Cryptocurrency Price Prediction

Using Time Series Forecasting

CARDANO

Group 11
Rifhad Pirani
Lilian Signey-Hall

DSCI 5340
Predictive Analytics
Dr. Mahdi Fathi

Introduction

Cardano also known as ADA (named after Ada Lovelace – 1st computer programmer & 19th-century mathematician) is a rising cryptocurrency founded by Charles Hoskinson, who was also one of the co-founders of Ethereum the second most popular cryptocurrency behind Bitcoin. It's a digital currency that can be used by anyone anywhere in the world without the use of a third-party mediator. Every transaction is tracked, recorded, and verified on the Cardano blockchain. Cardano is coined as "Built for the Future" by Investor Place and is unique in that it is "one of the biggest blockchains to successfully use a proof-of-stake consensus mechanism, which is less energy-intensive than the proof-of-work algorithm relied upon by Bitcoin" (Cardano, 2021). This proof-of-stake blockchain platform's main goal is to allow "changemakers, innovators, and visionaries" to bring about positive global change. It also aims to "redistribute power from unaccountable structures to the margins to individuals — helping to create a society that is more secure, transparent, and fair" (Hake, 2021).

Cardano is used in agriculture as a way to track fresh produce from field to fork, used in academics to store and verify credentials, and used in retail to overcome counterfeit products (Sirois, 2021).

Literature Review

I did not find journals written specifically on Cardano but there were many written when it comes to Bitcoin and cryptocurrency in general. Although Cardano has been around since 2017, it is not as popular as Bitcoin or other well-known digital currency.

Hypothesis Statement

Null hypothesis: To predict the future Low Price of Cardano using an ARMA model.

Alternative hypothesis: The Low Price of Cardano cannot be predicted using the ARMA model.

Motivation and Objectives

Cardano uses an ADA token that is configured that certifies its owners to be involved in the operation of the network. As a result of this, Cardano holders "have the right to vote on any proposed changes to the software" (3).

Analysis of this cryptocurrency is beneficial for those interested in this digital asset although this should not be viewed as investment advice.

We aim to achieve the following objectives in our project: Analyze the times series and use the different models discussed in chapters 6,7,8,9,10 and 11 in our textbook, to present the best model to forecast the future low price of Cardano.

Dataset

Fig. 1 represents the low price of Cardano from January 2021 to June 24, 2021. Our dataset was downloaded from coinmarketcap.com as daily and monthly CSV files.

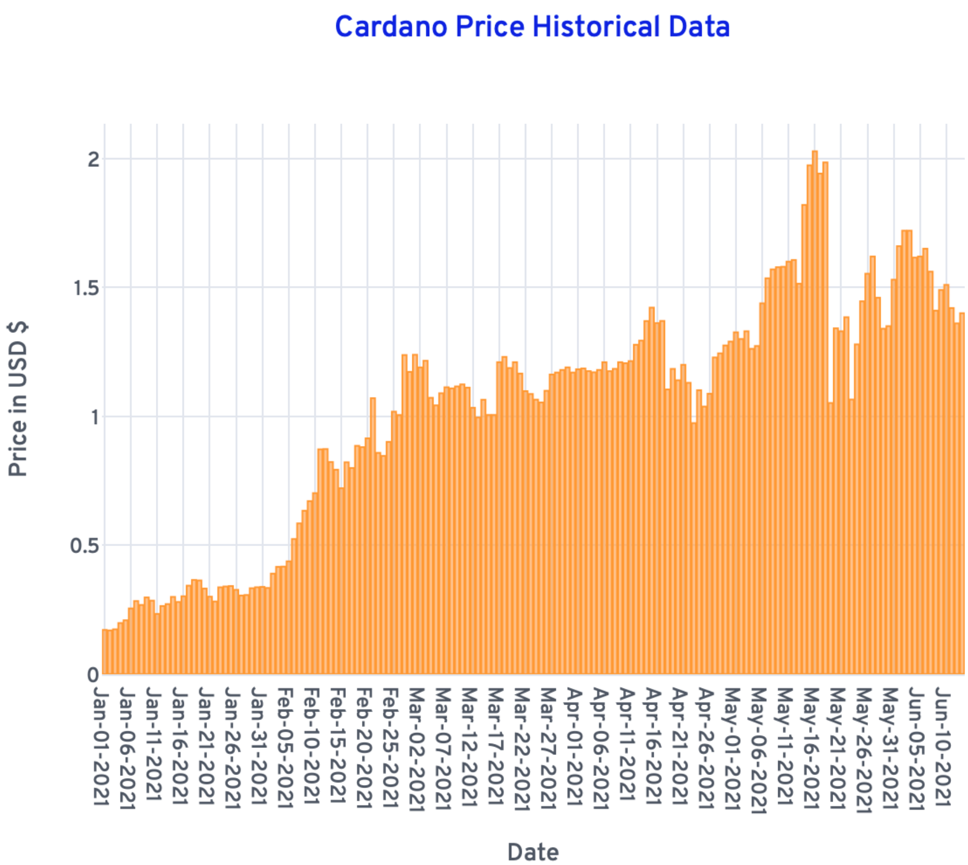


Figure 1 – Original Daily Low Price Time Series of Cardano

Methodology

Chapter 6 - Time Series Regression

In plotting our data to time series regression as seen in fig.1, it shows no cycle nor seasonality but shows an increasingly positive trend. The SAS output in fig.3 shows the time series model is significant because the P-value is less than alpha, Beta zero, and Beta one's P-value is less than alpha, so they are significant for the model. In fig.3 (b), we use the Durbin Watson Statistic to detect the first-order autocorrelation of the model. It shows d is less than DL, alpha, so we reject the null hypothesis and so the errors are positively autocorrelated. This model is inadequate.

The REG Procedure
Model: MODEL1
Dependent Variable: Low

Number of Observations Read	164
Number of Observations Used	164

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	28.50856	28.50856	771.40	<.0001
Error	162	5.98703	0.03696		
Corrected Total	163	34.49559			

Root MSE	0.19224	R-Square	0.8264
Dependent Mean	1.02488	Adj R-Sq	0.8254
Coeff Var	18.75744		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-195.91828	7.09092	-27.63	<.0001
Date	1	0.00881	0.00031709	27.77	<.0001

Figure 3(a) – SAS Output – P-value

The REG Procedure
Model: MODEL1
Dependent Variable: Low

Durbin-Watson D	0.334
Number of Observations	164
1st Order Autocorrelation	0.822

Figure 3(b) – SAS Output Durbin Watson Test

After performing a logarithmic transformation at order 1, the model $\mathbf{z_t = \delta + a_t}$ is an adequate model because the values of Q are greater than alpha.

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	7.44	6	0.2822	-0.123	0.027	-0.112	0.063	-0.106	0.025
12	14.98	12	0.2424	-0.011	-0.052	0.070	0.144	0.021	-0.117
18	17.78	18	0.4701	-0.096	-0.036	-0.016	0.036	0.033	0.049
24	26.08	24	0.3492	0.094	0.040	-0.038	-0.144	-0.024	0.102

Figure 3(c) – SAS Output BL values

In point estimation prediction, a model must be transformed to have linear parameters and the growth curve model, see fig. 4 (a), does this by taking the logarithms on both sides. The Durbin-Watson statistic for this model is .094 and shows evidence that the error terms for this model display first-order autocorrelation. In fig.4 (b) shows the p-values are less than alpha and are therefore significant.

Figure 4 (a) – SAS Output Growth Curve

The REG Procedure
Model: MODEL1
Dependent Variable: Lny

Durbin-Watson D	0.094
Number of Observations	164
1st Order Autocorrelation	0.930

Growth Curve

The REG Procedure
Model: MODEL1
Dependent Variable: Lny

Number of Observations Read	164
Number of Observations Used	164

Analysis of Variance

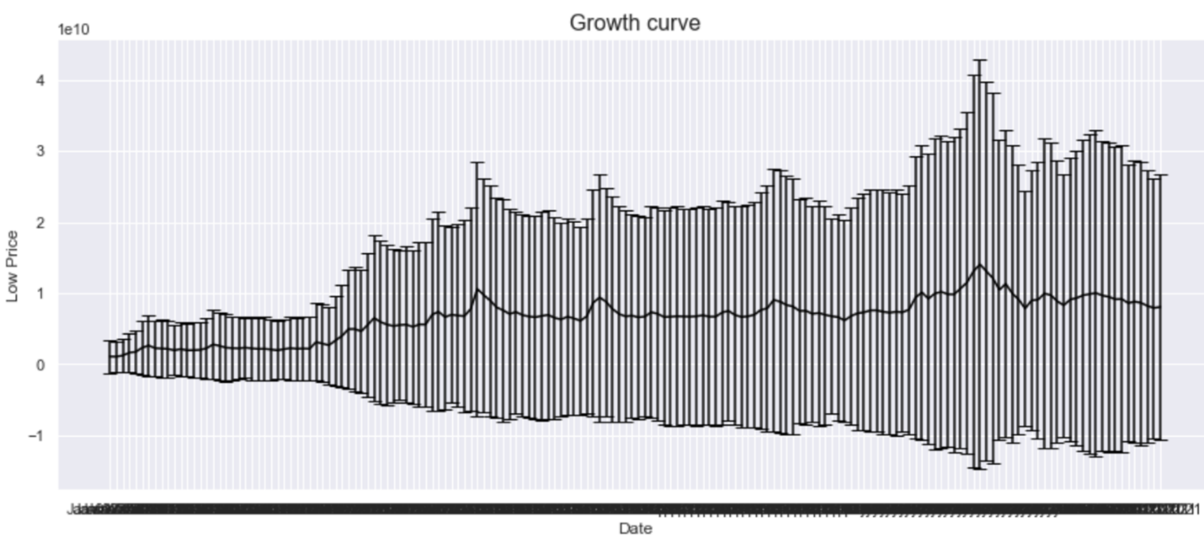
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	48.52638	48.52638	500.96	<.0001
Error	162	15.69236	0.09687		
Corrected Total	163	64.21874			

Root MSE	0.31123	**R-Square**	0.7556
Dependent Mean	-0.12918	**Adj R-Sq**	0.7541
Coeff Var	-240.92616		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-257.07539	11.47997	-22.39	<.0001
Date	1	0.01149	0.00051336	22.38	<.0001

Figure 4 (b) – Growth Curve SAS Output



Chapter 7 - Decomposition

To forecast time series that exhibit trend and seasonal effects, decomposition models such as multiplicative and additive decomposition models can be modeled adequately using moving averages and centered moving averages to eliminate such seasonal variations and irregular fluctuations. Multiplicative is useful when seasonal variation increases over time and additive are useful when the seasonal variation is relatively constant over time. Before decomposing the time series, we used the monthly dataset to apply it to the multiplicative model. The SAS output can be seen in fig. 5. and with a p-value, a little over alpha, and a very low R^2 signifies that this model is not adequate.

Low Price DeseasonalizedLow regressed on Time

The REG Procedure
Model: MODEL1
Dependent Variable: DeseasonalizedLow

Number of Observations Read	45
Number of Observations Used	45

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	7.57835	7.57835	8.62	0.0053
Error	43	37.80244	0.87913		
Corrected Total	44	45.38079			

Root MSE	0.93762	R-Square	0.1670
Dependent Mean	0.63039	Adj R-Sq	0.1476
Coeff Var	148.73547		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.09637	0.28427	-0.34	0.7362
Time	1	0.03160	0.01076	2.94	0.0053

Figure 5 – Decomposition Output in SAS

Chapter 8 – Exponential Smoothing

Exponential Smoothing is useful when a time series changes over time, has trend and seasonal factors because it gives higher weight on the most recent observed values. We trained our data using three models: Simple Linear, Winter's, and Holt Exponential Smoothing. Table 1 shows the comparison of how the three models fit our dataset using R squared, AIC, SBC, and MSE. Linear Holt's Exponential Smoothing model fared better than the other two models because it has the highest R^2 and lowest values for errors. This model is useful for a series with trend but no seasonal variation. The SAS output shown in fig. 6 shows the p-values of the level weight for all three models are significant. After training the dataset using these three models seen in table 2, the R^2 values were significantly low in value but the model with the least error is simple exponential smoothing.

	R^2	AIC	SBC	MSE
Simple	94.06%	-675.4494	-672.4125	.0123
Linear Holt's	94.15%	-675.1774	-669.1041	.0122
Winter's	93.71%	-662.15658	-653.04572	.0131

Table 1 - Fit after Exponential Smoothing

	R^2	AIC	SBC	MSE
Simple	-466.23%	-28.9151	-28.9151	.0555
Linear Holt's	-788.43%	-24.4105	-24.4105	.0871
Winter's	-722.08%	-25.1867	-25.1867	.0806

Table 2 - Performance after Exponential

Simple Exponential Smoothing Parameter Estimates

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|
Level Weight	0.83105	0.05636	14.74	<.0001

Linear Exponential Smoothing Parameter Estimates

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|
Level Weight	0.81612	0.05697	14.32	<.0001
Trend Weight	0.0010000	0.02033	0.05	0.9608

Winters Method (Multiplicative) Parameter Estimates

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|
Level Weight	0.78363	0.05337	14.68	<.0001
Trend Weight	0.0010000	0.02041	0.05	0.9610
Seasonal Weight	0.33964	0.12230	2.78	0.0062

Figure 6 – SAS Output with P-values

Chapter 9 – Non-seasonal Box-Jenkins Model - ARIMA

In time series analysis, an autoregressive integrated moving average or ARIMA model requires the dataset to be stationary. In pre-processing the dataset, data exploration through visualization using Python is a good way to see how the dataset looks like (see fig.1). The two methods we used to test stationarity are the p-values of the Augmented Dickey-Fuller Test and SAC. Fig. 7(a) shows the ADF test with p-values greater than alpha and Fig. 7 (b) shows SAC slowly dying down. As a result, data is non-stationary.

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	1	0.1899	0.7266	0.1762	0.7365		
Single Mean	1	-4.9650	0.4338	-1.8765	0.3425	2.2886	0.4876
Trend	1	-21.4118	0.0450	-3.0985	0.1102	5.0720	0.1636

Figure 7(a) – ADF Test

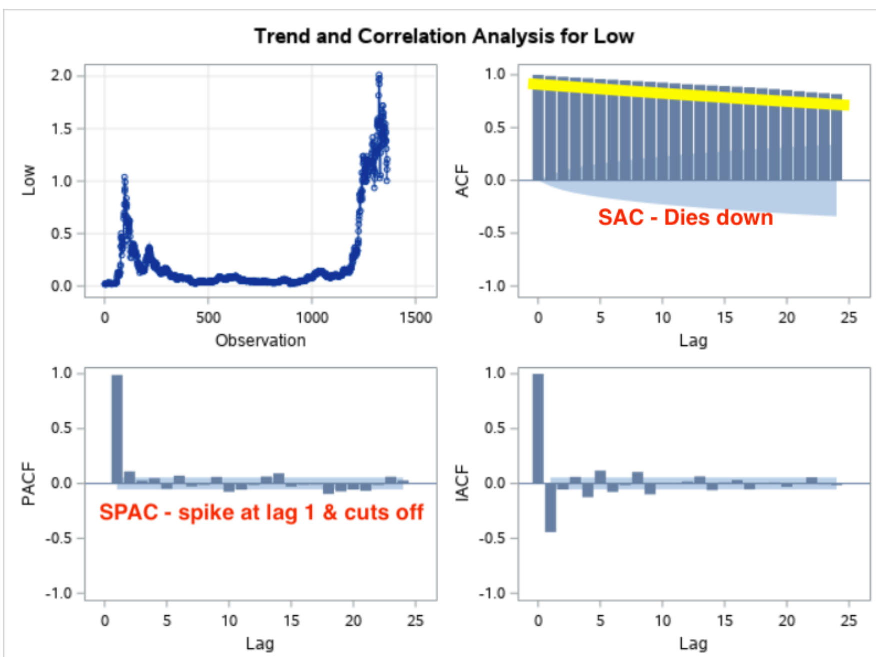


Figure 7(b) – SAC slowly dying down

To attain stationary, in the time series, first-order differencing at lag one was applied.

```
ADF Stastistic: -2.834158
p-value: 0.053558
The graph is non stationary
```

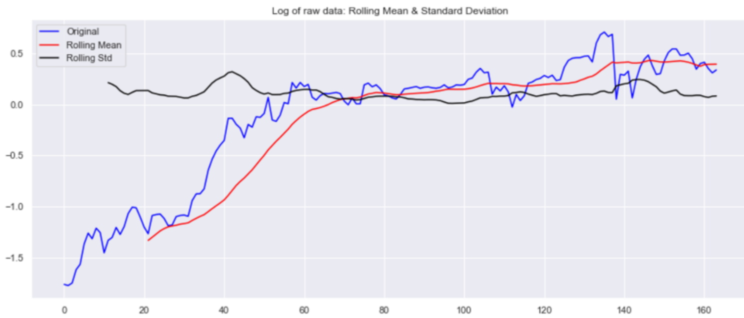


Figure 8 (a) Before Differencing

```
ADF Stastistic: -14.312968
p-value: 0.000000
The graph is stationary
```

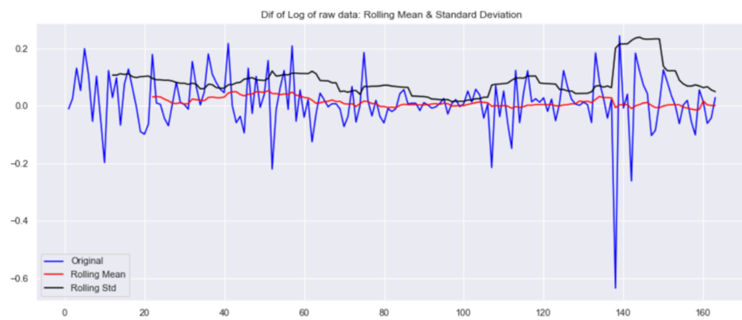


Figure 8 (b) After Differencing, data is stationary

Chapter 10 -Estimation, Diagnostic Checking & Forecasting

Box-Jenkins models require forecasting the time series to be stationary and invertible. The moving average model point estimate results are shown in fig. 9 (a) with $\hat{\theta}_1$ = -.95828 after 4 iterations, and since $|\theta_1| < 1$ this model satisfies the invertibility condition.

Conditional Least Squares Estimation				
Iteration	SSE	MA1,1	Lambda	R Crit
0	65.9111	-0.95828	0.00001	1
1	62.8229	-0.90873	1E-6	0.242307
2	62.8150	-0.89940	1E-7	0.02182
3	62.8056	-0.90387	0.1	0.018692
4	62.8055	-0.90354	0.01	0.000799

Figure 9 (a) - Invertibility

Since the highest absolute estimated correlation between the point estimates found in the correlation matrix in fig. 9 (b) is > .9, we conclude that the point estimates in this model are highly correlated.

The absolute t-value = 25.41 (fig.9 c) which is greater than 2 and p-value is less than .05. Thus, it is reasonable to retain this parameter in the model.

The Ljung-Box statistic tests the adequacy of the overall model as seen in fig. 9 (b) which shows the values as less than .05. Thus, this model is inadequate.

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	800.47	5	<.0001	0.822	0.975	0.816	0.952	0.812	0.927
12	1554.71	11	<.0001	0.813	0.903	0.809	0.881	0.802	0.855
18	2266.15	17	<.0001	0.791	0.836	0.786	0.823	0.782	0.807
24	2942.56	23	<.0001	0.771	0.789	0.761	0.772	0.757	0.759
30	3582.73	29	<.0001	0.749	0.745	0.745	0.721	0.729	0.698

Figure 9(b) – Correlation Matrix

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	-0.90354	0.03556	-25.41	<.0001	1

Variance Estimate	0.38531
Std Error Estimate	0.620733
AIC	310.0007
SBC	313.1006
Number of Residuals	164

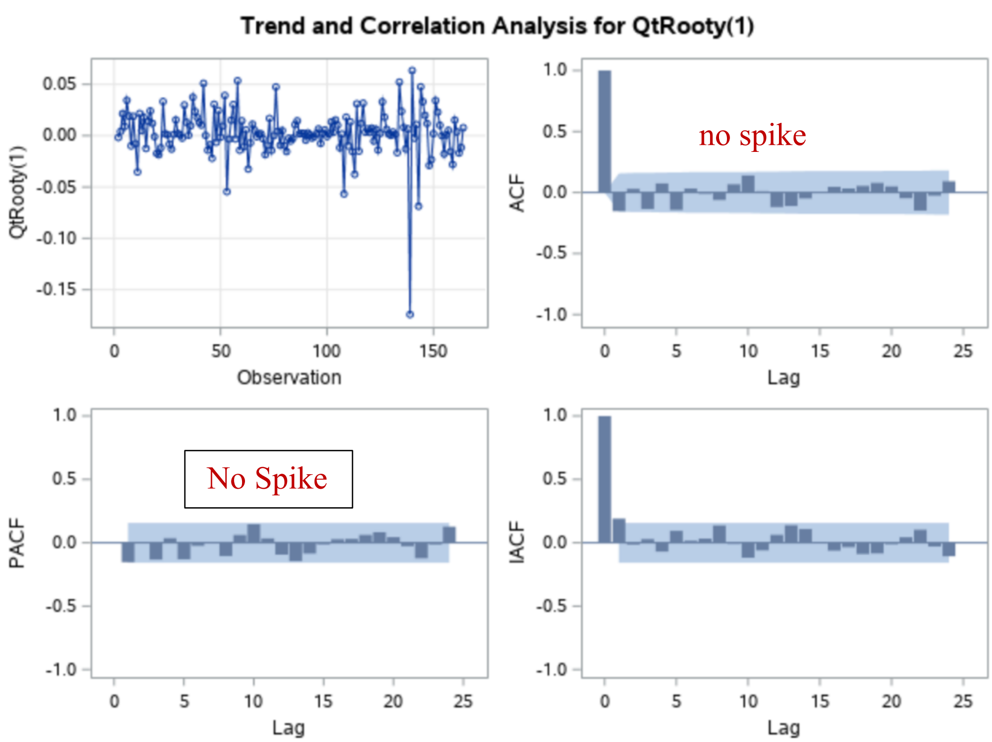
Figure 9(c) – T-values and P-values

For the Moving Average model, the resulting p-value is less than alpha and so we reject the null hypothesis, and as a result, the value is significant, but the Ljung-Box statistic is all less than alpha, showing significant autocorrelation, therefore, making the model inadequate.

Chapter 11 – Seasonal Box-Jenkins Model - SARIMA

The Seasonal Box-Jenkins model is used to improve the error structure and transform the time series regression into a stationary time series by using pre-differencing transformation. After performing the seasonal quartic root transformations to our dataset, the Q values in the Ljung-Box statistic show a greater than alpha and shows there no spike in the SAC and SPAC, we can conclude that this model is adequate.

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	11.72	6	0.0685	-0.153	0.030	-0.135	0.075	-0.143	0.033
12	19.15	12	0.0849	-0.010	-0.061	0.067	0.140	0.010	-0.119
18	22.92	18	0.1936	-0.110	-0.049	0.001	0.047	0.033	0.054
24	30.91	24	0.1564	0.079	0.050	-0.047	-0.148	-0.025	0.092



Conclusion

After building different models to train our dataset using the different models found in chapters 6-11, we found the following models to be adequate and they are represented in the table below. Using logarithmic transformation in Model 1 and quartic root transformations Model 2, this showed Q values that are greater than alpha making the models adequate. On the other hand, comparing Log, Linear, and Quadratic in Models 3-5, showed the following results. In comparing all 5 Models based on standard error, the model with the least error is Model 2 ARMA(1) on the first order.

MODEL	AIC	SBC	STD ERROR	R^2
Model 1 $z_t = \delta + a_t$	**-301.629**	**-298.536**	**.095633**	
Model 2 ARMA(1)	**-753.308**	**-750.214**	**.023927**	
Model 3 Lny = log(y)			**.31123**	**75.56%**
Model 4 $Y_t = \beta_0 + \beta_1 t + \varepsilon_t$			**.19224**	**82.64%**
Model 5 $Y_t = \beta_0 + \beta_1 t + \beta_1 t^2 \varepsilon_t$			**.19224**	**82.64%**

Here are the forecasted values with 95% confidence limits and standard errors. .

	Forecasts for variable QtRooty				
	Obs	Forecast	Std Error	95% Confidence Limits	
July 25, 2021	165	1.0905	0.0239	1.0436	1.1374
July 26, 2021	166	1.0932	0.0338	1.0269	1.1595
July 27, 2021	167	1.0959	0.0414	1.0147	1.1772
July 28, 2021	168	1.0986	0.0479	1.0049	1.1924
July 29, 2021	169	1.1014	0.0535	0.9965	1.2062
July 30, 2021	170	1.1041	0.0586	0.9892	1.2190
July 1, 2021	171	1.1068	0.0633	0.9827	1.2309
July 2, 2021	172	1.1095	0.0677	0.9769	1.2422
July 3, 2021	173	1.1123	0.0718	0.9716	1.2530
July 4, 2021	174	1.1150	0.0757	0.9667	1.2633

References:

Cardano. (2021). Cardano. https://www.cardano.org

Hake, M. (2021, April 19). *Why Cardano's Proof-of-Stake System Is Unique for a Cryptocurrency*. Investor Place. https://investorplace.com/2021/04/why-cardanos-proof-of-stake-system-is-unique-for-a-cryptocurrency/

Sirois, A. (2021, May 25). *Cardano Has a Different Use as Well as a Different Future From Bitcoin*. Investor Place. https://investorplace.com/2021/05/cardano-has-a-different-use-as-well-as-a-different-future-from-bitcoin/