SARIMA in Cryptocurrency Price Prediction

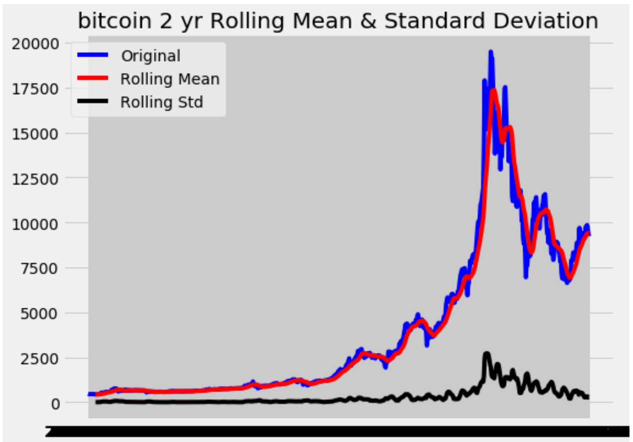
Chunhui Zhu
CUNY SPS Data 602 Spring 2018

Abstract:
Bitcoin is also the first cryptocurrency produced by Satoshi Nakamoto in 2008. It is currently the mainstream cryptocurrency nowadays. It is regulated on electric paper which encryption blockchain techniques are used to regulate the generation of units of currency and verify the transfer of funds, operating independently of a central bank. Virtual currency is circulated through the worldwide network through the use of regional blockchain technology methods. Due to cyber security and technical challenges, and most people who own cryptocurrency currencies are speculators, cryptocurrencies are not accepted and widely used by all countries. At present, cryptocurrency currencies can be speculative like stocks in US, which is extremely risky, and have high volatility. The results in this paper uses the traditional financial statistic models combine with machine learning, scarp live data and run in python program. The finding in this study is using SARMA model can give good prediction in short period cryptocurrency price. This research is focus on bitcoin and other 3 mainstream cryptocurrencies price prediction. The data in this article was collected from a real-time virtual currency trading website, bittrex.com, coinmarketcap.com.

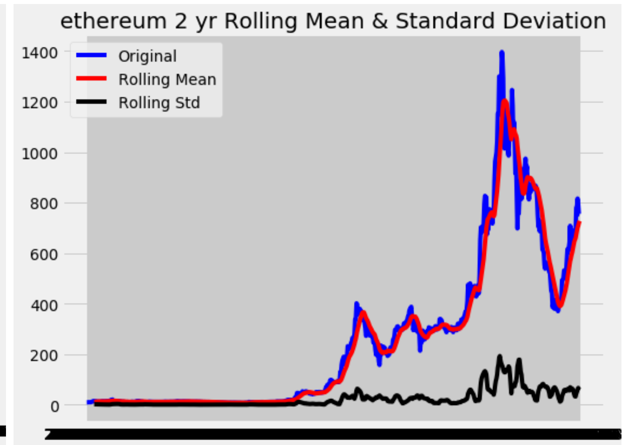
The following history data is from coinmarketcap.com API, using last 2-year close price for bitcoin, ehtereum, ripple, and zcash. From the plot of 2-year close price for four main-stream cryptocurrency, the changes in the prices over the past two years show that the apparent increase in the cryptocurrency market started in 2017. Zcash only has 550s day on the market. In just one year, the price of Bitcoin is 10 times higher. Its price from 2,000 USD over 20,000 USD last year. The price of the Ethereum was less than 100 USD, yet over 1,400 US dollars in last year, an increase of more than 14 times. The price of the ripple was 20 cents USD and reached 3.60 USD. The price of the zcash born at 6 cents and became 44 cents in the past. At the end of 2017, all prices were retracement. At present, the market price of bitcoin, ethereum, ripple and zcash have returned 9,300 USD, 780 USD, 0.88 USD and 277 USD respectively.

Time series model is to find a sequence from discrete-time data to forecasting the trend movement. The model assumes that the data that changes during a certain period of time is uncertain, but the sequence of the entire change is a certain rule, so that the mathematical model can be used to study the similarity of numerical changes. Time series data is basically divided into two categories, stationary and nonstationary. Usually in the financial stock price market, most of the data are nonstationary. There is a python statistical package program to help observe price changes in past 2 years and test the stationary of time series data. The following is to use Dickey-Fuller Test on bitcoin, ethereum , ripple and zcash price data. The hypothetical data for the Dukey-Fuller test are unit root and nonstationary which is assuming random walk in the data. Unit root is present in an autoregressive model (AR). A random walk is failed stationary because of the infinity variance while time increase. Time series can not able to predict the trend if data is random walk. If data fails to reject H0, all t-test results are not being trusted.

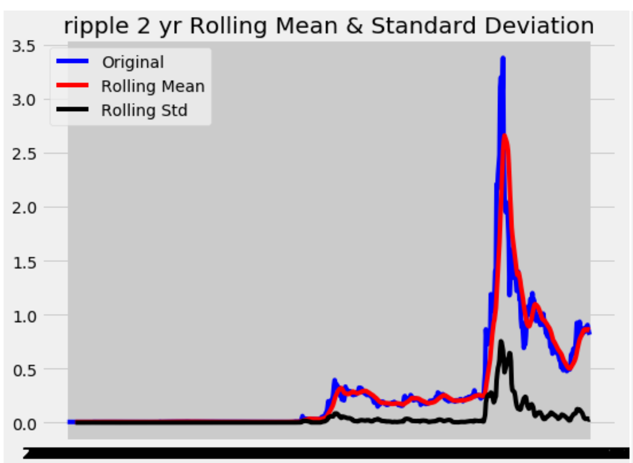
The following will be the assumption for Dukey-Fuller test:
The simple AR(1)model is $y_t = \beta y_{t-1} + u_t$.
Where assuming the $y_t = y_{t-1}$ when it is stationary, so that the expectation of error terms u_t is also 0.



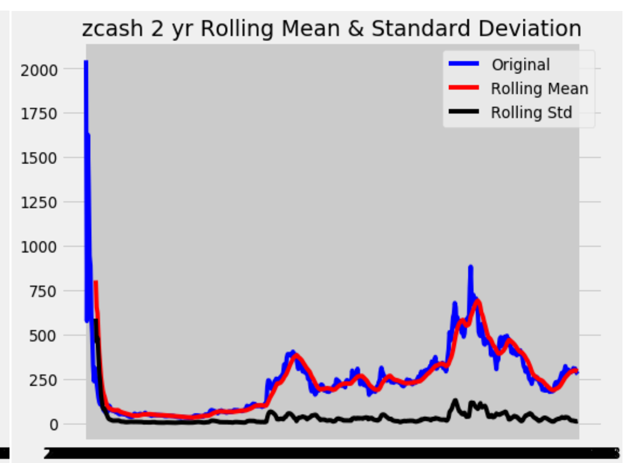

```
Results of Dickey-Fuller Test:
Test Statistic                  -1.319949
p-value                          0.619997
#Lags Used                      20.000000
Number of Observations Used    709.000000
Critical Value (1%)             -3.439607
Critical Value (5%)             -2.865625
Critical Value (10%)            -2.568945
dtype: float64
```




```
Results of Dickey-Fuller Test:
Test Statistic                  -1.510541
p-value                          0.528290
#Lags Used                      20.000000
Number of Observations Used    709.000000
Critical Value (1%)             -3.439607
Critical Value (5%)             -2.865625
Critical Value (10%)            -2.568945
dtype: float64
```




```
Results of Dickey-Fuller Test:
Test Statistic                  -2.470500
p-value                          0.122818
#Lags Used                      18.000000
Number of Observations Used    711.000000
Critical Value (1%)             -3.439581
Critical Value (5%)             -2.865614
Critical Value (10%)            -2.568939
dtype: float64
```




```
Results of Dickey-Fuller Test:
Test Statistic                  -1.773868
p-value                          0.393511
#Lags Used                      11.000000
Number of Observations Used    545.000000
Critical Value (1%)             -3.442406
Critical Value (5%)             -2.866858
Critical Value (10%)            -2.569602
dtype: float64
```

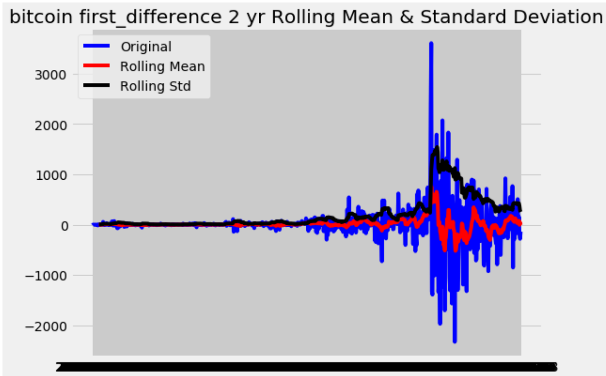
From Dukey-Fuller test of four cryptocurrencies, they all fail to reject H0 that data is a random walk since t-test and p-value fall within 5% boundary. All four price data sets of bitcoin, ethereum, ripple, and zcash have unit root and nonstationary which are random walk. Next to use first_difference method to breakdown random walk, and see if the drifts are stationary in four price data sets.

The following use 1-day difference of itself price and repeat Dukey-Fuller test.

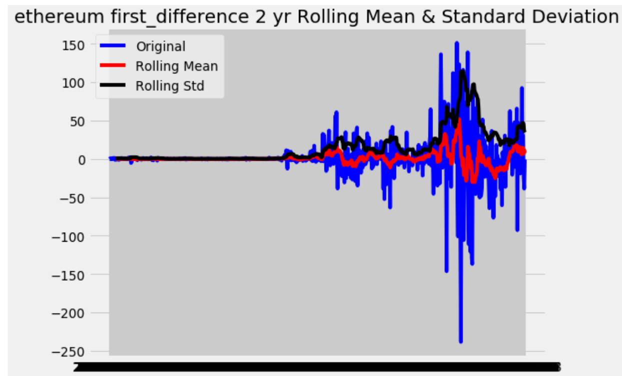
The following will be the assumption for Dukey-Fuller test:

The simple fist_diff AR(1)model is deltay_t= (β-1)y_t-1 + u_t.

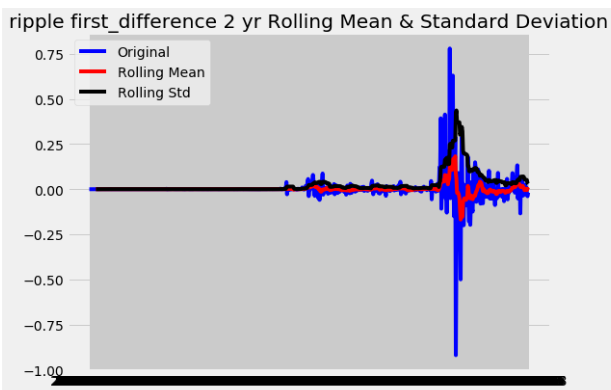
Where assuming the β <1 when it is stationary, and expectation of error terms u_t is also 0.



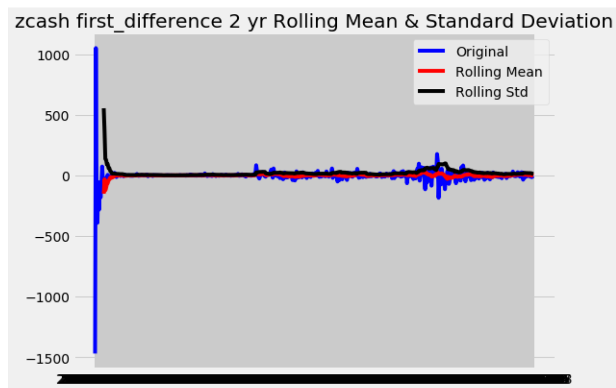

```
first_difference Results of Dickey-Fuller Test:
Test Statistic                  -4.916657
p-value                          0.000032
#Lags Used                      19.000000
Number of Observations Used    709.000000
Critical Value (1%)             -3.439607
Critical Value (5%)             -2.865625
Critical Value (10%)            -2.568945
dtype: float64
```




```
first_difference Results of Dickey-Fuller Test:
Test Statistic                  -3.951137
p-value                          0.001689
#Lags Used                      20.000000
Number of Observations Used    708.000000
Critical Value (1%)             -3.439620
Critical Value (5%)             -2.865631
Critical Value (10%)            -2.568948
dtype: float64
```



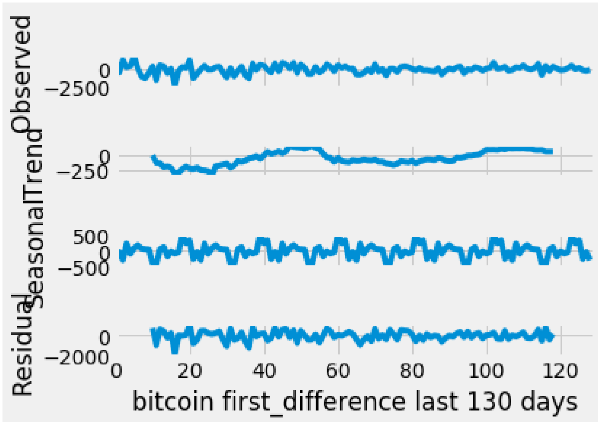

```
first_difference Results of Dickey-Fuller Test:
Test Statistic                 -6.025281e+00
p-value                         1.460947e-07
#Lags Used                      1.700000e+01
Number of Observations Used     7.110000e+02
Critical Value (1%)            -3.439581e+00
Critical Value (5%)            -2.865614e+00
Critical Value (10%)           -2.568939e+00
dtype: float64
```

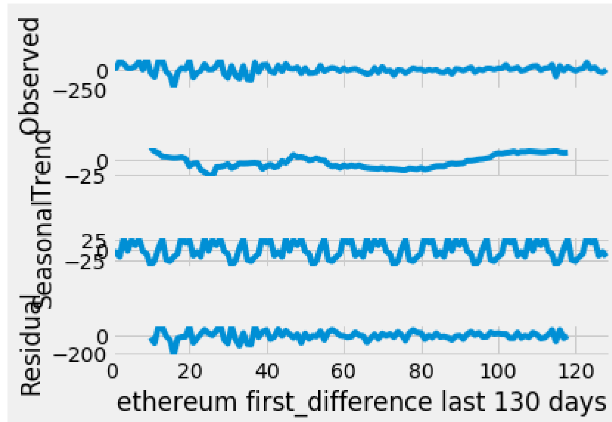


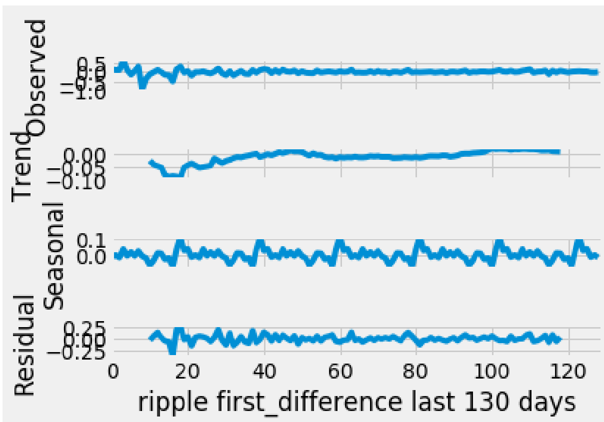

```
first_difference Results of Dickey-Fuller Test:
Test Statistic                 -9.541306e+00
p-value                         2.722015e-16
#Lags Used                      1.000000e+01
Number of Observations Used     5.450000e+02
Critical Value (1%)            -3.442406e+00
Critical Value (5%)            -2.866858e+00
Critical Value (10%)           -2.569602e+00
dtype: float64
```

From Dukey-Fuller test of four first_difference cryptocurrency's prices, all reject H0 that data is not a random walk since t-test and p-value fall out of 5% boundary. All four price data sets of bitcoin, ethereum, ripple, and zcash close to stationary. Now time series price would be possibly predictive.

The trend time series models are not suitable if the data is seasonal. And to check for seasonal data, the best way is to plot the data and observe the patterns from the chart. Python metaplot is easy to get patterns of residual, seasonal, trend and observed data. The following charts present bitcoin, Ethereum, ripple, and zcash price patterns.







The finding from bitcoin, ethereum, ripple and zcash chars, it is easily to find all of them have significant seasonal changes. All else residual, trend and observed are not showing patterns in 2 years data. Therefore, seasonal time series models will be considered to use.

The SARIMA model is a time serials model adding seasonal part of the model consists of terms. It also called seasonal ARIMA models. The SARIMA model adds a seasonally varying coefficient to the ARMA. SARIMA has 7 parameters, p,d,q,P,D,Q, S. For example, ARIMA is a special case of SARIMA, (1,1,1)(1,1,1,0). The method to find the best SARIMA model for the time series data set. Fist is to simulate the possible parameters in SARIMA model, then compare p-value and AIC. After taking out all models which fail 5% p-value test, selecting the model has the smallest AIC.

Due to computer limitation, it is impossible to select a large number of the parameter set. From previous observation, the data becomes close to stationary in first different itself price. I may assume AR(1) will be work. To left some uncertain assumption for the machine to figure it out. Since the data is seasonal data, parameters could be set monthly or weekly.

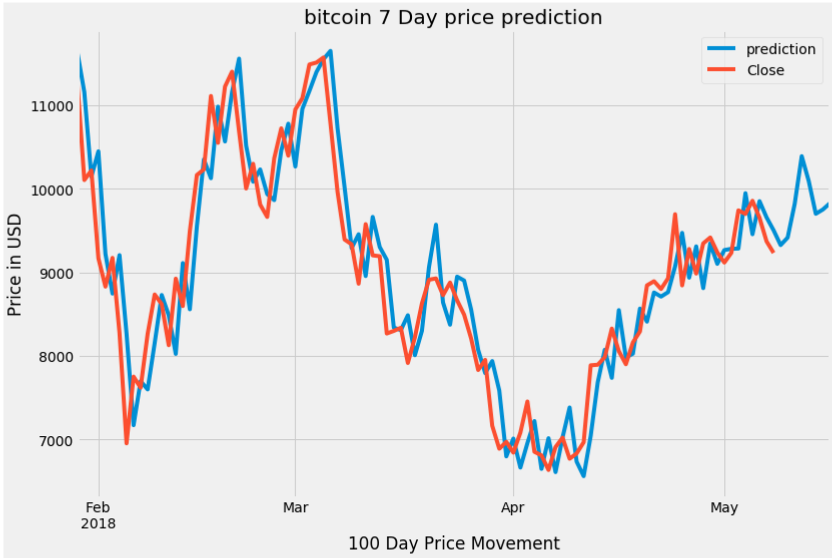
```
                                 Statespace Model Results
==========================================================================================
Dep. Variable:                              Close   No. Observations:                  730
Model:             SARIMAX(1, 1, 1)x(1, 1, 1, 12)   Log Likelihood               -5298.657
Date:                            Wed, 09 May 2018   AIC                          10607.314
Time:                                    18:34:05   BIC                          10630.279
Sample:                                05-09-2016   HQIC                         10616.174
                                     - 05-08-2018
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.9998      0.126     -7.914      0.000      -1.247      -0.752
ma.L1          0.9999      0.146      6.837      0.000       0.713       1.287
ar.S.L12      -0.0627      0.019     -3.256      0.001      -0.100      -0.025
ma.S.L12      -0.9928      0.042    -23.641      0.000      -1.075      -0.910
sigma2      1.426e+05   4.13e-06   3.45e+10      0.000    1.43e+05    1.43e+05
===================================================================================
Ljung-Box (Q):                      178.13   Jarque-Bera (JB):             12862.07
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):             775.41   Skew:                             0.87
Prob(H) (two-sided):                  0.00   Kurtosis:                        23.68
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 1.25e+26. Standard errors may be unstable.
```

After running the python program, it suggests SARIMAX(1,1,1)(1,1,1,12) by setting in monthly patterns. However, it is still uncertain if patterns breakdown in weekly would better.

```
                                 Statespace Model Results
==========================================================================================
Dep. Variable:                              Close   No. Observations:                  730
Model:             SARIMAX(1, 1, 1)x(1, 1, 1, 52)   Log Likelihood               -5039.391
Date:                            Wed, 09 May 2018   AIC                          10088.782
Time:                                    18:34:48   BIC                          10111.747
Sample:                                05-09-2016   HQIC                         10097.642
                                     - 05-08-2018
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.9978      0.437     -2.283      0.022      -1.854      -0.141
ma.L1          0.9981      0.445      2.241      0.025       0.125       1.871
ar.S.L52      -0.0970      0.047     -2.046      0.041      -0.190      -0.004
ma.S.L52      -0.8161      0.044    -18.700      0.000      -0.902      -0.731
sigma2      1.577e+05   3247.453     48.553      0.000    1.51e+05    1.64e+05
===================================================================================
Ljung-Box (Q):                      182.60   Jarque-Bera (JB):              9707.59
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):             733.64   Skew:                             0.79
Prob(H) (two-sided):                  0.00   Kurtosis:                        21.48
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

After running the python program, it suggests SARIMAX(1,1,1)(1,1,1,52) by setting in weekly patterns. The AIC is smaller than in monthly patterns. SARIMAX(1,1,1)(1,1,1,52) model is a better option for bitcoin. And Warnings for error in the weekly model also lesser than in the monthly model.

In the following, it is to use SARIMAX(1,1,1)(1,1,1,52) to predict 7 days bitcoin future price. From last 100 days, the models are able to generate the similar path as the real data.



The following steps will repeat the same steps in finding best SARIMA model in predicting ethereum, ripple, and zcash.

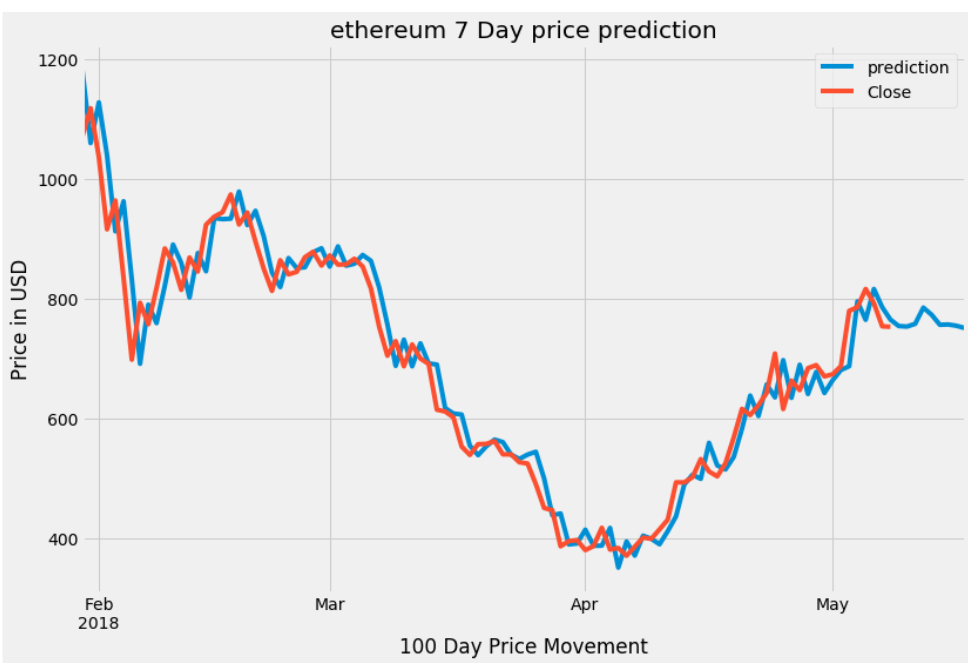
For Ethereum, the python program suggests SARIMAX (0,1,1)(1,1,1,12) in monthly setting. Now try to set the model in weekly, then compare AIC.

```
                                 Statespace Model Results
==========================================================================================
Dep. Variable:                              Close   No. Observations:                  730
Model:             SARIMAX(0, 1, 1)x(1, 1, 1, 12)   Log Likelihood               -3416.472
Date:                            Thu, 10 May 2018   AIC                           6840.945
Time:                                    03:29:05   BIC                           6859.317
Sample:                                05-10-2016   HQIC                          6848.033
                                     - 05-09-2018
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1          0.0600      0.017      3.528      0.000       0.027       0.093
ar.S.L12      -0.0724      0.018     -4.091      0.000      -0.107      -0.038
ma.S.L12      -0.9987      0.307     -3.256      0.001      -1.600      -0.398
sigma2       751.5980    226.992      3.311      0.001     306.702    1196.494
===================================================================================
Ljung-Box (Q):                      203.27   Jarque-Bera (JB):              7077.29
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):            4224.87   Skew:                            -0.66
Prob(H) (two-sided):                  0.00   Kurtosis:                        18.33
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

```
                           Statespace Model Results
==========================================================================================
Dep. Variable:                          Close   No. Observations:                     730
Model:             SARIMAX(0, 1, 1)x(1, 1, 1, 52)   Log Likelihood                 -3267.484
Date:                        Thu, 10 May 2018   AIC                              6542.969
Time:                                03:22:26   BIC                              6561.341
Sample:                            05-10-2016   HQIC                             6550.057
                                 - 05-09-2018
Covariance Type:                          opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1          0.0445      0.018      2.542      0.011       0.010       0.079
ar.S.L52      -0.1534      0.048     -3.229      0.001      -0.247      -0.060
ma.S.L52      -0.8014      0.046    -17.438      0.000      -0.891      -0.711
sigma2       826.0427     17.218     47.975      0.000     792.295     859.790
===================================================================================
Ljung-Box (Q):                      186.41   Jarque-Bera (JB):              5026.55
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):            4380.66   Skew:                            -0.52
Prob(H) (two-sided):                  0.00   Kurtosis:                        16.31
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

After running the python program, it suggests SARIMAX(1,1,1)(1,1,1,52) by setting in weekly patterns, since the AIC in the weekly model is smaller than in monthly's. SARIMAX(1,1,1)(1,1,1,52) model is a better option for predicting ethereum.

In the following, it is to use SARIMAX(1,1,1)(1,1,1,52) to predict 7 days ethereum future price. From last 100 days, the models are able to generate the similar path as the real data.



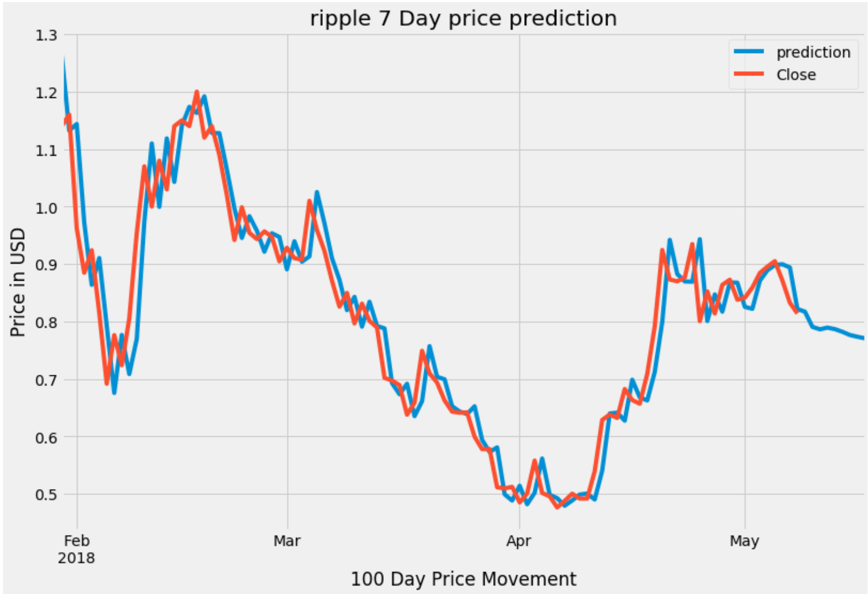
```
                           Statespace Model Results
==========================================================================================
Dep. Variable:                              Close    No. Observations:                730
Model:             SARIMAX(1, 1, 1)x(1, 0, 0, 12)    Log Likelihood                921.421
Date:                            Thu, 10 May 2018    AIC                         -1834.843
Time:                                    03:50:29    BIC                         -1816.471
Sample:                                05-10-2016    HQIC                        -1827.755
                                     - 05-09-2018
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.8552      0.031     27.930      0.000       0.795       0.915
ma.L1         -0.7906      0.040    -19.870      0.000      -0.869      -0.713
ar.S.L12      -0.1579      0.023     -6.874      0.000      -0.203      -0.113
sigma2         0.0047   4.88e-05     95.744      0.000       0.005       0.005
===================================================================================
Ljung-Box (Q):                      238.74   Jarque-Bera (JB):            198886.63
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):          191624.28   Skew:                            -0.84
Prob(H) (two-sided):                  0.00   Kurtosis:                        83.90
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

For ripple, the python program suggests SARIMAX (1,1,1)(1,0,0,12) in monthly setting. Now to try set model in weekly, then compare AIC.

```
                           Statespace Model Results
==========================================================================================
Dep. Variable:                              Close    No. Observations:                730
Model:             SARIMAX(1, 1, 1)x(1, 0, 0, 52)    Log Likelihood                912.918
Date:                            Thu, 10 May 2018    AIC                         -1817.837
Time:                                    03:51:21    BIC                         -1799.464
Sample:                                05-10-2016    HQIC                        -1810.748
                                     - 05-09-2018
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.8167      0.036     22.405      0.000       0.745       0.888
ma.L1         -0.7486      0.046    -16.402      0.000      -0.838      -0.659
ar.S.L52      -0.0382      0.063     -0.611      0.541      -0.161       0.084
sigma2         0.0048   4.05e-05    118.110      0.000       0.005       0.005
===================================================================================
Ljung-Box (Q):                      270.00   Jarque-Bera (JB):            214803.66
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):          204712.66   Skew:                            -0.92
Prob(H) (two-sided):                  0.00   Kurtosis:                        87.07
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

After running the python program, it suggests SARIMAX(1,1,1)(1,0,0,12) by setting in monthly setting, since the AIC in monthly model is smaller than in weekly's. SARIMAX(1,1,1)(1,0,0,12) model is a better option for predicting ripple.

In the following, it is to use SARIMAX(1,1,1)(1,0,0,12) to predict 7 days ripple future price. From last 100 days, the model is able to generate the similar path as the real data.



For zcash, the python program suggests SARIMAX (1,1,0)(0,1,1,12) in monthly setting. Now to try the set model in weekly, then compare AIC.

```
                                 Statespace Model Results
==========================================================================================
Dep. Variable:                              Close   No. Observations:                  558
Model:             SARIMAX(1, 1, 0)x(0, 1, 1, 12)   Log Likelihood               -3073.612
Date:                            Thu, 10 May 2018   AIC                           6153.224
Time:                                    04:00:18   BIC                           6166.197
Sample:                                10-29-2016   HQIC                          6158.291
                                     - 05-09-2018
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.8492      0.005   -155.633      0.000      -0.860      -0.838
ma.S.L12      -0.7470      0.032    -22.996      0.000      -0.811      -0.683
sigma2      4569.5225    147.771     30.923      0.000    4279.898    4859.147
===================================================================================
Ljung-Box (Q):                      347.06   Jarque-Bera (JB):             15980.63
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):               0.24   Skew:                             2.10
Prob(H) (two-sided):                  0.00   Kurtosis:                        29.19
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

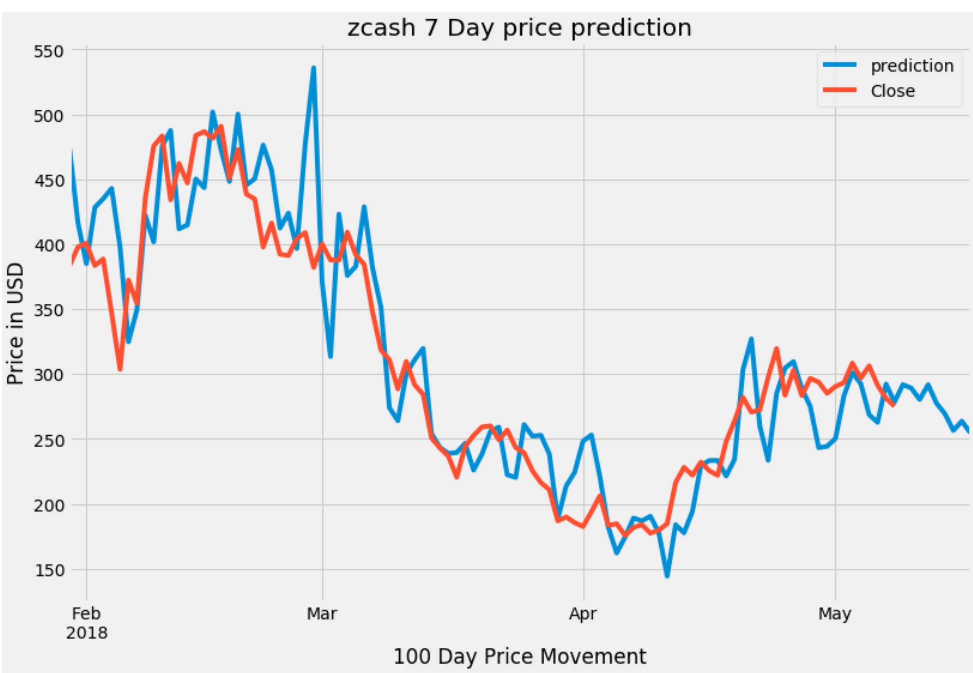
After running the python program, it suggests SARIMAX(1,1,1)(1,0,0,12) by setting in monthly setting, since the AIC in the monthly model is smaller than in weeklies. SARIMAX(1,1,1)(1,0,0,12) model is a better option for predicting ripple.

In the following, it is to use SARIMAX(1,1,1)(1,0,0,12) to predict 7 days ripple future price. From last 100 days, the model is able to generate the similar path as the real data.

```
                                 Statespace Model Results
==========================================================================================
Dep. Variable:                              Close   No. Observations:                  558
Model:             SARIMAX(1, 1, 0)x(0, 1, 1, 52)   Log Likelihood               -2879.464
Date:                            Thu, 10 May 2018   AIC                           5764.927
Time:                                    04:01:08   BIC                           5777.900
Sample:                                10-29-2016   HQIC                          5769.994
                                     - 05-09-2018
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.8367      0.004   -205.677      0.000      -0.845      -0.829
ma.S.L52      -0.5409      0.030    -18.245      0.000      -0.599      -0.483
sigma2      5084.3342    141.133     36.025      0.000    4807.719    5360.950
===================================================================================
Ljung-Box (Q):                      110.11   Jarque-Bera (JB):             22752.97
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):               0.24   Skew:                             2.63
Prob(H) (two-sided):                  0.00   Kurtosis:                        35.46
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

After running the python program, it suggests SARIMAX(1,1,0)(0,1,1,52) by setting in weekly setting, since the AIC in the weekly model is smaller than in monthly's. SARIMAX(1,1,0)(0,1,1,52) model is a better option for predicting zcash.

In the following, it is to use SARIMAX(1,1,0)(0,1,1,52) to predict 7 days zcash future price. From last 100 days, the model is able to generate the similar path as the real data.



Conclusion, bitcoin, ethererum, ripple, zcash have strong significant seasonal pattern, for which SARIMAX model will be a good option in predicting cryptocurrency price. Since cryptocurrency prices has high volatile and dose not have long history to observe, therefore, breaking down the season periods will get a better perdition. However, SARIMAX model is very unstable, so the model will not work in predicting long term price movement. The parameters need to be adjust often. In the real world, cryptocurrency volatility created by inspirators, not relative to economic in real world. cryptocurrency prices affect by every time regulation announcement and risk of being hack. Combining all reasons, SARIMA should be reasonable in short-term price prediction.