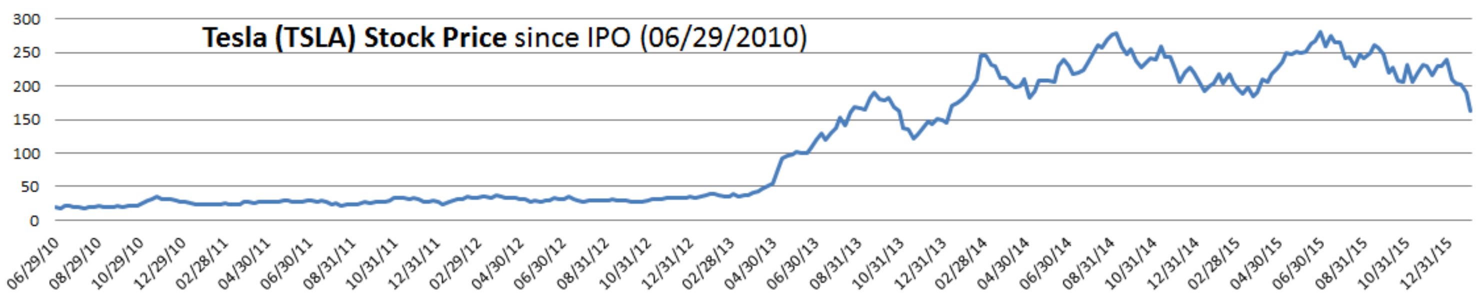
Joseph Sobieralski - Student Project
Time Series, Winter 2016

Tesla Motors (TSLA) Stock Pattern

Introduction

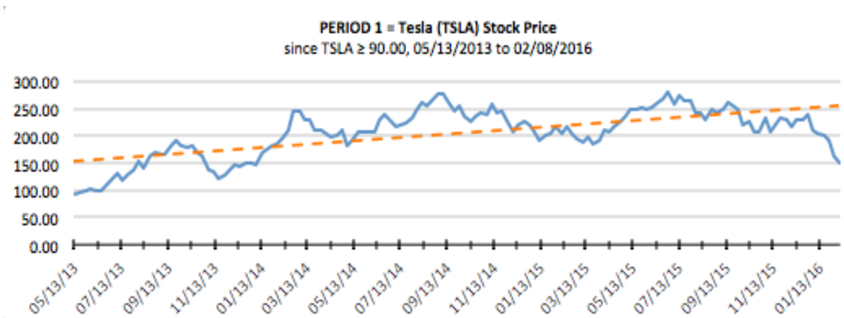
I have been dreaming of investing in my future and wish I had some extra money to buy a few stocks. One of the companies I have been thinking about "going long" on is Tesla Motors (TSLA on the NASDAQ exchange), founded in 2003 by famed inventor Elon Musk. Its traded stock had an initial public offering (IPO) on June 29th, 2010 of $19.00. The stock performance from 06/29/10 to 02/08/16 is shown. Notice the stock's recent dip towards $150. If there is an upward trend I can both model and validate---a model and trend I can trust my money with--- this dip has me thinking of "buying low and selling high" with my imaginary amount to invest!

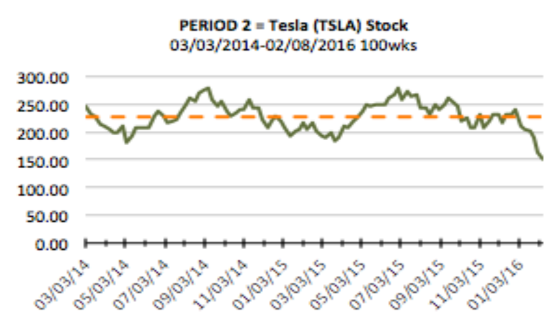
I got the data from Yahoo Finance (http://finance.yahoo.com/q/hp?s=TSLA&a=05&b=26&c=2010&d=01&e=8&f=2016&g=w).



Method

Step 1 - *"Choosing a Period"* - To analyze a Time Series, it is best to look at specific periods to see whether a pattern (or trend) can be best suited for a model with which to forecast. I notice after Tesla's sharp dip upward, there is an up and down pattern with a positively sloped trend line (*left*; ***Period 1***; TSLA ≥ 90.00). I also see a shorter trend being much less positively sloped and nearly flat (*right*; ***Period 2***; the past 100 weeks up to Feb. 8, 2016).





Step 2 - *"Finding a Pattern through Transformation"* -- Period 1 (143 weeks), Period 2 (100 weeks) -- After picking out these 2 periods, I wanted to look for stationarity and begin to identify what type of model I might use. I did this by finding the sample autocorrelation function, r_k, at lag k (by week) in my series. The values of r_k (ACF) versus lag k give us a *correlogram* with which to diagnose our time series pattern.

$$r_k = \frac{\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}$$

Period 1	Date	Price
0	05/13/13	$91.50
1	05/20/13	$97.08
2	05/28/13	$97.76
3	06/03/13	$102.04
4	06/10/13	$100.30
5	06/17/13	$99.55
6	06/24/13	$107.36
7	07/01/13	$120.09
8	07/08/13	$129.90
9	07/15/13	$119.68
10	07/22/13	$129.39
11	07/29/13	$138.00
12	08/05/13	$153.00
13	08/12/13	$142.00
14	08/19/13	$161.84
15	08/26/13	$169.00
16	09/03/13	$166.97
17	09/09/13	$165.54
18	09/16/13	$183.39
19	09/23/13	$190.90
20	09/30/13	$180.98
21	10/07/13	$178.70
22	10/14/13	$183.40
23	10/21/13	$169.66

Period 1	Period 2	Date	Price
24		10/28/13	$162.17
25		11/04/13	$137.95
26		11/11/13	$135.45
27		11/18/13	$121.38
28		11/25/13	$127.28
29		12/02/13	$137.36
30		12/09/13	$147.65
31		12/16/13	$143.24
32		12/23/13	$151.12
33		12/30/13	$149.56
34		01/06/14	$145.72
35		01/13/14	$170.01
36		01/21/14	$174.60
37		01/27/14	$181.41
38		02/03/14	$186.53
39		02/10/14	$198.23
40		02/18/14	$209.60
41		02/24/14	$244.81
42		03/03/14	$246.21
43	0	03/10/14	$230.97
44	1	03/17/14	$228.89
45	2	03/24/14	$212.37
46	3	03/31/14	$212.23
47	4	04/07/14	$203.78

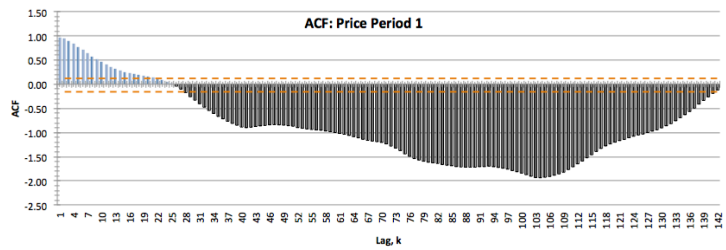
Period 1	Period 2	Date	Price
48	5	04/14/14	$198.12
49	6	04/21/14	$199.85
50	7	04/28/14	$210.91
51	8	05/05/14	$182.26
52	9	05/12/14	$191.56
53	10	05/19/14	$207.30
54	11	05/27/14	$207.77
55	12	06/02/14	$208.17
56	13	06/09/14	$206.42
57	14	06/16/14	$229.59
58	15	06/23/14	$239.06
59	16	06/30/14	$229.25
60	17	07/07/14	$218.13
61	18	07/14/14	$220.02
62	19	07/21/14	$223.57
63	20	07/28/14	$233.27
64	21	08/04/14	$248.13
65	22	08/11/14	$262.01
66	23	08/18/14	$256.78
67	24	08/25/14	$269.70
68	25	09/02/14	$277.39
69	26	09/08/14	$279.20
70	27	09/15/14	$259.32
71	28	09/22/14	$246.60

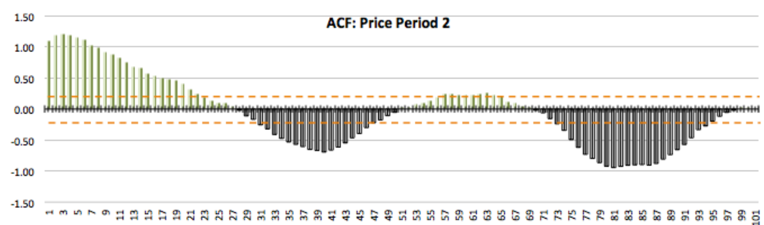
Period 1	Period 2	Date	Price
72	29	09/29/14	$255.21
73	30	10/06/14	$236.91
74	31	10/13/14	$227.48
75	32	10/20/14	$235.24
76	33	10/27/14	$241.70
77	34	11/03/14	$240.20
78	35	11/10/14	$258.68
79	36	11/17/14	$242.78
80	37	11/24/14	$244.52
81	38	12/01/14	$223.71
82	39	12/08/14	$207.00
83	40	12/15/14	$219.29
84	41	12/22/14	$227.82
85	42	12/29/14	$219.31
86	43	01/05/15	$206.66
87	44	01/12/15	$193.07
88	45	01/20/15	$201.29
89	46	01/26/15	$203.60
90	47	02/02/15	$217.36
91	48	02/09/15	$203.77
92	49	02/17/15	$217.11
93	50	02/23/15	$203.34
94	51	03/02/15	$193.88
95	52	03/09/15	$188.68

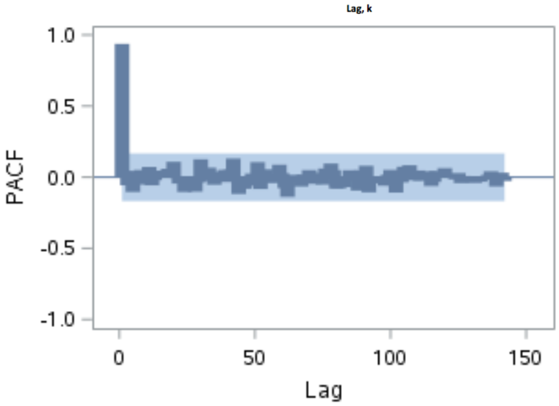
Period 1	Period 2	Date	Price
96	53	03/16/15	$198.08
97	54	03/23/15	$185.00
98	55	03/30/15	$191.00
99	56	04/06/15	$210.90
100	57	04/13/15	$206.79
101	58	04/20/15	$218.43
102	59	04/27/15	$226.03
103	60	05/04/15	$236.61
104	61	05/11/15	$248.84
105	62	05/18/15	$247.73
106	63	05/26/15	$250.80
107	64	06/01/15	$249.14
108	65	06/08/15	$250.69
109	66	06/15/15	$262.51
110	67	06/22/15	$267.09
111	68	06/29/15	$280.02
112	69	07/06/15	$259.15
113	70	07/13/15	$274.66
114	71	07/20/15	$265.41
115	72	07/27/15	$266.15
116	73	08/03/15	$242.51
117	74	08/10/15	$243.15
118	75	08/17/15	$230.77
119	76	08/24/15	$248.48

Period 1	Period 2	Date	Price
120	77	08/31/15	$241.93
121	78	09/08/15	$250.24
122	79	09/14/15	$260.62
123	80	09/21/15	$256.91
124	81	09/28/15	$247.57
125	82	10/05/15	$220.69
126	83	10/12/15	$227.01
127	84	10/19/15	$209.09
128	85	10/26/15	$206.93
129	86	11/02/15	$232.36
130	87	11/09/15	$207.19
131	88	11/16/15	$220.01
132	89	11/23/15	$231.61
133	90	11/30/15	$230.38
134	91	12/07/15	$217.02
135	92	12/14/15	$230.46
136	93	12/21/15	$230.57
137	94	12/28/15	$240.01
138	95	01/04/16	$211.00
139	96	01/11/16	$204.99
140	97	01/19/16	$202.55
141	98	01/25/16	$191.20
142	99	02/01/16	$162.60
143	100	02/08/16	$151.04

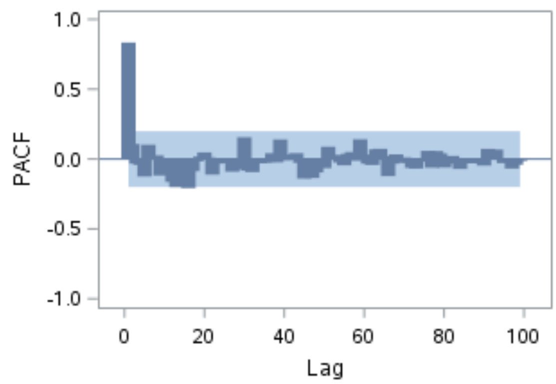
Periods 1 and 2 (Date; Price): *Note these values are closing prices (usually Friday) for the week with the given starting date (usually Monday).*



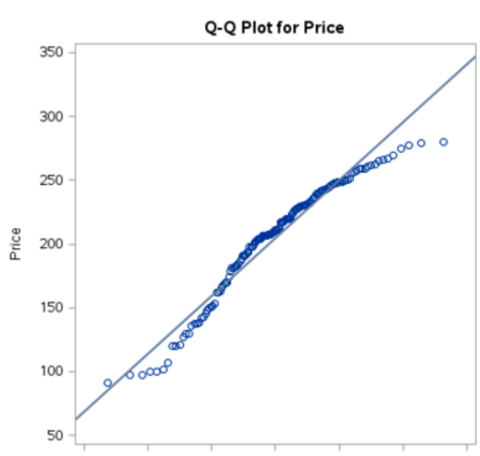




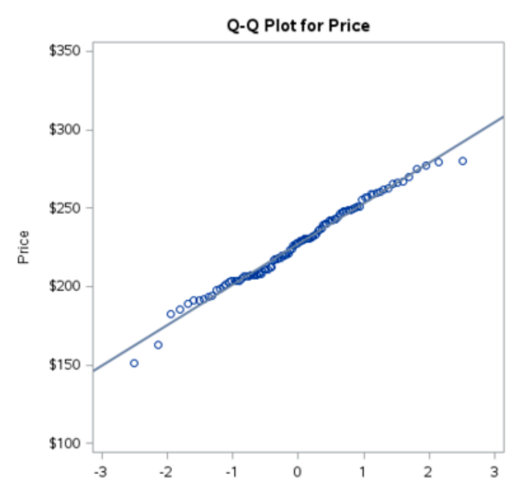
I used SAS to create the **Partial Autocorrelation Function (PACF)** graphs for the closing Prices of Period 1 (*left*), Period 2 (*right*).



Though Period 1's ACF appears out of control in the negative region, Period 2's ACF decay is more linear than exponential and suggesting either an *autoregressive* parameter (p = 1) or an ARMA model be used (p =1, q =1). The PACF graphs both cut off rather than tailing off, which suggests a strict *autoregressive* model approach.

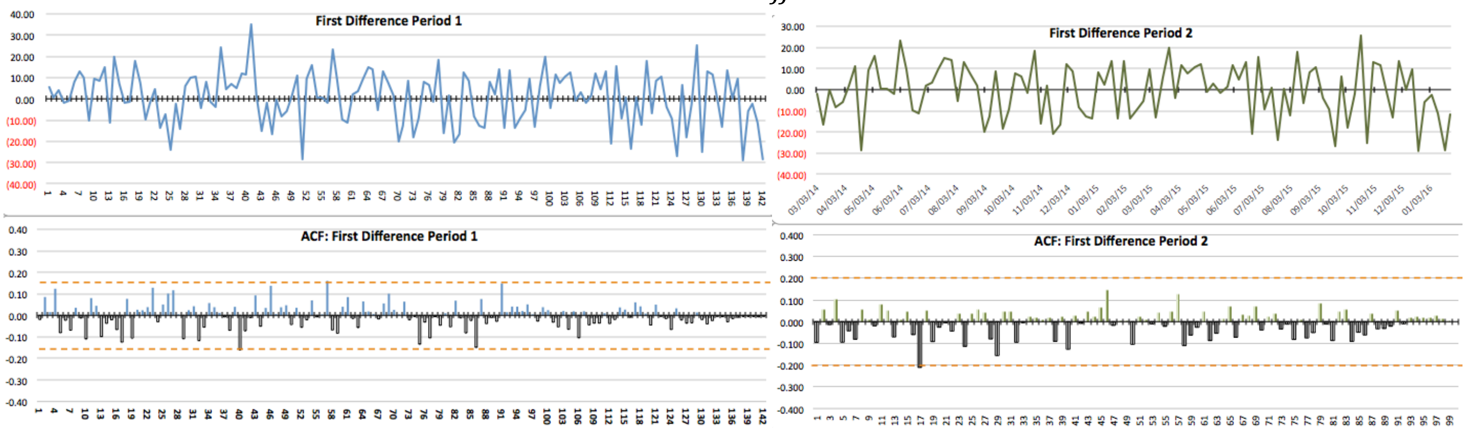


I used SAS to create Normal Quantile (Q-Q) plots to understand the distribution of the prices. Period 1 (*left*) appears heavy tailed but random. Period 2 (*right* follows a Normal Distribution quite well.



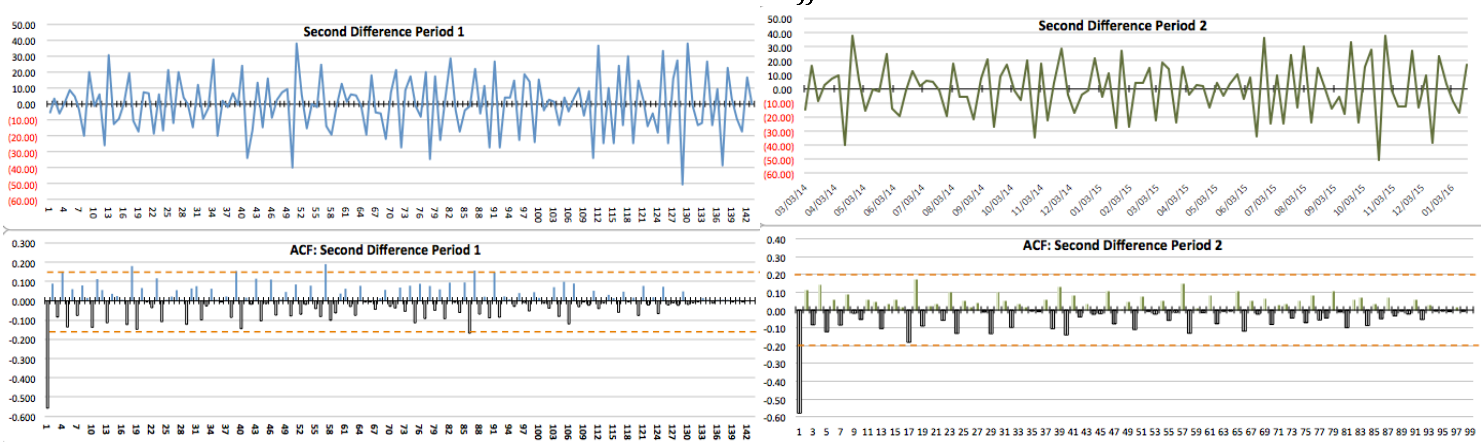
As for a stationary series, the ACF should decay to zero. More so, to be stationary, as the lags increase these values should oscillate within 2 approximate standard errors of the sample autocorrelations, $\pm 2/\sqrt{n}$ (here Period 1 = ±0.167; Period 2 = ±0.199; the *orange dashed lines* shown in every *correlogram*). Since neither Period 1 (*left*) nor Period 2's data (*right*) have autocorrelations that oscillate and stabilize around the zero-axis, they are both non-stationary. But I must make this time series stationary in order to produce a model that can forecast future terms accurately. So I will perform transformations on my data---First Difference and Second Difference (among other available options, i.e. logarithm and/or standardized residuals)---to make it stationary in order to diagnose what my model needs are.

First Differences



Period 2's First Difference ACFs oscillate near zero within the 2 standard error range giving us a new stationary set of data (with one exception), while Period 1's correlogram has various lags testing the limits of this range. I can see how using shorter period of data (especially with stocks) would be easier to find a trend to model. A further transformation is needed. Let's see a second difference and keep in mind that if we create an ARIMA(p,d,q) model, the number of differences, d should equal 1: ARIMA(p,1,q).

Second Differences



Notice in both correlograms, before the remaining lags oscillate near zero, the first lag is a negative value outside the error range (showing statistical significance). This is a sign of *overdifferencing* and that one difference will be enough in my model (d = 1), not two. The Period 1 ACF again, however, has later lags outside the error range, while Period 2 is contained within the 2 standard error range. Seeing that we need a stationary series to model and Period 2's Q-Q plot (*on page* 2) shows a Normal Distribution, let's focus on Period 2 for the rest of the project.

Overall, I now have a clearer and general idea of what type of model I should use for Period 2: ARI(1,1) a *differenced first-order autoregressive model*:

$$\hat{Y}_t = \mu + Y_{t-1} + \phi(Y_{t-1} - Y_{t-2}) + e_t$$

or equivalently,

$$\hat{Y}_t = \mu + (1+\phi)Y_{t-1} - \phi Y_{t-2} + e_t$$

Step 3 - *"Creating the Model"* - Let's try a few different models and see what works best, even though we have a hunch that an ARIMA(1,1,0) should be better than most.

AR(1): I used Excel's regression analysis and lag my series by one, letting (Y, X) become (Y_t, Y_{t-1}).

Regression Statistics

R	0.87559
R-square	0.76665
Adjusted R-square	0.76427
S	12.59399
N	100

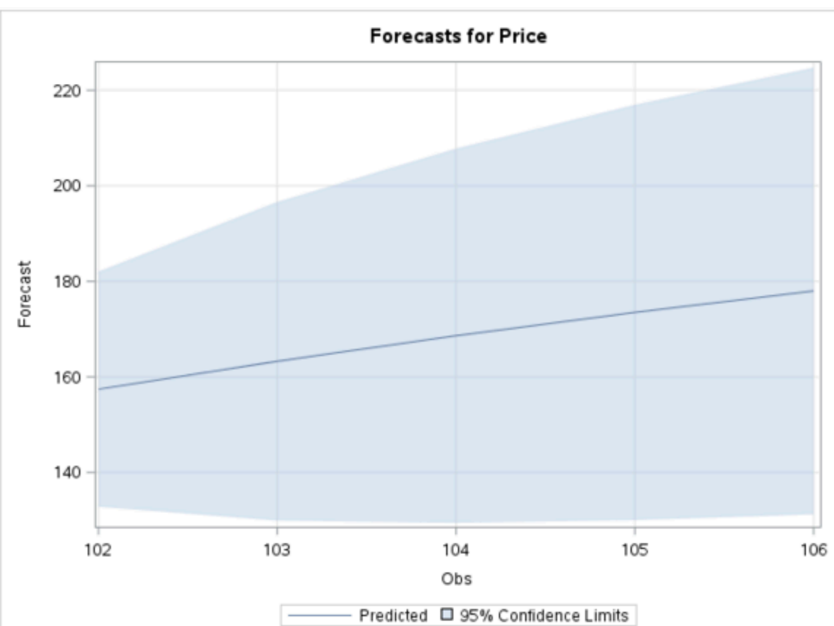
ANOVA

	d.f.	*SS*	*MS*	*F*	*p-level*
Regression	1.	51,067.51337	51,067.51337	321.9718	0.
Residual	98.	15,543.6479	158.60865		
Total	99.	66,611.16128			

	Coefficient	*Standard Error*	*Lower 95%*	*Upper 95%*	*t Stat*	*P-value*
Intercept	18.18871	11.71291	-5.05518	41.43261	1.55288	0.12368
Y_t-1	0.9167	0.05109	0.81532	1.01808	17.94357	0.

	Box Pierce Q Statistic					
		sumproduct		*sum*100/*	*chi2*	*chi2*
lag	*sumproduct*	*squared*	*sums*	*um-rsd-square*	*inv*	*dist*
1	-615.5	378824.9	378824.9	0.16		
2	1421.4	2020241.0	2399065.8	0.99	2.71	0.32
3	277.9	77238.8	2476304.6	1.02	4.61	0.60
4	1919.1	3682836.8	6159141.4	2.55	6.25	0.47
5	-1038.8	1079149.9	7238291.4	3.00	7.78	0.56
6	-462.8	214167.1	7452458.5	3.08	9.24	0.69
7	-1001.8	1003591.3	8456049.8	3.50	10.64	0.74
8	969.2	939412.6	9395462.4	3.89	12.02	0.79
9	328.6	107987.1	9503449.5	3.93	13.36	0.86
10	-163.9	26873.2	9530322.7	3.94	14.68	0.92
11	1214.4	1474777.8	11005100.5	4.55	15.99	0.92
12	887.9	788454.4	11793554.9	4.88	17.28	0.94
13	-886.6	786055.0	12579609.9	5.21	18.55	0.95
14	68.3	4660.3	12584270.2	5.21	19.81	0.97
15	744.7	554625.5	13138895.7	5.44	21.06	0.98

The R-square (0.767) and Adjusted R-square (0.764) values cannot be seen as good or bad in modeling a stock with much volatility. The AR(1) equation becomes $\hat{Y}_t = 18.189 + 0.917Y_{t-1} + e_t$. The Durbin Watson statistic for this AR(1) is 2.06, which is very close to 2, showing no serial correlation among the residuals. The Box Pierce Q statistics give p-values (in the Chi^2 Distribution) well above 10%, so we don't reject the null hypothesis. A white noise process exists in this AR(1) model. From SAS's tool for producing an AR(1) model, I gained the following 5 forecast estimates, standard errors, ranges, and the graph thereof:



Forecasts for variable Price

Obs	Forecast	Std Error	95% Confidence Limits	
102	157.4307	12.5564	132.8206	182.0408
103	163.2747	17.0149	129.9262	196.6232
104	168.6188	19.9939	129.4314	207.8061
105	173.5057	22.1800	130.0338	216.9776
106	177.9746	23.8547	131.2202	224.7289

The SAS output of 5 forecasted values fitting the **AR(1)** to the data (*above* and *left*). Trend = Upward. Note "Obs" represents the number of observations, and our time t started at zero (0). So really these values are forecasted as t = 101 (the 101st week), 102, 103, 104, 105 after our 100

AR(2): I used Excel's regression analysis again, lagging my series twice and setting (Y, X_1, X_2) to be (Y, Y_{t-1}, Y_{t-2}).

Regression Statistics

R	0.87586
R-square	0.76714
Adjusted R-square	0.76228
S	12.71101
N	99

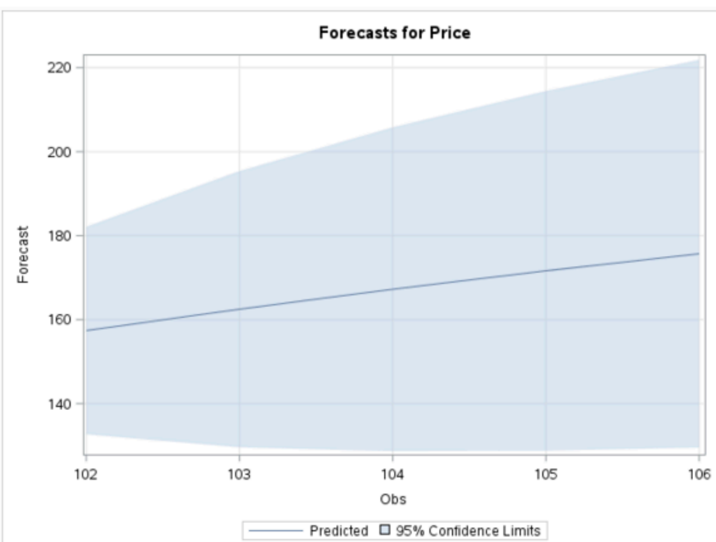
ANOVA

	d.f.	*SS*	*MS*	*F*	*p-level*
Regression	2.	51,097.38562	25,548.69281	158.12802	0.
Residual	96.	15,510.68858	161.56967		
Total	98.	66,608.0742			

	Coefficient	*Standard Error*	*Lower 95%*	*Upper 95%*	*t Stat*	*P-value*
Intercept	16.36207	12.51528	-8.48055	41.2047	1.30737	0.19421
Y_t-1	0.87719	0.1029	0.67293	1.08145	8.52439	0.
Y_t-2	0.04743	0.10675	-0.16447	0.25933	0.44428	0.65784

	Box Pierce Q Statistic					
		sumproduct		sum*99/	chi2	chi2
lag	sumproduct	squared	sums	m-rsd-squar	inv	dist
1	47.1	2214.2	2214.2	0.00		
2	1264.7	1599411.5	1601625.8	0.66	2.71	0.42
3	325.6	106041.1	1707666.9	0.70	4.61	0.70
4	1803.6	3253018.2	4960685.1	2.04	6.25	0.56
5	-1059.3	1122207.8	6082892.9	2.50	7.78	0.64
6	-624.8	390348.1	6473241.0	2.66	9.24	0.75
7	-1052.0	1106611.9	7579852.9	3.12	10.64	0.79
8	920.2	846747.3	8426600.1	3.47	12.02	0.84
9	361.3	130525.1	8557125.2	3.52	13.36	0.90
10	-115.4	13317.8	8570443.0	3.53	14.68	0.94
11	1258.9	1584901.8	10155344.8	4.18	15.99	0.94
12	913.5	834405.0	10989749.7	4.52	17.28	0.95
13	-827.6	684950.9	11674700.7	4.80	18.55	0.96
14	79.4	6296.9	11680997.6	4.81	19.81	0.98
15	739.4	546647.6	12227645.2	5.03	21.06	0.99

The R-square (0.767) and Adjusted R-square (0.762) values are nearly the same as AR(1)'s. The AR(2) equation becomes $\hat{Y}_t = 16.362 + 0.877Y_{t-1} + 0.047Y_{t-2} + e_t$. The Durbin Watson statistic for this AR(2) is 1.96, which is again very close to 2, showing no serial correlation among the residuals. And the Box Pierce Q statistics again gives *p*-values (in the Chi^2 Distribution) well above 10%, so we don't reject the null hypothesis. A white noise process exists in this AR(2) model. I also used SAS to produce an AR(2) model forecast for Period 2.



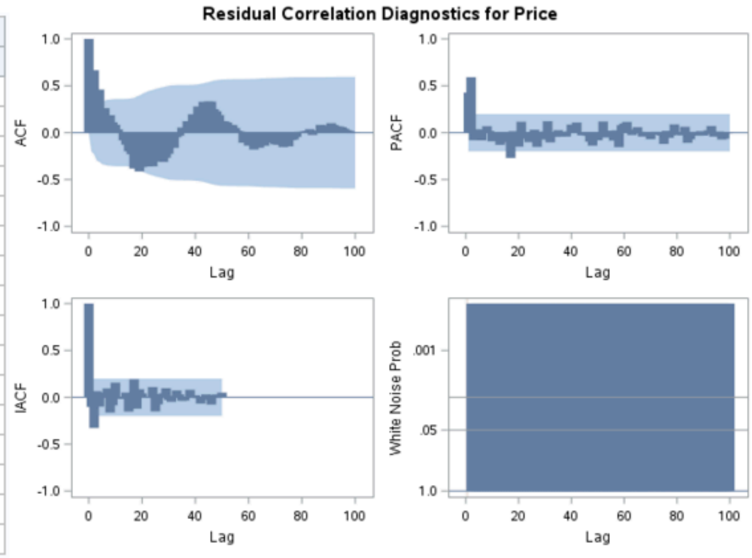
Forecasts for variable Price

Obs	Forecast	Std Error	95% Confidence Limits	
102	157.4026	12.6095	132.6884	182.1168
103	162.4927	16.7832	129.5982	195.3872
104	167.2382	19.6776	128.6707	205.8056
105	171.6261	21.8514	128.7981	214.4541
106	175.6851	23.5527	129.5227	221.8476

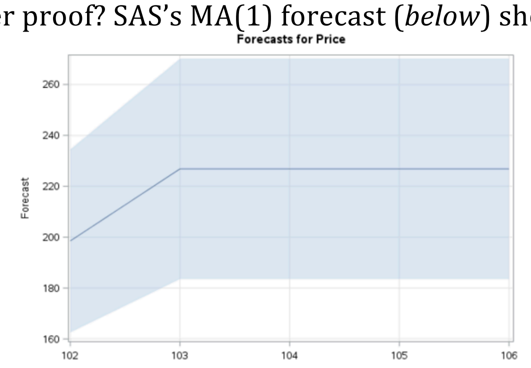
The SAS output of 5 forecasted values fitting the **AR(2)** to the data (*above* and *left*). Trend = Upward. Note "Obs" represents the number of observations, and our time *t* started at zero (0). So really these values are forecasted as *t* = 101 (the 101st week), 102, 103, 104, 105 after our 100 weeks of Period 2.

MA(1): Reaffirming my choice of an autoregressive parameter instead of a moving average, my SAS-produced MA(1) model that rejects the null hypothesis and does *not* create a white noise process:

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	111.87	5	<.0001	0.426	0.666	0.325	0.460	0.215	0.261
12	120.29	11	<.0001	0.141	0.186	0.111	0.084	0.035	0.005
18	163.60	17	<.0001	-0.082	-0.127	-0.199	-0.230	-0.382	-0.299
24	265.09	23	<.0001	-0.414	-0.367	-0.359	-0.361	-0.289	-0.359
30	328.68	29	<.0001	-0.257	-0.307	-0.263	-0.312	-0.268	-0.226
36	336.79	35	<.0001	-0.139	-0.162	-0.048	-0.033	0.054	0.041
42	374.14	41	<.0001	0.122	0.075	0.195	0.145	0.279	0.245
48	473.43	47	<.0001	0.329	0.302	0.331	0.330	0.250	0.228
54	491.81	53	<.0001	0.188	0.122	0.120	0.108	0.079	0.081
60	501.86	59	<.0001	0.050	0.019	0.011	-0.107	-0.084	-0.138
66	535.48	65	<.0001	-0.092	-0.179	-0.153	-0.160	-0.109	-0.139
72	558.31	71	<.0001	-0.075	-0.107	-0.071	-0.124	-0.128	-0.118
78	588.33	77	<.0001	-0.153	-0.133	-0.147	-0.078	-0.074	-0.046
84	589.80	83	<.0001	0.009	0.008	0.012	0.017	0.036	-0.026
90	604.99	89	<.0001	0.020	0.001	0.082	0.041	0.077	0.060
96	644.19	95	<.0001	0.100	0.081	0.064	0.064	0.058	0.041

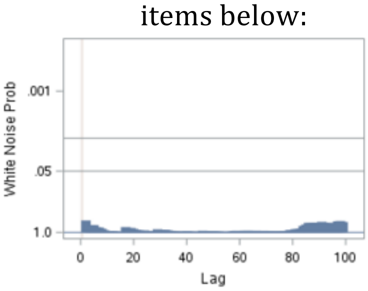


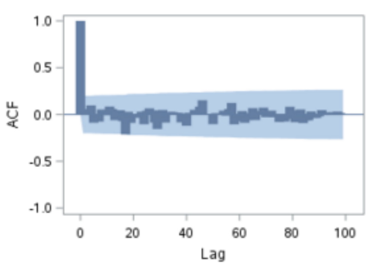
Notice the 'Autocorrelation Check for Residuals' table (*above left*) produces Pr < ChiSq probabilities all at the <0.0001 level, making the 'White Noise Prob' graph (*above, bottom right*) become a full blue rectangle (since the graph goes up from 1.0 to 0.001 and measures statistical significance). Also, the ACF of MA(1)'s 'Residual Correlation Diagnostics for Price' graph (*above, top left*) is similar to Period 2's original correlogram (*page 2, top right*), further telling us MA(1) is not the way to go. Need further proof? SAS's MA(1) forecast (*below*) shows a flat line for Obs 103, 104, 105 and 106.



Forecasts for variable Price				
Obs	Forecast	Std Error	95% Confidence Limits	
102	198.5314	18.3516	162.5630	234.4998
103	226.8460	22.1005	183.5299	270.1622
104	226.8460	22.1005	183.5299	270.1622
105	226.8460	22.1005	183.5299	270.1622
106	226.8460	22.1005	183.5299	270.1622

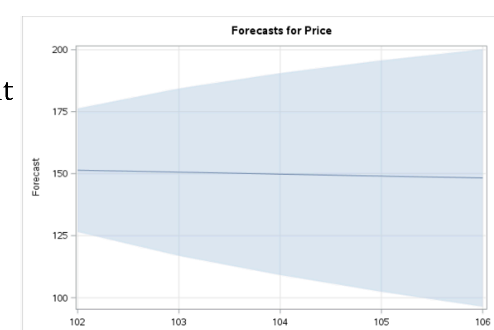
Moving on, looking at SAS's model for an IMA(1,1), that includes a first difference, a moving average parameter makes more sense---comparing the same MA(1) items highlighted above to the IMA(1,1) items below:





Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2.50	5	0.7758	-0.005	0.057	0.002	0.097	-0.088	-0.054
12	4.72	11	0.9440	-0.076	0.053	0.015	-0.008	0.085	0.059
18	11.70	17	0.8178	-0.061	0.008	0.046	-0.072	-0.213	0.028
24	14.57	23	0.9093	-0.090	-0.030	-0.003	-0.036	0.029	-0.104

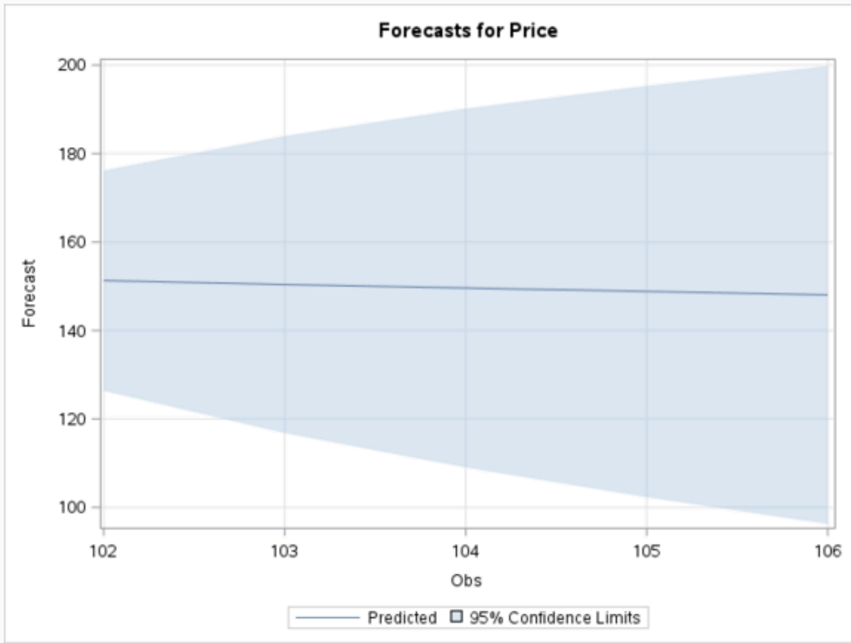
Further notice MA(1)'s Obs 102 forecast unrealistically starts at 198.53 after Period 2's last actual value (Obs 101) equals 151.04, while the IMA(1,1) below has Obs 102 = 151.39.



Forecasts for variable Price				
Obs	Forecast	Std Error	95% Confidence Limits	
102	151.3965	12.7116	126.4823	176.3107
103	150.6107	17.2195	116.8610	184.3604
104	149.8249	20.7711	109.1142	190.5356
105	149.0392	23.7985	102.3949	195.6834

ARIMA(1,1,0): Finally, the *differenced first-order autoregressive model* where my findings pointed me to. It is also known as an ARI(1,1): $\hat{Y}_t = \mu + Y_{t-1} + \phi(Y_{t-1} - Y_{t-2}) + e_t$

Using SAS, I came up with $\hat{Y}_t = -0.7889 + Y_{t-1} - 0.09535(Y_{t-1} - Y_{t-2}) + e_t$



Forecasts for variable Price				
Obs	**Forecast**	**Std Error**	**95% Confidence Limits**	
102	151.2781	12.7060	126.3749	176.1813
103	150.3913	17.1337	116.8098	183.9728
104	149.6117	20.6968	109.0468	190.1767
105	148.8219	23.7254	102.3211	195.3228
106	148.0331	26.4093	96.2718	199.7945

The SAS output of 5 forecasted values fitting the **ARIMA(1,1,0)** to the data (*above* and *left*). Trend = Downward (slightly). Note "Obs" represents the number of observations, and our time *t* started at zero (0). So really these values are forecasted as *t* = 101 (the 101st week), 102, 103, 104, 105 after our 100 weeks of Period 2.

Conclusion

After seeing Period 2's original data's correlogram tail off, its PACF cutting off, and its first difference making the series stationary, I am convinced an ARIMA(1,1,0) model would be the best available option. These clues were fun to follow as my first time series model developed. This ARI(1,1) forecasted values, however, have a downward trajectory. My conclusion for investing my part of my imaginary retirement nest egg in Tesla Motors at this time is "No, it is too risky". I know my model is not perfect, but of the options I have shown---AR, MA, ARI, IMA---the analysis I performed pointed to a specific model (ARI). I know stocks are volatile, and Tesla has been on a roller coaster the past year (52 weeks; *see below)*. Following the principle of parsimony that the simplest solution is best, the simple clues I followed make me confident in what I learned and found in my data. Though I know there is no right or wrong answer here, I do have more clarity on the time series techniques I touched upon.

This project has been a powerful way to piece most of the material in the course together with a topic that interests me. My final verdict for Tesla Motors is that there is a pattern in Period 2 (the last 100 weeks), but the model does not have enough concrete evidence that "it should" go up. Instead there is evidence that it cannot be definitively predicted "in the long run" at this time whether it will go up or down. And that's not good for an investor. I will stay away. Too risky!