# PREDICTION OF CRYPTOCURRENCY PRICES USING SEASONAL AND AUTOREGRESSIVE INTEGRATED MOVING AVERAGES

**Mrs.P.Naga Durga,** Assistant Professor, Department of Statistics, Loyola Academy, Secunderabad, Telangana
**Dr.K.Vijayalakshmi,** Assistant Professor, Department of Mathematics, Loyola Academy, Secunderabad, Telangana

## ABSTRACT
In recent times we have observed a new kind of "asset", as people are treating it, to be legitimate, a kind of an intangible asset, called crypto currency. A crypto, as it is called in the market, is a decentralized virtual currency, supervised by no central authority, it is basically used for business and day-to-day transactions. It is highly encrypted, by cryptography and is secured. This crypto paved the way for a secured transaction of large amounts. This feature of huge amounts of money transfer is the reason for its high use especially for business related transactions. Now, due to its huge popularity and wavering price of a crypto coin sometimes spiking up of its prices due time and being able to trade or cash in the coin for more than it was bought, this advantage made people to consider it more of an investment option than as a currency. So, we want to predict the future prices of the stocks and crypto currency prices and compare them based on the risk factor in future. This paper mainly focuses on the various features of the crypto Currency using various statistical models like ARIMA(Autoregressive Integrated Moving Average) and SARIMA Models using R-Language.

**Key words:** Cryptography, Virtual Currency, Crypto Currency, ARIMA, SARIMA.

**Introduction:** A common person looks to progress and prosper in life, so he will be careful in investing his time and efforts, especially money. In recent, times we have seen a new kind of asset or to be more precise a popular investing place area called crypto currency. Now, a day's people are more intrigued towards crypto currency because of its popularity and high returns. This high returns also comes hand in hand with a high risk of loss and profits. Now, due to its huge popularity and wavering price of a crypto coin sometimes spiking up of its prices due time and being able to trade or cash in the coin for more than it was bought, this advantage made people to consider it more of an investment option than as a currency. As defined above it is considered as an intangible asset, this asset is treated more as a stock, where in we buy it at a particular price and sell it at a particular price, which is pretty similar to crypto as well. The person who wants to invest his/her money has a new option in the form of crypto currency, in contrast to old stock.

## Objective of the study
The objective of the study:

(i) To forecast the range of prices of crypto-currency taken into consideration at random for the years 2022-24, those are ADA, BNB, BTC and USDT
(ii) To apply a chi-square test to validate the forecasted high value with stock or crypto high price and do the same for forecasted low value.
(iii) To determine where to invest based on the variance.

## Research Hypothesis:
1. H0C: There will be a great difference of standard deviation between lower and upper predicted values of stocks.
2. H1C: There will be a less difference of standard deviation between lower and upper predicted values of stocks.

**Significance of the study**

Things change over time and people adapt to them. This change may be permanent or temporary, so is the rise of crypto in the market, which can be considered as a new trend, a change that happened in the market with a thirty rise of usage of crypto-currency and its popularity in the recent past and its increasing prices people are considering it as a good investment option. Moreover, the comparison between stocks and crypto is something people have been skeptical about and this study tries to resolve that. This can help people have an idea on how different or similar stocks and crypto can be in the near future.

**Scope of the study**:

The study is purely based on statistical analysis and the inference that can be drawn from them. There are many stocks listed in the world market, that too in a specific economy. We are considering the stocks listed in the Nation Stock Exchange (NSE) market. We are not considering the stock not listed in NSE, on contrary to crypto, whose scope of its ascribable factor is not limited to India. We have taken five stocks and crypto currency price data from yahoo finance website, the subjects under study were considered absolutely at random, their choice is pretty irrational. We are predicting the price based on the past, recent past and present data for both stock and crypto, but not on the factor that might influence them.

**Limitation of study**

The core motive behind this project is to predict the future prices of random stocks and crypto-currency. The prediction may or may not lie in the range of value of future stocks and crypto-currencies. Moreover, these may be induced by market forces as well. This study is limited to statistical prediction based on ARIMA and SARIMA modeling.

**METHODOLOGIES**

ARIMA & SARIMA:

The word ARIMA stands for "Autoregressive Integrated Moving Average", in statistics and econometric and in particular in time series analysis, an autoregressive integrated moving average model is a generalization of an autoregressive moving average model. Both of these models fitted to time series data either to better understand the data or predict future points in the series. In SARIMA the 'S', stands for season, if seasonality exists in the data.

ARIMA models are generally denoted ARIMA ($p$, $d$ , $q$) where parameters $p$, $d$, and $q$ are non-negative integers, $p$ is the order (number of time lags) of the autoregressive model, $d$ is the degree of differencing (the number of times the data have had past values subtracted), and $q$ is the order of the moving-average model. Seasonal ARIMA models are usually denoted ARIMA ($p$, $d$ ,$q$)($P$,$D$,$Q$), where $m$ refers to the number of periods in each season, and the uppercase $P$,$D$,$Q$ refer to the autoregressive differencing, and moving average terms for the seasonal part of the ARIMA model.

A stationary time series properties do not depend on the time at which the series is observed. Specifically, for a wide-sense stationary time series, the mean and the variance/auto-covariance keep constant over time so, basically the time series will not have a predictable pattern in the long term. To know whether a time stationary or not we use adf(), whose acronym is Augmented Dickey Fuller test , where in p- value should be less the 0.05, if not the time series is non- stationary

We can also visually depict the trends and seasonality using acf() acf acronym to autocorrelation function, in this function if we observe a trend in the lines and seasonality in it we have to make it stationary. So, to make a time series stationary we need to make a difference. The pacf() is used to depict how well a particular lag is partially correlated with the other lag.

Ats function :

The ts() is used to convert a data frame into a time series data in R. ts(data, frequency, start,end)

where in the parameter definition

data - (in our case the column that should be converted into time series the column from the excel sheet can be accessed by using '$');
frequency (No. of prices in a time cycle (year) in our case)=52; start (starting date);
end (ending date) .

Differencing

Differencing in statistics is a transformation applied to a non-stationary time-series in order to make it stationary ie., to remove trend and seasonality in a time series, but having nothing to do with the non-stationary of the variance. Likewise, the seasonal differencing is applied to a seasonal time-series to remove the seasonal component.Since we are using R programming language the amount of seasonal and non-seasonal differencing can be found out by using the ndiff() and nsdiff().

ACF() and PACF()

After differing accordingly and knowing d and D value using the above mentioned function we need to find p, q values, since we are trying the model using R language, we can find those values using acf() and pacf(). The p and q value is chosen when the spike in acf and pacf graphs exceeds more than the significant bond which appears in the form of a linear line across the X-axis, after seasonal and non-seasonal differing.Now try a different combination of (p,d,q),(P,D,Q) values that we have got from acf(), pacf(), ndiffs(), nsdiffs(). The model which gets comparatively lower AICs value is considered the best fit model for the given time series data.

AICs

The Akaike information criterion (AIC) is an estimator of out-of-sample prediction error and thereby relative quality of statistical models for a given set of data. Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models. Thus, AIC provides a means for model selection.In plain words, AIC is a single number score that can be used to determine which of multiple models is most likely to be the best model for a given data set. It estimates models *relatively*, meaning that AIC scores are only useful in comparison with other AIC scores for the same data set. A lower AIC score is better.

forecast()

The forecast() is used to forecast future values for a specific time period for a given series of data

checkresidual()

This function is used to verify the residuals of a time series, residuals means, the difference between the actual value and fitted value. This function gives a histogram, ACF and graph showing residuals.

ADF test:

This is used to test the stationary of a time series data. It's hypothesis are : H0: The time series is non-stationary ie., p-value > 0.05
H1: The time series is stationary ie., p-value < 0.01

**Data Analysis:**

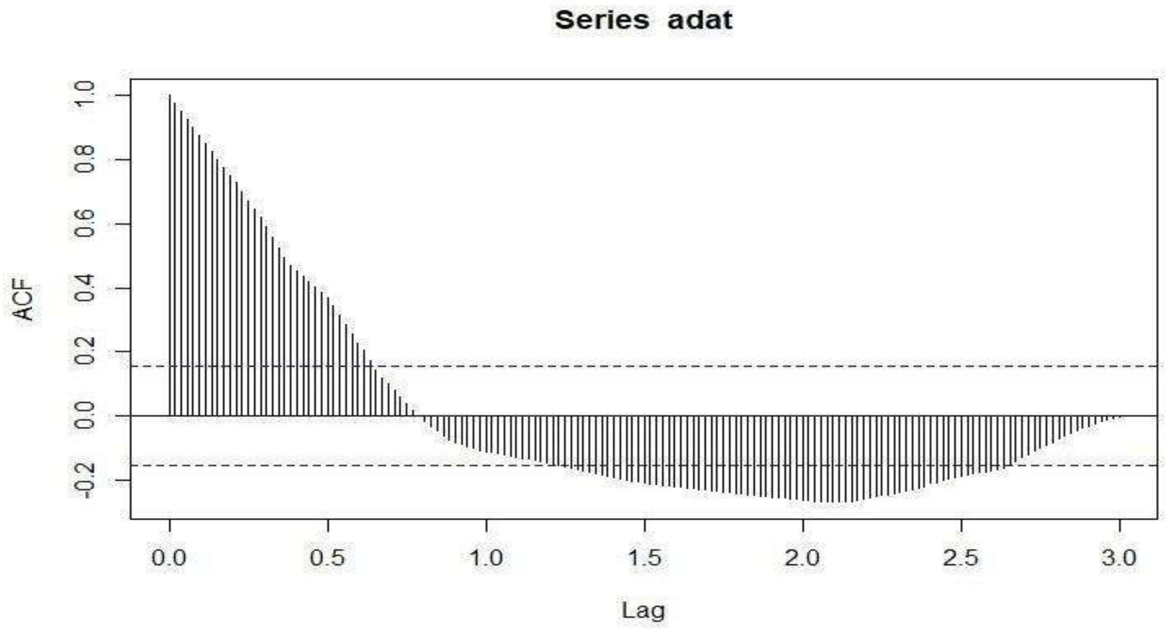
We have collected weekly prices of five random stocks and crypto-currency each from Yahoo Finance, from 1th January 2019 to 31th December 2021. The analysis has been done using the R programming language in R studio.

**CRYPTOCURRENCY:**

First we import the data and convert it into a time series data, as the data imported is in data frame format, to convert it into time series we use ts function. Now, check for stationarity of the time series by using ADF function

```
        Augmented Dickey-Fuller Test

data:  adat
Dickey-Fuller = -2.0121, Lag order = 5, p-value = 0.5713
alternative hypothesis: stationary
```




We can see that the p-value is greater than 0.05. So, we reject the null hypothesis that it is stationary and conclude that the given series is non-stationary. We can also check for it or depict it visually by acf() plot.

We can observe a pattern in the correlogram which depicts very high autocorrelation with the $n^{th}$ lag term. So, the given series cannot be stationary.

We now use the auto arima function to find an arima model to fit the data.We are not confining ourselves to the model given by auto arima function, but try out other combinations as well, based on the criterion of least AICs value. So, in order to do so we first find the 'd', parameter, which is the amount of non-seasonal or sometimes seasonal differing required.

```
Series: adat
ARIMA(0,1,0)

sigma^2 = 115.8:  log likelihood = -592.01
AIC=1186.01   AICc=1186.04   BIC=1189.06
> ndiffs(adat)
[1] 1
> nsdiffs(adat)
[1] 0
```

After ascertaining the 'd' and 'D' value and differing it we now test for stationarity using the ADF function. The result we got is that the p-value, which is 0.01 is less than 0.05, So we accept the null hypothesis and therefore the series is now stationary.

We now need to find p,q values to fit an arima model. We find them using ACF and PACF plot.

The 'p', values are0,11,25 and 'q', values are 0,20 Now we try out different p, d, q combinations. The (0,1,0) combination is the best as it has the least value among all the other combinations. The

ISSN: 2278-4632
Vol-12 Issue-12 No.03, December 2022

forecast based on the model is done and it is exported to an excel sheet.The forecasted value along high and low prices is plotted on a graph in the excel sheet.After forecasting checks for the residual and calculate lower and upper values standard deviation.

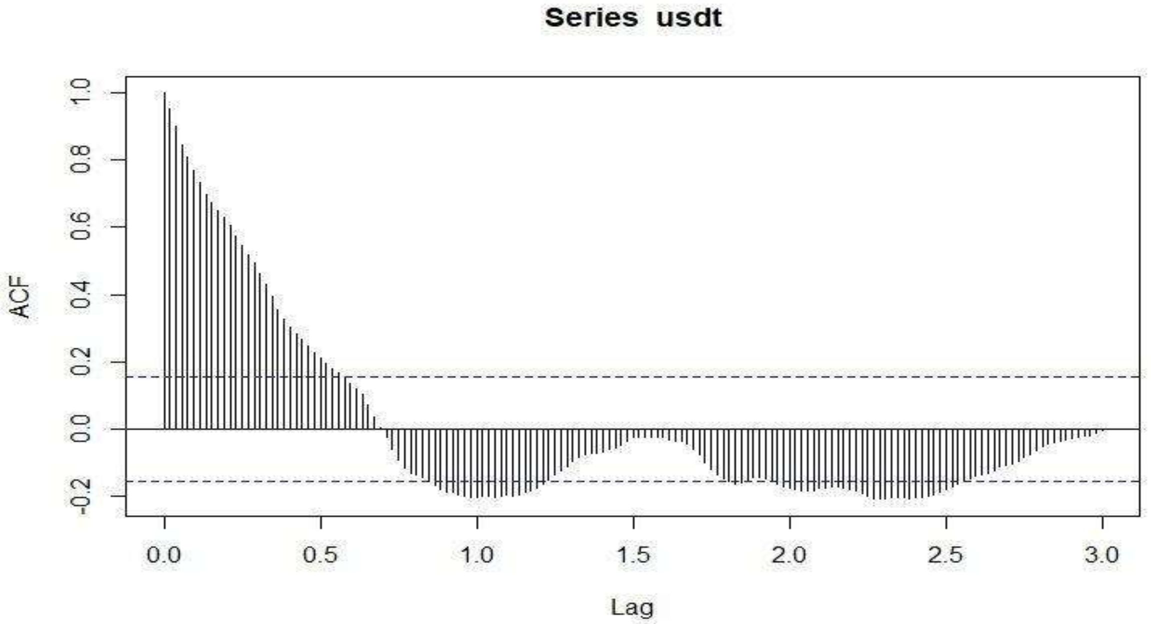
The correlogram indicates that the model has used most of the data we can observe one line is a bit above the significant bound when we come to histogram the curve is skewed to center indicating that the model has used most available data and is equally weighted between past and recent values, but more no past data. We can also observe outliers which is inevitable.

**Conclusion:** So, the best arima model fitted for the given data is (0,1,0). The given output has two prediction one is the lower one and another is upper, these indicate the highest and lowest the price can reach at a specific point of time. The variance is equal for both lower and upper range of values which are almost the same with a slit decimal difference indicating an equal possibility of rate of increase or decrease in the future.

**USDT**

First, we import the data and convert it into a time series data, as the data imported is in data frame format, to convert it into time series we use ts function. Now, check for stationarity of the time series by using ADF function.

We can see that the p-value is greater than 0.05. So, we reject the null hypothesis that it is stationary and conclude that the given series is non-stationary. We can also check for it or depict it visually by acf() plot.




We can observe a pattern in the correlogram which depicts very high autocorrelation with the $n^{th}$ lag term. So, the given series cannot be stationary.We now use the auto arima function to find an arima model to fit the data. We are not confining ourselves to the model given by auto arima function, but try out other combinations as well, based on the criterion of least AICs value. So, in order to do so we first find the ‘d and ‘D’, parameter, which is the amount of non-seasonal or sometimes seasonal differing required.

After ascertaining the ‘d’ value and differing it we now test for stationarity using the ADF function. The result shows that the p-value, which is 0.01 is less than 0.05, So we accept the null hypothesis and therefore the series is now stationary.

We now need to find p, q values to fit an arima model. We find them using ACF plot and PACF plot. The ‘p’, values are4,35 and ‘q’, values are 0,4,33 Now we try out different p, d, q combinations. The

(3,1,33) combination is the best as it has the least value among all the other combinations. The forecast based on the model is done and it is exported to an excel sheet. The forecasted values along with low and high prices are plotted on a graph in the excel sheet.

After forecasting check for the residual and calculate standard deviation of both lower and upper predicted values. The correlogram indicates that the model has no line above the significant bound when we come to histogram the curve is skewed to center indicating that the model has used most available data and is equally weighted between past and recent values, but more no past data. We can also observe outliers which is inevitable.

**Conclusion:** So the best arima model fitted for the given data is (3,1,33). The given output has two prediction one is the lower one and another is upper, these indicate the highest and lowest the price can reach at a specific point of time. The variance of upper value is slightly greater than the lower values and there is a low level of variations in both lower and upper value. So, there can be low level of increase or decrease in the both regions.

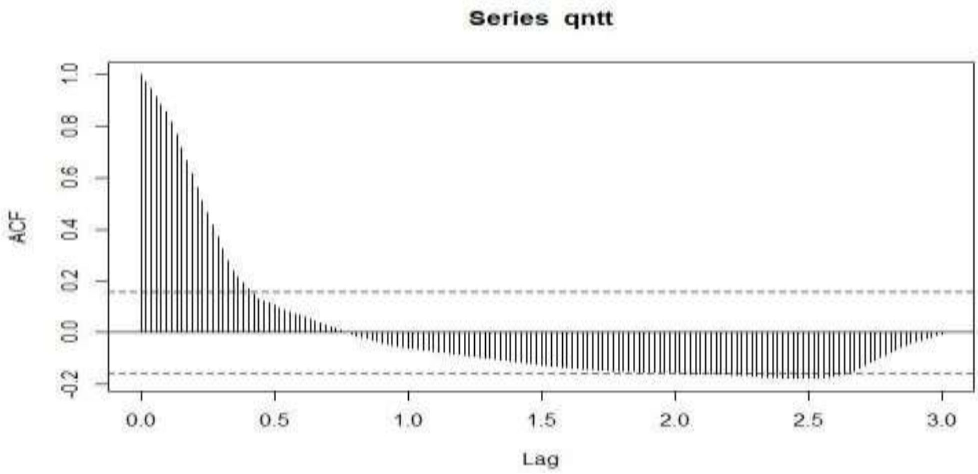
**QNT**

First we import the data and convert it into a time series data, as the data imported is in data frame format, to convert it into time series we use ts function. Now, check for stationarity of the time series by using ADF function

```
        Augmented Dickey-Fuller Test

data:  qntt
Dickey-Fuller = -2.2085, Lag order = 5, p-value = 0.4894
alternative hypothesis: stationary
```

We can see that the p-value is greater than 0.05. So, we reject the null hypothesis that it is stationary and conclude that the given series is non-stationary. We can also check for it or depict it visually by acf() plot.




We can observe a pattern in the correlogram which depicts very high autocorrelation with the n th lag term. So, the given series cannot be stationary.We now use the auto arima function to find an arima model to fit the data.We are not confining ourselves to the model given by auto arima function, but try out other combinations as well, based on the criterion of least AICs value. So, in order to do so we first find the 'd', parameter, which is the amount of non-seasonal or sometimes seasonal differing required.The forecasted values along with lower and upper values are plotted on a graph in the excel sheet.

ISSN: 2278-4632
Vol-12 Issue-12 No.03, December 2022

**COMPARISON:**

CRYPTOCURRENCY:

| | LOWER | UPPER | DIFFERENCE |
|---|---|---|---|
| ADA | 50.31402 | 50.31402 | 0 |
| BNB | 10501.8 | 11590.81 | -1089.01 |
| BTC | 977011.3 | 989285.9 | -12274.6 |
| QNT | 9708.367 | 14862.13 | -5153.763 |
| USDT | 3.012775 | 3.184634 | -0.17186 |

The table depicts the standard deviation of each lower and upper predicted stocks and crypto currencies values .In the table we can see that difference between the standard deviation of lower and upper values for crypto currency are very high, indicating high risk and higher returns, but where as we can observe very sight significant difference in the difference between standard deviation of lower and upper values of stocks, which says that the risk is low, so is the returns.

**CONCLUSION:**

- ✓ For this we can conclude that, there is a very high risk in investing in future and the return might also not be stable, due to unstable variation in standard deviation in each crypto as well.
- ✓ So over all the volatility is high among crypto prices.
- ✓ We conclude that, if a person wants to look for safe and stable returns on a long run, they can opt for stocks
- ✓ If there want high returns, in long, it comes with an inevitable high risk, then they can opt for crypto.

**References:**

**1.**Raji Pillai and Dr Lokanadha Reddy M, the paper titled "Cryptocurrency in India: A Study with reference to the Bitcoin Prices", Paper published in Journal of Advance Management Research, UGC Index No 42888, ISSN: 2393-9664, Feb 2018, Volume 6, Issue 2, pp 203-208.

2.Raji Pillai and Dr Lokanadha Reddy. M, the paper titled, "Cryptocurrency and the Unsteady financial System of developing countries" International Journal of Business and Administration Research (Impact Factor 4.729) Review ISSN No 2348-0653, E-ISSN No 2347-856X, Volume 3, Issue 21, Jan-Mar 2018. Peer Reviewed Quarterly Journal. Pp 96-101.

3.Raji Pillai and Dr Lokanadha Reddy. M, the paper titled, "Investor perception of Investments and fund management through Cryptocurrencies", International Journal of Management, IT & Engineering. ISSN:2249-0558, Impact Factor :7.119, Special Issue April 2019, Pp 223 -238.

4.Raji Pillai and Dr Lokanadha Reddy. M, the paper titled, "Digitization of Financial Markets: A Literature Review on White-collar crimes", Springer Nature Switzerland AG 2021 P. K. Singh et al. (eds.), Innovations in Information and Communication Technologies (IICT-2020), Advances in Science, Technology & Innovation. Ch-6, Pg 1/6.