8408

R-squared	0.081294	Mean dependent var	0.036685
Adjusted R-squared	0.042365	S.D. dependent var	0.159619
S.E. of regression	0.156201	Akaike info criterion	-0.828171
Sum squared resid	2.879047	Schwarz criterion	-0.691706
Log likelihood	57.34660	Hannan-Quinn criter.	-0.772736
F-statistic	2.088295	Durbin-Watson stat	1.810329
Prob(F-statistic)	0.071615	Wald F-statistic	4.977780
Prob(Wald F-statistic)	0.000358		

(Figure 12: Regression of AR(4) model including the effect of Elon Musk on log_return)

H_0: $\beta_5 = \beta_6 = \beta_7$

H_1: at least two of the B_i $are\ not\ equal,\quad for\ 5 \leq i \leq 7$

$$F = \frac{(SSR_r - SSR_{ur})/2}{SSR_{ur}/(124-7-1)} \sim F_{2,116}\ under\ H_0$$

$$F = \frac{(2.879 - 2.816)/2}{2.816/(124-7-1)} \simeq 1.298$$

$F_{crit} \simeq 3.074$

$=> F_{crit} > 1.298$, hence we cannot reject the null hypothesis that all the actions (pic 1, 2, 3) had the same effect on Tesla's log return.

$$\underset{}{log\_return_t} = \underset{(0.0133)}{0.0255} + \underset{(0.0888)}{0.1020}\, log\_return_{t-1} - \underset{(0.0769)}{0.0581}\, log\_return_{t-2} - \underset{(0.0932)}{0.0083}\, log\_return_{t-3} + \underset{(0.0603)}{0.2582}\, log\_return_{t-4} - \underset{(0.0942)}{0.0190}(PIC1 + PIC2 + PIC3)$$

Short-term: In the short term, the effect on the log_returns is equivalent to the coefficient of 'Musk_effect', which is: -0.019. This means that keeping every other variable in the model constant, Musk's tweets 1-3 on average resulted in a decrease of the return of Tesla stock by around 1.9%.

Long-term: For the long-run effect, the following formula is used:

$$\frac{sum\ of\ the\ coefficients\ x_t\ and\ its\ lags}{1 - sum\ of\ the\ coefficients\ of\ lags\ of\ y_t}$$

$$\frac{-0.019}{1-(0.102 - 0.058 - 0.008 + 0.258)} \simeq -0.027$$

Hence, the long run effect of each of these actions by Elon Musk that attracts media's attention is a 2.7% decrease in Tesla returns.

Overall, it was found that the given tweets had negative effects on TSLA stack return both on short and long run.

In order to test whether the Covid-19 breakout has changed the dynamic of Tesla's return, a number of new variables were added to the model. A dummy variable called 'covid', was added to the model, which holds the value of 1 only from 1/1/20 onwards. This variable allows for a constant difference between the covid & no-covid times (intercept). Meanwhile, in order to detect whether the interaction between our explanatory and response variable has changed in Covid times, new 'interactive' variables ($Covid * log(return)$) were added as well to detect any difference in the slope.

The following model was obtained after adding the new regressors:

Dependent Variable: LOG_RETURN
Method: Least Squares
Date: 05/26/21 Time: 13:10
Sample (adjusted): 1/12/2010 1/03/2021
Included observations: 124 after adjustments
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed
bandwidth = 5.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.019732	0.012418	1.588975	0.1149
LOG_RETURN(-1)	0.088335	0.135098	0.653858	0.5145
LOG_RETURN(-2)	-0.019147	0.090318	-0.211996	0.8325
LOG_RETURN(-3)	-0.118503	0.104312	-1.136038	0.2583
LOG_RETURN(-4)	0.236613	0.074709	3.167112	0.0020
MUSK_EFFECT	-0.142761	0.175447	-0.813698	0.4175
COVID	0.368655	0.089105	4.137293	0.0001
COVID*LOG_RETURN(-1)	-0.388315	0.185949	-2.088295	0.0390
COVID*LOG_RETURN(-2)	-0.667276	0.249947	-2.669667	0.0087
COVID*LOG_RETURN(-3)	-0.197586	0.149685	-1.320018	0.1895
COVID*LOG_RETURN(-4)	-0.306745	0.198155	-1.548004	0.1244

R-squared	0.182109	Mean dependent var	0.036685
Adjusted R-squared	0.109729	S.D. dependent var	0.159619
S.E. of regression	0.150607	Akaike info criterion	-0.863763
Sum squared resid	2.563112	Schwarz criterion	-0.613577
Log likelihood	64.55330	Hannan-Quinn criter.	-0.762131
F-statistic	2.516019	Durbin-Watson stat	1.770611
Prob(F-statistic)	0.009045	Wald F-statistic	7.095071
Prob(Wald F-statistic)	0.000000		

(Figure 13: Regression of the AR(4) model including the effect of COVID on log_returns)

$$log\_return_t = \underset{(0.0124)}{0.0197} + \underset{(0.1351)}{0.0883} log\_return_{t-1} - \underset{(0.0903)}{0.0191} log\_return_{t-2} - \underset{(0.1043)}{0.1185} log\_return_{t-3}$$
$$+ \underset{(0.0747)}{0.2366} log\_return_{t-4} - \underset{(0.1754)}{0.1428} MUSK\_EFFECT + \underset{(0.0891)}{0.3687} COVID$$
$$- \underset{(0.2499)}{0.3883} COVID * log\_reutrn_{t-1} - \underset{(0.1859)}{0.6673} COVID * log\_reutrn_{t-2} - \underset{(0.1497)}{0.1976} COVID * log\_retur$$
$$- \underset{(0.1982)}{0.3067} COVID * log\_return_{t-4}$$

Finally, an F-test was conducted to evaluate whether the new variables are significant or not.

H_0: $\beta_6 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = 0$

H_1: at least two of the $B_i \neq 0$, $for\ 6 \leq i \leq 10$

$$F = \frac{(SSR_r - SSR_{ur})/5}{SSR_{ur}/(124 - 10 - 1)} \sim F_{5,113}\ under\ H_0$$

$$F = \frac{(2.879 - 2.563)/5}{2.563/(124 - 10 - 1)} \simeq 2.786$$

$F_{crit} \simeq 2.295$

=> $F_{crit} < 2.786$, hence we can reject the null hypothesis that Covid19 had no effect on dynamics of Tesla stock returns, meaning that the dynamics of TSLA's return has changed since the outbreak started in January 2020.

Even though our OLS can give us an estimate for the Tesla stock return in April, yet this prediction itself has some uncertainty in it.

$$log(return) \ = \ log\widehat{\_return} + \ \hat{u}$$

Since the above is true, then the uncertainty (or standard error) of log(returns) is:

$$se(log(return)) \ = \ \sqrt{se(log\widehat{\_return})^2 \ + \ \sigma^2}$$

=> the standard error of regression (σ^2) is given with our regression. However, in order to get the standard error of $log\widehat{\_return}$, we need to transform our regressors and run regressions so that in the below model the standard error for 'c' is representing the standard error for our prediction($log\widehat{\_return}$).

Dependent Variable: LOG_RETURN
Method: Least Squares
Date: 05/27/21 Time: 18:12
Sample (adjusted): 1/12/2010 1/03/2021
Included observations: 124 after adjustments
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 5.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.449946	0.101253	4.443769	0.0000
LOG_RETURN(-1)+0.0113	0.088335	0.135098	0.653858	0.5145
LOG_RETURN(-2)+0.1610	-0.019147	0.090318	-0.211996	0.8325
LOG_RETURN(-3)-0.1173	-0.118503	0.104312	-1.136038	0.2583
LOG_RETURN(-4)-0.2177	0.236613	0.074709	3.167112	0.0020
MUSK_EFFECT	-0.142761	0.175447	-0.813698	0.4175
COVID-1	0.368655	0.089105	4.137293	0.0001
COVID*LOG_RETURN(-1)+0.0113	-0.388315	0.185949	-2.088295	0.0390
COVID*LOG_RETURN(-2)+0.1610	-0.667276	0.249947	-2.669667	0.0087
COVID*LOG_RETURN(-3)-0.1173	-0.197586	0.149685	-1.320018	0.1895
COVID*LOG_RETURN(-4)-0.2177	-0.306745	0.198155	-1.548004	0.1244

R-squared	0.182109	Mean dependent var	0.036685
Adjusted R-squared	0.109729	S.D. dependent var	0.159619
S.E. of regression	0.150607	Akaike info criterion	-0.863763
Sum squared resid	2.563112	Schwarz criterion	-0.613577
Log likelihood	64.55330	Hannan-Quinn criter.	-0.762131
F-statistic	2.516019	Durbin-Watson stat	1.770611
Prob(F-statistic)	0.009045	Wald F-statistic	7.095071
Prob(Wald F-statistic)	0.000000		

(Figure 14: Adjusted regression of the AR(4) model including the effect of COVID on log_returns)

This way we have:

$log\widehat{\_return}$: 0.4499

$se(log\widehat{\_return})$: 0.1013

σ^2: 0.1506

$$se(log(return)) \ = \ \sqrt{se(log\widehat{\_return})^2 \ + \ \sigma^2}$$

$$se(log(return)) \ = \ \sqrt{(0.1013)^2 \ + \ (0.1506)^2} \ \simeq \ 0.1815$$

$$prediction\ interval\ : log\widehat{\_return} \ \pm \ t_{n-k-1}(0.025).\ 0.1815$$

$$log\widehat{\_return} \ \pm \ (1.98) \ \times \ (0.1815)$$

$prediction\ interval : [0.0905, 0.8093]$

Based on this calculations and assuming that all our assumptions have been valid so far, there is 95% chance that TSLA's return on first of April lies anywhere between around +9% to +125%$(1 - e^{0.8093})$.