PROJECT SUBMISSION - FINAL

Presented by: Phillip Béliveau (40098503) and Matthew Cuyler (40124941)

Presented to: Professor Mohsen Farhadloo

Concordia University
BSTA 450 – Section A

Introduction

When it comes to technical data analysis, there is no topic more relevant than the prediction of prices in the equity market. The ways with which individuals make these predictions will always vary, however, in the scope of purely statistical models for data analysis, the ARIMA and ARIMA(X) approach is very common. Our project attempts to address why using this prevalent method in its traditional form is vastly unsuccessful in accurately predicting a future closing stock price. It may, in fact, be impossible.

In doing research on different forecasting methods, we found several written-up attempts on Medium.com from those who have used ARIMA and ARIMA(X) modelling in similar ways. In all cases, we deduced that using the traditional, textbook form of autoregression is unsuccessful because most of the classic assumptions are not usually met. This is most likely as a result of the stock market's extreme complexity.

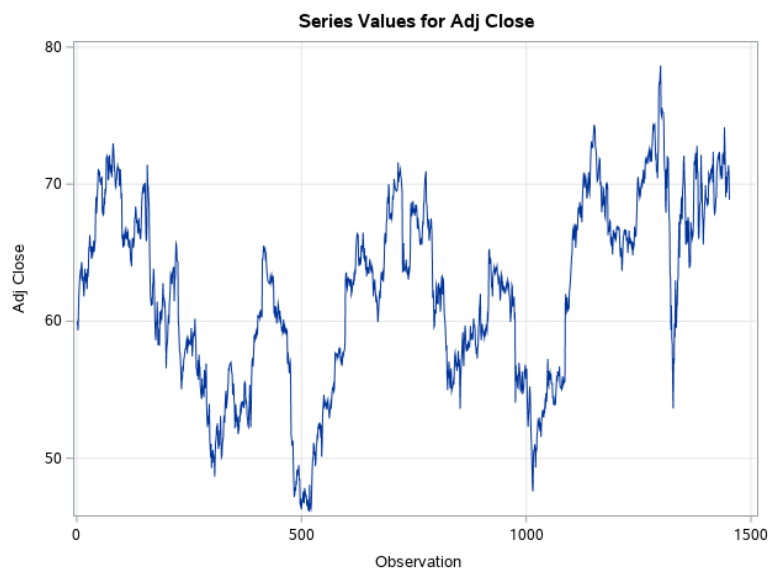
Nonetheless, we will conduct a series of exploratory tests on our own data set, and see for ourselves the effectiveness of a traditional ARIMA(X) model. We will assume a 5% significance level for all hypothesis testing.

Phase 1 Data Exploration:

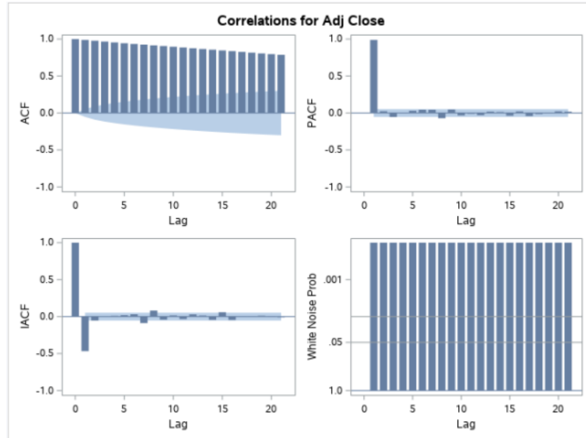
Original Data Series

A total of 1,452 observations were used.

Selected Series Range: January, 2016 – November, 2020



Next, we must stabilize the series. Our ACF shows a consistent decreasing pattern but stays well above 0. This is important because it gives us an indication of the time series stationarity.



<u>Target:</u>

The Adjusted Closing Price of Apple stock in United States Dollars (USD).

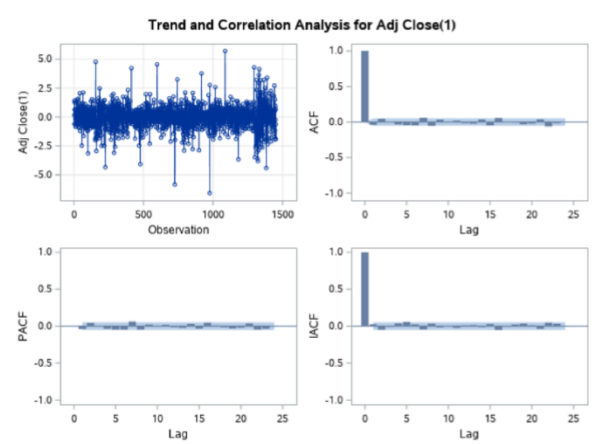
<u>Variables to be tested:</u>

We will be testing correlation on the following:

- A WilliamsR indicator, MACDhist, quantified stock momentum values, an Ultimate Oscillator, DMI+ (Directional movement index), DMI-, RSI over last 10 days, ATR (Average True Range), TSF (Total Social Financing), and RSI_Bollinger.

All of these variables have been standardized as they were not on the same range.

Phase 2 Stationarity:



Here we test two transformations: a log transformation and a square root transformation. The results had no serious impact stabilizing our series. Thus, we must apply a difference of one. (Lag =1)

<u>Difference of One:</u>

As visible here, our differencing has somewhat stabilized the series, as there is no lag spike moving over or under the blue interval.

We can now check if this D of 1 will fill the assumptions of white noise and residuals by testing for the autocorrelation in the residuals.

(We additionally tried testing a difference of 2, but we were unsuccessful in adding stability.)

<u>Autocorrelation for white noise:</u>

The p-values are very low, meaning that in order to remove significant autocorrelation in the model, we must add terms such as autoregression (AR) and moving average (MA). Although, when looking at the ACF and PACF plot, no spike or patterns are being perceived, thus no AR or MA terms seem suited for the job. We will dive deeper into those terms in the phase 7.

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	12.92	6	0.0444	-0.039	0.047	0.001	-0.033	-0.043	-0.047
12	25.11	12	0.0143	0.061	-0.054	0.034	-0.005	0.021	-0.015
18	35.93	18	0.0072	-0.022	0.037	-0.047	0.057	0.001	-0.007
24	46.14	24	0.0043	-0.030	-0.021	0.038	-0.062	-0.014	-0.000

<u>Autocorrelation for residuals (</u>Ljung-Box test)<u>:</u>

H_0: The model does not show lack of fit

H_a: The model does show a lack of fit.

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	12.92	6	0.0444	-0.039	0.047	0.001	-0.033	-0.043	-0.047
12	25.11	12	0.0143	0.061	-0.054	0.034	-0.005	0.021	-0.015
18	35.93	18	0.0072	-0.022	0.037	-0.047	0.057	0.001	-0.007
24	46.14	24	0.0043	-0.030	-0.021	0.038	-0.062	-0.014	-0.000
30	49.34	30	0.0145	0.013	-0.000	-0.013	0.035	0.023	-0.009
36	57.16	36	0.0139	-0.020	-0.005	-0.003	0.062	0.010	0.031
42	60.02	42	0.0352	-0.004	0.004	0.028	-0.023	-0.014	-0.019
48	63.53	48	0.0659	-0.016	-0.017	0.002	-0.038	-0.007	-0.017

A significant p-value ($p<0.05$) in this test rejects the null hypothesis that the time series is not autocorrelated, or does not show lack of fit. Therefore, we reject the null hypothesis at each lag from 6 to 42, that the series isn't autocorrelated and that the series still have autocorrelation in the residuals.

(We perform this analysis at this stage, as those assumptions of no autocorrelation in the residuals and that the series proves to be white noise needs to be fill before testing for the significance of the independent variables. Without those assumptions fill, we cannot assure the significance of our independent variables on the dependent variable.) On top of looking at the plot, residuals, and white noise. We must perform a Dicker-Fuller Test to test if our series is stationary, and test the existence of a unit root. Even though there is still autocorrelation in the white noise and residuals, we will perform the test and continue as planned.

<u>Dicker-Fuller Test:</u>

Ho: The series is non-stationary. A unit root is present in the autoregressive time-series model.

Ha: The series is stationary. There are no unit roots present in the series

We will Reject Ho if $p<0.005$

Here we are analyzing the p-values at Single Mean: Pr < Tau

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-0.0331	0.6755	-0.05	0.6651		
	1	-0.0203	0.6784	-0.03	0.6717		
	2	-0.0294	0.6763	-0.05	0.6673		
	3	-0.0525	0.6710	-0.09	0.6546		
	4	-0.0583	0.6697	-0.10	0.6502		
	5	-0.0492	0.6718	-0.09	0.6541		
	6	-0.0409	0.6737	-0.08	0.6579		
Single Mean	0	-15.4076	0.0360	-2.74	0.0680	3.79	0.0996
	1	-14.3515	0.0467	-2.63	0.0871	3.50	0.1728
	2	-15.9992	0.0311	-2.78	0.0622	3.89	0.0930
	3	-16.2487	0.0292	-2.79	0.0611	3.90	0.0923
	4	-15.1689	0.0382	-2.68	0.0785	3.61	0.1449
	5	-13.8491	0.0529	-2.55	0.1045	3.27	0.2322
	6	-12.6077	0.0718	-2.42	0.1355	2.95	0.3128

We notice that these p-values for each lag are all greater than α = 0.05 and we therefore, at the 5% level of significance, do not have enough evidence to confirm that no unit roots are present in our time-series. Thus, our series is still not stationary, even after trying the following combinations:

- Normal (0,0,0) (0,1,0) (0,2,0)
- Log (0,0,0) (0,1,0) (0,2,0)
- Square-Root (0,0,0) (0,1,0) (0,2,0)

The selected combination is: Normal (0,1,0)

This phase does not support the application of the terms, as we wanted to stabilize the series as much as possible to then be able to test the significance of the X variables on the independent variables.

Stationarity in X variables:

After the decision of applying a differencing to my Adj closing price variables, we apply the same transformation (D of 1) in the X variables, so that we get rid of the noise in them, thus stabilizing the variables.

Phase 3 Causality:

The Granger Test of Causality must be performed to examine: "The directions of significant causality between an exogenous variable and the dependent variable (…). If reverse causality is detected, the exogenous variable must be removed from the pool of independent-variable candidates. This test must be performed on the dependent and independent variable in their current form (untransformed or transformed)". (1)

For the time being, we cannot perform this analysis as we do not know how to perform it on SAS.

Hypothesis and test:

Any variable with a p-value below 0.05 led to rejection of the null hypothesis of no reverse causality. Given the p-value of the test, it will lead us to reject or not the hypothesis of reverse causality and force us to remove certain variable. This test arrives in the third phase, following the second in which we decided to Difference the series by 1 for the dependent and all the independent variables. It is important to do it in order, as this test needs to be performed in the transformed (current) form of the variables.

Phase 4 Correlation:

Next, we must perform correlation analysis on the transformed dependent variable with our transformed exogenous variables. This is the first screening of our X variables. Features that do not display significant correlation with p-value less than 0.05, will be removed. Here are the selected exogenous variables:

Pearson Correlation Coefficients, N = 1451 Prob > |r| under H0: Rho=0	
	Adj Close
Standardized_willr willr	0.74690 <.0001
Standardized_macdhist macdhist	0.74232 <.0001
Standardized_momentum momentum	0.70996 <.0001
Standardized_Ultimate oscillator Ultimate oscillator	0.58215 <.0001
Standardized_DMI + DMI +	0.69391 <.0001
Standardized_DMI - DMI -	-0.69321 <.0001
Standardized_rsi_col10 rsi_col10	0.90788 <.0001

Noticeably, our results display high correlation and very low p-values. Therefore, there is a high level of significance from our variables.

H_0: There is no significant linear relationship between x and y.

H_a: There is a significant linear relationship between x and y.

We will reject Ho if $p<0.05$

Decision:

These 7 variables all display a significance of $p=0.0001<0.05$. Thus, we can reject Ho, stating <u>a strong significance</u> between the variables.

Phase 5 VIF:

In this phase, we conduct an examination of Variance Inflation Factors, or VIFs. These values measure the degree of multicollinearity in a set of variables in multiple regression. When highly collinear relationships are present, there is no reduction in the explanatory power of the model, but there is a reduction in the statistical significance of the independent variables. Axiomatically, this creates an issue. It is also important to note that the acceptable threshold for VIF values is highly subjective. Some argue that VIF values just under 10 are acceptable, while some take a more conservative approach and believe 5 to be a VIF maximum. For our analysis, we will stick with the rule of thumb and use 10 as our VIF threshold. Therefore, with all our variables maintaining a variance inflation value of less than 10, we may proceed with caution.

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	0.00580	0.00936	0.62	0.5356	0
Standardized_willr	willr	1	-0.05462	0.03424	-1.60	0.1109	4.52272
Standardized_macdhist	macdhist	1	0.56179	0.05493	10.23	<.0001	2.47299
Standardized_momentum	momentum	1	0.26050	0.02813	9.26	<.0001	2.00648
Standardized_Ultimate oscillator	Ultimate oscillator	1	-0.00188	0.02799	-0.07	0.9465	1.95288
Standardized_DMI +	DMI +	1	0.18736	0.03591	5.22	<.0001	2.18094
Standardized_DMI -	DMI -	1	-0.22396	0.03985	-5.62	<.0001	2.22396
Standardized_rsi_col10	rsi_col10	1	1.37660	0.05983	23.01	<.0001	8.92761

Phase 6 Selection:

Thus far, we have assessed the correlation significance in which we eliminate a substantial number of variables, and we have conducted VIF analysis to find signs of multicollinearity. Next, we may perform a “Stepwise” selection process, which is characterized by the ability to remove variables as we add others. In other words, we start by adding the variable with the lowest p-value to the model, then we add the second and third. As predictors increase, the value of the first p-value changes.

We use the Akaike information criterion to select our best model, the AIC is the conventional selection criteria for time series analysis. It is a: “mathematical method for evaluating how well a model fits the data it was generated from. In statistics, AIC is used to compare different possible models and determine which one is the best fit for the data.” (2)

Here are our results:

Stepwise Selection Summary											
Step	Effect Entered	Effect Removed	Number Effects In	Model R-Square	Adjusted R-Square	AIC	AICC	BIC	CP	SBC	PRE
0	Intercept		1	0.0000	0.0000	1429.0316	1429.0399	-25.2251	9757.8079	-18.6884	1427.22
1	Standardized_rsi_col		2	0.8242	0.8241	-1091.7201	-1091.7035	-2543.9122	522.7207	-2534.1601	252.22
2	Standardized_macdhis		3	0.8550	0.8548	-1369.1270	-1369.0993	-2820.7914	179.7112	-2806.2870	208.88
3	Standardized_momentu		4	0.8634	0.8631	-1453.5753	-1453.5337	-2904.9974	87.7407	-2885.4552	197.48
4	Standardized_DMI -		5	0.8680	0.8676	-1500.7451	-1500.6869	-2951.9400	38.7377	-2927.3450	192.01
5	Standardized_DMI +		6	0.8710	0.8706	-1532.6213	-1532.5437	-2983.5764	6.5908	-2953.9412*	188.42
6	Standardized_willr		7	0.8712	0.8707*	-1533.2195*	-1533.1197*	-2984.1420*	6.0045*	-2949.2595	188.44
* Optimal Value of Criterion											

The selected model:

Selected Model

The selected model, based on AIC, is the model at Step 6.

Effects:	Intercept Standardized_willr Standardized_macdhis Standardized_momentu Standardized_DMI + Standardized_DMI - Standardized_rsi_col

Note: The p-values for parameters and effects are not adjusted for the fact that the terms in the model have been selected and so are generally liberal.

Now knowing that this set of variables should give us the lower AIC, we will use them to perform my first ARIMAX model. We will have to come back to this step, because we will have to find a new set of variables that fit the assumptions of the model, and when we will mix those variables with different AR and MA term, it will change the significance of some of them. Thus, we will have to remove some probably.

Following the selection process, we can now perform proper statistical procedure looking the t-value of each variable.

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.005801	0.009359	0.62	0.5355
Standardized_willr	1	-0.054832	0.034084	-1.61	0.1079
Standardized_macdhis	1	0.560852	0.053106	10.56	<.0001
Standardized_momentu	1	0.260624	0.028059	9.29	<.0001
Standardized_DMI +	1	0.187730	0.035464	5.29	<.0001
Standardized_DMI -	1	-0.224453	0.039143	-5.73	<.0001
Standardized_rsi_col	1	1.375429	0.057234	24.03	<.0001

Example Hypothesis: Standardized_Momentum

H_0: $\beta_1 = 0$

H_a: $\beta_1 \neq 0$

Decision Rule

We will reject H_0 if $t_{calculated} > t_{(0.025,1452)}$ - or - $t_{calculated} < -t_{(0.025,1452)}$

We do not reject H_0 if $-t_{(0.025,1452)} <= t_{calculated} <= t_{(0.025,1452)}$

$t_{(0.025,1452)} = 1.96$ (Calculated using t-table, our degree of freedom is large enough however, to assume a Z-value of 1.96.)

Test Statistic

A t-Test is being performed.

$t_{calculated}$ = (β_1 – 0 = estimate) / ($S*\beta_1$ = standard error) → 0.260624 / 0.028059 = 9.29

Decision

$t_{calculated} > t_{(0.025,1452)}$ → 9.29 > 1.96

Thus, we reject H_0.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	1241.74378	206.95730	1628.42	<.0001
Error	1444	183.51884	0.12709		
Corrected Total	1450	1425.26262			

Root MSE	0.35650
Dependent Mean	0.00625
R-Square	0.8712
Adj R-Sq	0.8707
AIC	-1533.21952
AICC	-1533.11966
BIC	-2984.14201
C(p)	6.00451
PRESS	188.44204
SBC	-2949.25946
ASE	0.12648

Conclusion

Based on the results of our hypothesis test, we can conclude that there is a linear relationship between standardized stock momentum and our target variable, adjusted closing price. We can see that the X variables seems to be doing a good job in explaining 87% the variation in the model with an adjusted R-squared of 0.87.

Phase 7 Assumption of stationarity:

In ARIMAX, the significance level of the independent variables is calculated under the assumption that the residuals simulate white noise. To properly make the assessment of significance, the residuals must be tested for stationarity through the Dickey-Fuller test of phase 2 and for autocorrelation through the Ljung-box test of phase 2 also.

Result Summary of the 2 Tests:

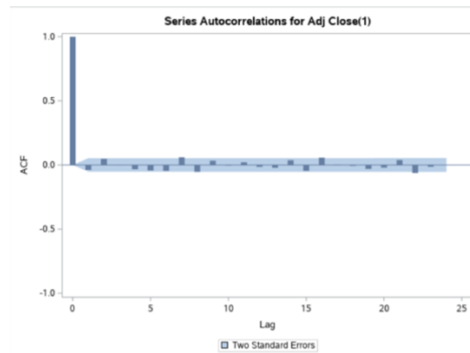
1. Dickey-Fuller has been proven non-stationary, which might confirm the possibility of autocorrelation in the residuals.
2. We then do a Ljung-Box test, in which we conclude to reject the null hypothesis that the series isn't autocorrelated and that the series still have autocorrelation in the residuals.

Conclusion:

Because those 2 assumptions are not meet, we cannot trust the significance level of our recently tested X variables. Thus, we will have to attempt to model the series through the addition of AR and MA terms. (This sort of conclusion is reinforcing our initial hypothesis that we cannot use ARIMAX to answer this difficult task of stock market prediction.)

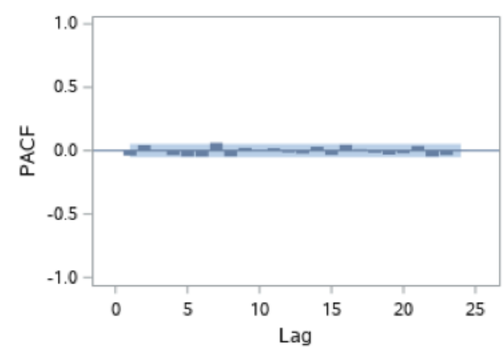
Phase 8 AR and MA terms:

We can add autoregressive and moving average terms and see if we can reduce the noise in the series. An AR terms for a dependent variable is a lagged value of that same variable, this lagged value has a significant relationship with its most recent value. On the other hand, MA terms are the residuals from the previous estimates. (1). The AR and MA terms are being decided based on the value of autocorrelation at each lag of the ACF and PACF.



Looking at the ACF plot:
The MA terms is concern about the ACF plot. We see here that there is no significant spike that would lead us to add an MA term.

Let's look at the PACF Plot:



On the other hand, the AR terms is concern about the PACF plot. Again, no significant spike at any lag, which seems to tell us that this series does not have any seasonality or relationship between his data points. As time series can be understood as the dependence of the data point in a time order.

Although this analysis is not truly relevant, as it is more of a trial an error in finding the best combination of terms to stabilize as much the series. On SAS, when trying around 15 different combinations of AR and MA terms, we successfully get p-value > 0.05 for lag [6, 18] in the Ljung-Box test (see results below). Meaning there is no presence of autocorrelation in the residuals in those lags.
Here are the results of an ARIMA (2, 1, 3) . AR = 2, I = 1, MA = 3.

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	0.82	1	0.3662	0.000	-0.004	0.004	0.002	-0.009	-0.021
12	12.41	7	0.0879	0.057	-0.063	0.012	-0.007	0.021	-0.010
18	21.61	13	0.0617	-0.016	0.034	-0.043	0.053	0.003	-0.014
24	31.90	19	0.0321	-0.036	-0.016	0.040	-0.059	-0.019	0.002
30	35.82	25	0.0743	0.013	-0.006	-0.017	0.036	0.029	-0.007
36	43.74	31	0.0642	-0.020	-0.006	0.002	0.064	0.014	0.025
42	46.93	37	0.1269	-0.007	0.005	0.031	-0.020	-0.017	-0.020
48	50.71	43	0.1957	-0.018	-0.016	0.000	-0.038	-0.010	-0.020

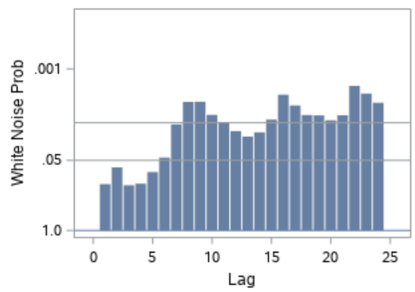


Figure 1

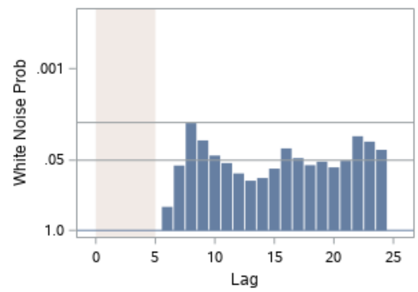


Figure 2

The white noise prob (fig 2) shows that our residuals have a higher probability of being random compared to before adding the terms (fig 1).

In the Q-Q plot, there is no improvement, what we would like to see if all the residuals align on the line. Which makes us thing our residuals are not white noise. We can clearly see autocorrelation.

Conclusion: We somewhat stabilize the series as much as we could without adding the X variables in it. When adding the variables, we might face challenges as it will change the model. Thus, we would have to loop over the process so to find which variables will fit properly the model in combination of the AR and MA terms.

Notice: In SAS studio, we can't seem to be able to perform variable selection process with my AR and MA terms, thus we cannot exactly find the good combination of variable with those terms.

Phase 9 Adding the variables to the model:

Using a combination of (2,1,3), let's see how our selected variables will behave on the model. Enormous amount of autocorrelation appears in the residuals given the 7 predictors.

Autocorrelation Check of Residuals

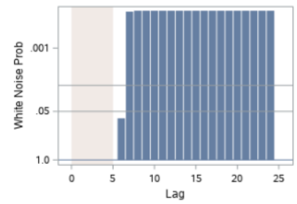
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.14	1	0.0764	-0.004	0.016	-0.018	0.002	0.002	-0.040
12	41.21	7	<.0001	0.102	-0.072	-0.041	0.093	-0.004	0.010
18	55.96	13	<.0001	-0.039	0.047	-0.065	0.031	0.030	-0.013
24	74.15	19	<.0001	-0.007	-0.035	0.032	-0.095	-0.023	0.023
30	78.46	25	<.0001	0.006	0.005	-0.021	0.030	0.037	-0.011
36	86.77	31	<.0001	0.015	-0.030	-0.025	0.024	-0.009	0.056
42	100.62	37	<.0001	0.038	-0.050	-0.049	-0.013	0.041	0.031
48	111.42	43	<.0001	-0.016	-0.037	0.006	-0.058	0.018	-0.043

The Ljung-Box test shows that the residuals possess autocorrelation. Meaning that the series is not stationary and can't be predict with those predictors. Thus, we tried different set of combination for the predictors and the terms. The AR and MA terms haven't change, we stayed with a (2, 1, 3) but the number of predictors yes, passing from 7 to 1.

Full 7 predictors:

AIC: 839

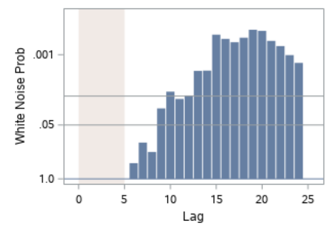
Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.14	1	0.0764	-0.004	0.016	-0.018	0.002	0.002	-0.040
12	41.21	7	<.0001	0.102	-0.072	-0.041	0.093	-0.004	0.010
18	55.96	13	<.0001	-0.039	0.047	-0.065	0.031	0.030	-0.013
24	74.15	19	<.0001	-0.007	-0.035	0.032	-0.095	-0.023	0.023
30	78.46	25	<.0001	0.006	0.005	-0.021	0.030	0.037	-0.011
36	86.77	31	<.0001	0.015	-0.030	-0.025	0.024	-0.009	0.056
42	100.62	37	<.0001	0.038	-0.050	-0.049	-0.013	0.041	0.031
48	111.42	43	<.0001	-0.016	-0.037	0.006	-0.058	0.018	-0.043



Trimming 1 Predictor with (2, 1, 3):

AIC: 2898

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	0.66	1	0.4175	-0.000	0.001	-0.007	0.013	-0.008	-0.012
12	18.57	7	0.0096	0.048	-0.015	0.071	0.052	0.022	0.038
18	37.18	13	0.0004	0.060	0.034	-0.069	0.028	0.028	-0.040
24	42.37	19	0.0016	-0.044	0.032	-0.001	-0.022	-0.001	0.009
30	45.28	25	0.0078	-0.032	-0.011	-0.011	0.014	0.023	0.002
36	49.75	31	0.0178	-0.034	-0.008	-0.003	0.012	-0.023	0.033
42	59.39	37	0.0112	0.063	0.001	-0.043	-0.024	0.003	0.002
48	66.03	43	0.0135	0.006	0.029	0.011	-0.020	0.021	-0.050



We see that by reducing the number of predictors, we can stabilize partially the series, but not enough as the assumption on white noise residuals is violated due to P-value<0.05. Looking at both AIC, the one with all predictors is considerably lower than the one with one predictor, thus here is better.

Conclusion:

We see that by using an ARIMAX instead of an ARIMA, we de-stabilize the series when adding new variables, which then delegitimize the significant of any exogenous variables that we add to the model due to the violation of the assumption of residuals being random. Thus, we will need to stick to an ARIMA model in the form (2,1,3) instead of an ARIMAX.

Phase 10 Assessing the significance of terms:

In this phase, we need to assess the significance of each parameter of our ARIMA model.

To assess the significance of each parameter in the ARIMA model, we conduct several T-tests, as shown in our SAS results above. The T-Test procedure for significance is the same here as used for each financial indicator listed in Phase 6. Therefore, we

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.0061609	0.02447	0.25	0.8012	0
MA1,1	1.16855	0.21218	5.51	<.0001	1
MA1,2	-0.71765	0.20253	-3.54	0.0004	2
MA1,3	0.08644	0.02719	3.18	0.0015	3
AR1,1	1.12805	0.21223	5.32	<.0001	1
AR1,2	-0.61924	0.20573	-3.01	0.0026	2

notice that the T-value for each of our ARIMA term parameters are all either > than $T_{critical}$ = 1.96 or they are < than $-T_{critical}$ = -1.96, and we can assume significance in all parameters tested.

Phase 11 Residual Assumption:

Now that we have chosen our model and made the prediction, we must see if the residuals satisfy the condition of normality through Kolmogorov-Smirnov test, and homoscedasticity through White's test.

K-S Test for Normality:

Ho: the residuals follow a normal distribution

Ha: the residuals do not follow a normal distribution

Decision Rule: Reject if $D > D_{critical}$

In this test we hope to fail to reject our null hypothesis and conclude that model residuals are normally distributed. We use test statistic D to measure the maximum vertical distance between our residual's distribution overlapped to a normal distribution. With N = 1452 and alpha = 0.05, our $D_{critical}$ would equal: (n cross α) / $\sqrt{n}$ = 0.0356

Conclusion: Due to stock market complexity and other data factors, the residuals of our model most likely do not follow a normal distribution. This adds to evidence against using ARIMA for stock market prediction.

White's test:

This test examines whether the variance of our data is approximately equal to the variance of our model. If the test is indeed significant, then our data is heteroscedastic.

Hypothesis:

H_0: The data is homoscedastic.

H_a: The data is heteroscedastic.

Decision Rule: If calculated Chi-Square value is greater than the Chi-Square critical, we would reject our null hypothesis in favor of heteroscedasticity, meaning most residuals vary largely from a fitted regression line, and there is large difference among variances.

Conclusion: Once again this test shows the difficulty in predicting stock market price using ARIMA, since it is incredibly difficult to ensure homoscedasticity in large time series data sets.

Phase 12 Model Performance:

Given the fact that on the phase 11, we haven't been able to meet the two assumptions about normality for the residuals, we have to say that the prediction is not viable and cannot be taken seriously.

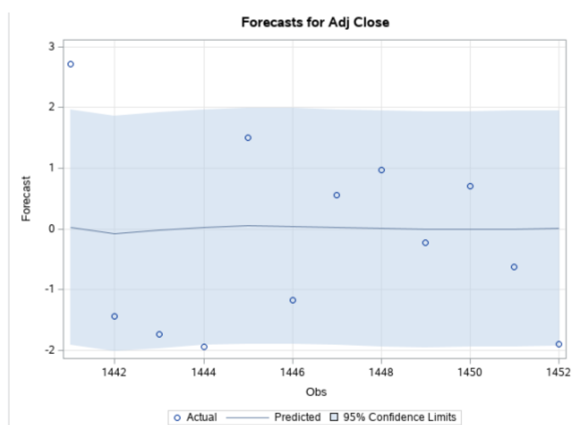
We will still show the performance indicators.

- Final model ARIMA (2, 1, 3)
- AIC: 4092.40

The AIC is the Akaike information criterion. As an estimator of prediction errors, this statistic relatively estimates how much information is lost by using a given model. In other words, it helps with model selection, estimating the quality of a model against others.

SSE: 1414.09

The SSE is the residual sum of squares, also known as the sum of squared estimate of errors. It is significant in statistics because it is a measure of the similarity between data and a model of estimation. The smaller the value, the tighter the fit of the model to the data.



Predictions on a 12 horizon with a hold back period of 12.

Conclusion:

In conclusion, the goal of this analysis was to counter the numerous attempts on Medium of using ARIMA to predict the stock market. We wanted to outline that the nature of this task is difficult and in fact cannot be predicted simply by applying the algorithm on the series. We showed this through the many assumptions and criteria that needs to be filled to be able to have trustworthy prediction. Without those assumptions met, ARIMA model cannot be efficiently use, especially in the context of prediction.

References

Andrews, Bruce H. Dean, Matthew D. Swain, Robert. Cole, Caroline. “Building ARIMA and ARIMAX Models for Predicting Long-Term Disability Benefit Application Rates in the Public/Private Sectors”. University of Southern Maine. August 2013. https://www.soa.org/globalassets/assets/files/research/projects/research-2013-arima-arimax-ben-appl-rates.pdf

Bevans, R. (2021, June 18). An introduction to the Akaike information criterion. Retrieved from https://www.scribbr.com/statistics/akaike-information-criterion/

Elbahloul, Karim. (2019). Stock Market Prediction Using Various Statistical Methods Volume I. 10.13140/RG.2.2.13235.17446.

Gandhi, P. (2020, August 19). 7 Statistical Tests to validate and help to fit the ARIMA model. Retrieved from https://towardsdatascience.com/7-statistical-tests-to-validate-and-help-to-fit-arima-model-33c5853e2e93

Malato, G. (2020, September 15). Statistical analysis of a stock price. Retrieved from https://towardsdatascience.com/statistical-analysis-of-a-stock-price-e6d6f84ac2cd

Peters, K. (2021, January 24). Time series analysis for predictive maintenance of turbofan engines. Retrieved from https://towardsdatascience.com/time-series-analysis-for-predictive-maintenance-of-turbofan-engines-1b3864991da4

Peters, K. (2020, October 04). The importance of problem framing for supervised predictive maintenance solutions. Retrieved from https://towardsdatascience.com/the-importance-of-problem-framing-for-supervised-predictive-maintenance-solutions-cc8646826093

Prabhakaran, S. (2021, November 22). Augmented Dickey-Fuller (ADF) Test - Must Read Guide - ML. Retrieved from https://www.machinelearningplus.com/time-series/augmented-dickey-fuller-test/

Stephanie. (2021, November 18). White Test: Definition, Examples. Retrieved from https://www.statisticshowto.com/white-test/

Ullah, M. I. (2021, January 20). White test for Heteroskedasticity Detection. Retrieved from https://itfeature.com/heteroscedasticity/white-test-for-heteroskedasticity