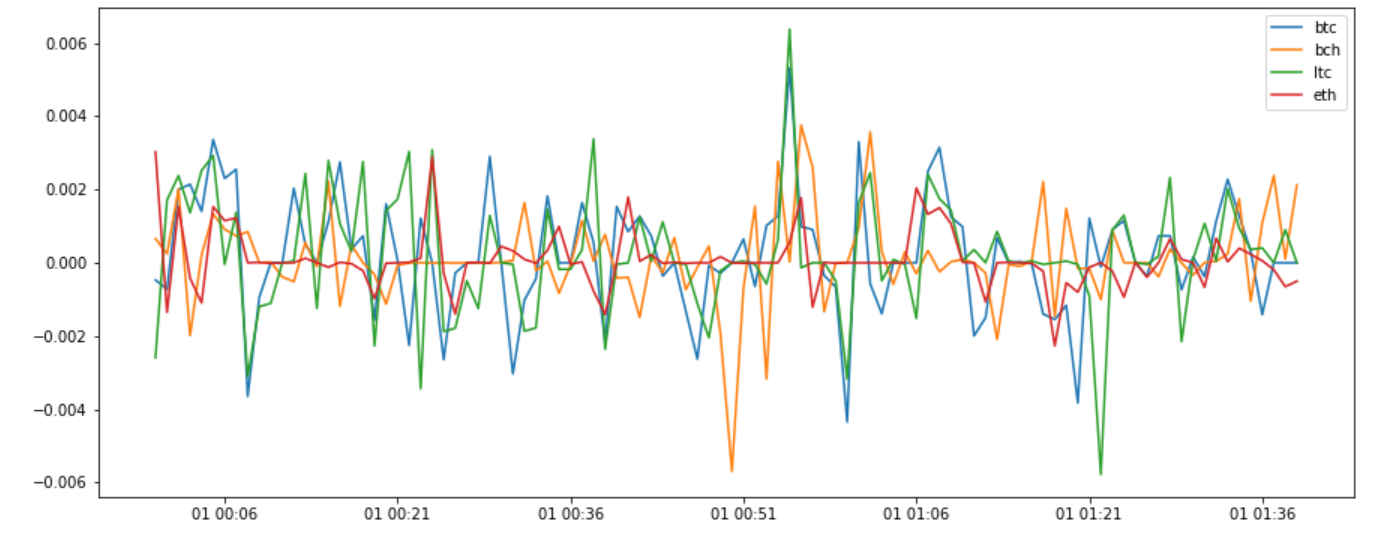
**VAR(vector autoregression) model:**

A naïve observation of the log returns of four crypto currencies implies the possible existence of high correlation among different currencies. Below is a plot of the four data’s log return from a random slice of timestamp in 1-minute scale, it is seen that in certain moments, the change in log return of one currency leads that of the others.



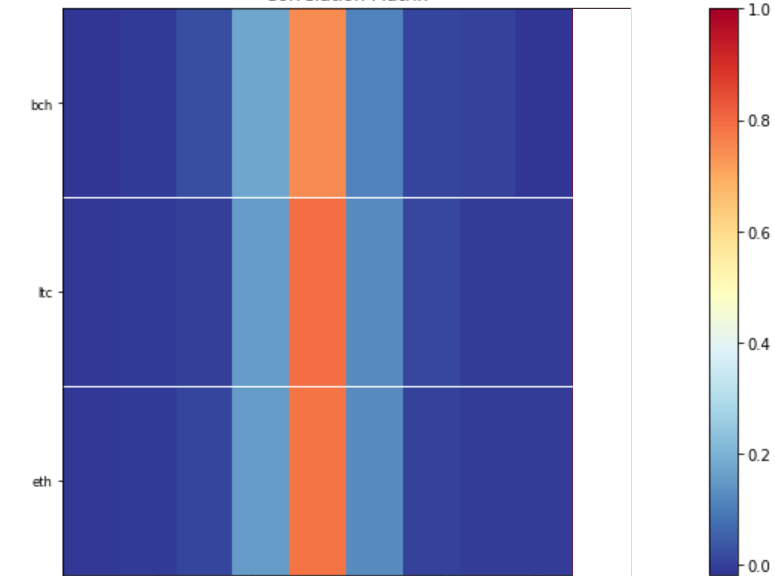
To further prove this hypothesis, we investigate the cross-correlation between different currencies, the graph below shows the matrix of the cross correlation between bitcoin and the other three currencies with different leads/lags in 1-minute scales, the x-axis represents the shifts of the other currency in the order of [-4, -3, -2, -1, 0, 1, 2, 3, 4], for example, the element in position [1, 1] of matrix represents the correlation between BTC and BCH.shift(-4),

and the element in position [1, 2] represents the correlation between BTC and BCH.shift(-3), etc. It is seen that the bitcoin has maximum correlation with all other currencies in zero shift, aka in-phase; and that litecoin has higher in-phase correlation with bitcoin than other currencies.

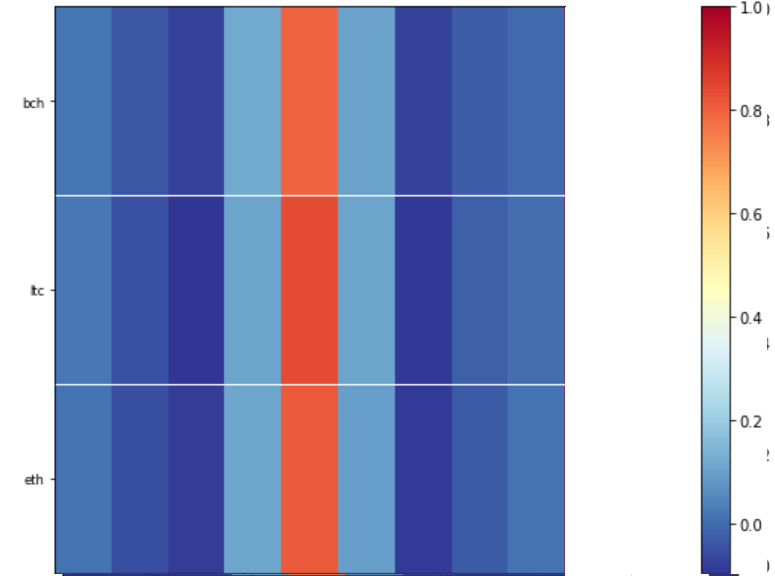
A screenshot of a cell phone

Description automatically generated

Expanding the time scale from 1-minute to 15-minute:



Higher correlation values are shown in all relationships. This trend continues for expanding the time scale to 60-minute:



A note-worthy fact in 60 minute-scale graph is that as the phase get closer, the correlation initially drops down and then rise up, showing a sigh of seasonality within the patterns.

Model building:

A VAR(vector autoregressive model) is one of the foundamental multivariant models in time series analysis. It could be viewed as the vector version of the univariant autoregressive(AR) model, which has the following general form:

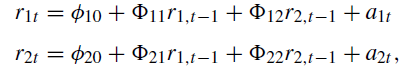


The AR model represents time serie values at time t as a function of weighted combination of its previous time values.

In VAR model, all the variables are represented in vector form, a VAR(1) model is in the form:

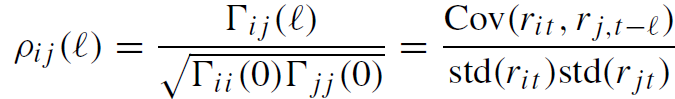
  
where rt Φ0 rt-1 and at are all of k-dimension, and **Φ** is a matrix of shape k\*k

when k =2, for example, the expanded VAR(1) model is:



The model shows that the value of time series rk is a combination of the previous time values of all {r1, r2, ..rk}

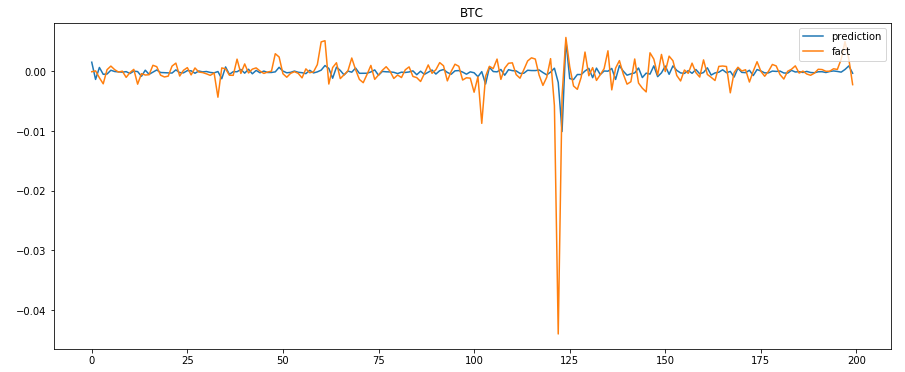
Regarding to our pratical implementation, since there are four crypto currencies, we choose k=4 for the model. To determine value of p is determined by the cross-correlation matrix of rr:

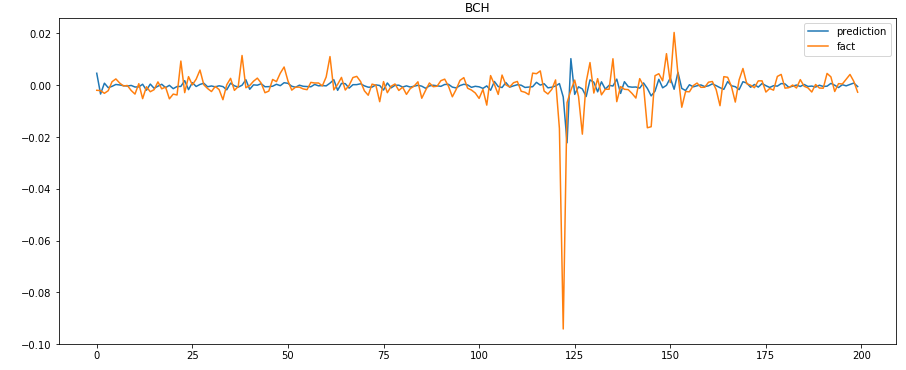


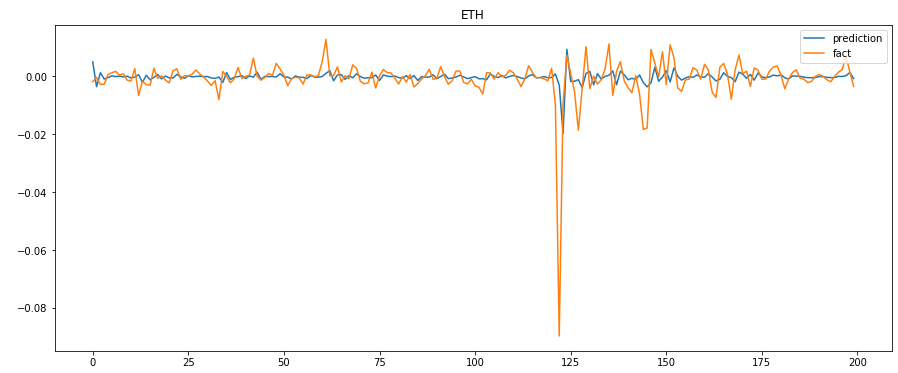
In which Γ(l) is the lag-l cross-covariance matrix of rt

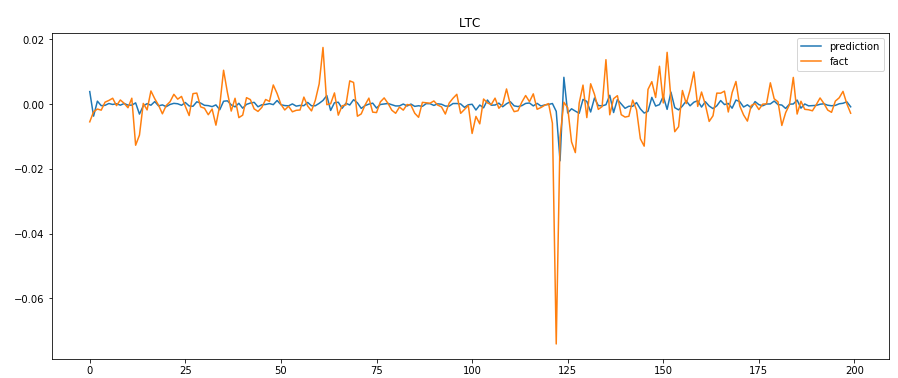
Performance:

In practice we fit a VAR(3) model with hourly resampled data. The following is the model’s prediciton of log return compared with ground truth in 200 time units









If treating the prediction as a classification task, which means predicting whether the log return is positive or negative at next moment, the accuracy for the 4 kinds of crypto currencies are:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | BTC | BCH | ETH | LTC |
| Accuracy | 0.545 | 0.61 | 0.575 | 0.605 |