ISYE6402 Final Project

Cryptocurrencies: Pricing, Volatility and Trading Strategy Using Time Series Analysis

Group 7

Minghan Xu

Nirmit Chetwani

Tianyi Liu

Task Assignment

1. Minghan Xu: Data collection, background on blockchain and cryptocurrencies, exploratory analysis, ARIMAX, VAR
2. Nirmit Chetwani: Data wrangling, ARIMA, programming on utility functions, dynamic trading strategy using ARIMA and XG-Boost
3. Tianyi Liu: literature on cryptocurrency pricing, volatility and external market indices, Heteroscedasticity models, abstract and conclusion

Table of Contents

Abstract

As the popularity of cryptocurrency grows over time, we are interested in exploring the patterns of their prices and use those to develop trading strategies. This report explores two components: first is performing time series analysis for forecasting the baseline price and volatility of cryptocurrency: Dynamic and rolling-window implementation on ARIMA, ARIMAX with exogenous factors for price forecasting; VAR for the lead-lag effect within cryptocurrencies and between cryptocurrency and exogenous factors; Heteroscedasticity model for conditional volatility. The second is using those forecast-models to develop dynamic trading strategies under some assumptions, and compare the profits generated through those on the cryptocurrencies and dow-jones index. The models used for trading were ARIMA and Gradient Boosting.

Keywords

Cryptocurrencies, Time Series, Bitcoin, Pricing, Volatility, Forecast, Prediction, Trading Strategy

1. Introduction

The advancement of blockchain technology has enabled the development, distribution, and transaction on cryptocurrencies. Cryptocurrency is a virtual currency that uses cryptography for currency creation and transaction verification. Relying on a distributed system for transactional activities, cryptocurrency is a completely decentralized system and has many benefits including accurate tracking, transaction cost reduction and a permanent ledger [1]. As an emerging technology and a possibly reliable medium of exchange in the near future, cryptocurrencies are drawing huge attention from global investors as an alternative financial instrument to diversify their investment portfolios. This project collects, analyzes and predicts pricing and volatility of major cryptocurrencies and propose a simulated trading strategy between traditional securities and cryptocurrencies.

2. Background

2.1. Blockchain

The blockchain is the underlying technology that enables creation and transactions of digital currencies. Unlike the traditional credit card transaction clearance requires a centralized agency such as Visa or Master for payment settlements, participants of blockchain can confirm the transaction by using the pier to pier (P2P) network. When someone in the network initiates a request of a transaction, the message is broadcasted to the joint network comprises millions of user nodes. These distributed nodes respond and serve as the agency for verification, creates a ledger block and add such block to the permanent ledger blockchain. Hence the transaction record is permanent and secured. The applications of blockchain are not restricted to digital currencies. It can become an enabler for more secure and cost-effective systems for digital transactions and records such as healthcare records and election voting [2].

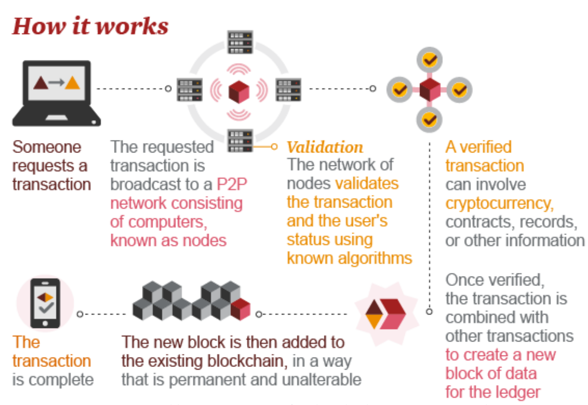

Figure 1 Illustration of Blockchain Network [3]

2.2. Cryptocurrencies

Different from gold, cryptocurrency is a virtual currency that is believed to carry no intrinsic value [4]. Its market value reflects market's assessment of bitcoin's value as a medium of exchange and a store of values, while the realization of the second value usually depends on the utilities of the first one. The supply of cryptocurrencies is not administered by any central bank or government body. However, the supply of any given cryptocurrency is not unlimited. Bitcoin (BTC), for example, has specified supply limit of 21 million units. Generally speaking, the creation of cryptocurrency is pre-defined by its underlying generation algorithm [5].

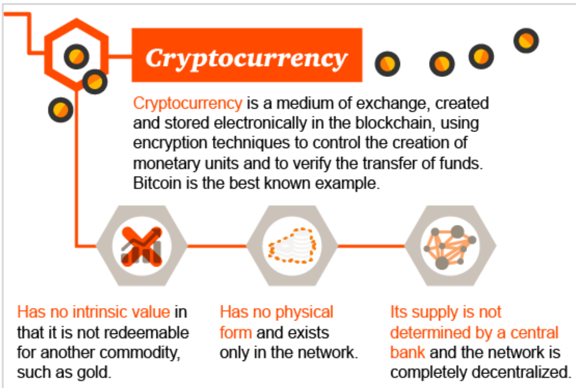

Figure 2 Characteristics of Cryptocurrency [3]

2.3. Production of Cryptocurrencies

New cryptocurrencies are created as a reward for verifying previous transactions (defined as mining). It involves using computer systems in the distributed network to compile recent transactions into the blockchain and solve extremely challenging mathematical problems. Such process is extremely expensive in terms of computational time and power consumption from the computer system. Though lacking physical form, the production/mining of new units of cryptocurrency incurs energy and other related costs. The relative cost and easiness of production between cryptocurrencies are the determinants of their market values.

2.4. Market Capitalization and Bitcoin

As of end February 2018, the total market capitalization of top 5 cryptocurrencies is almost equivalent market capitalization of Bank of America ($301B). Bitcoin, being the first decentralized virtual currency, has the longest history of the transaction and largest market capitalization ($190B) among all cryptocurrencies. It is observed that other cryptocurrencies follow similar trends in their market capitalization movements [6].

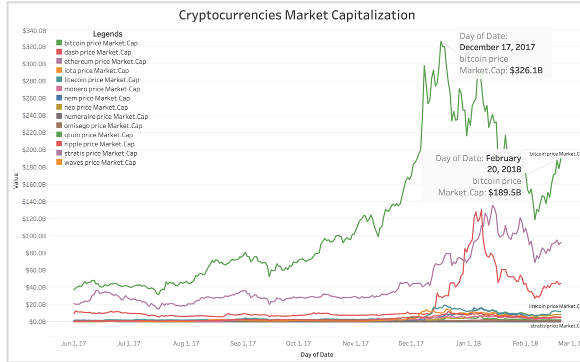

Figure 3 Cryptocurrencies Market Capitalization

2.5. Cryptocurrency Pricing

The price of cryptocurrencies, specifically Bitcoin, gained a huge momentum from the mid-2017. On Jan. 2017, China central bank first tightened the controls on the mining and transaction of bitcoins, followed by U.S. Securities and Exchange Commission's rejection on bitcoin exchange-traded fund. Despite early challenges, bitcoin climbed above $2,000 for the first time in end May 2017 and passed $3,000 just weeks after. In the first week of Sep. 2017, bitcoin price exceeded $5,000 mark and continued its upward trajectory with the bullish sentiment in the market. On Dec.17th 2017, Bitcoin reached its historical high of $19,783. Its year-over-year return at that time was approximately 1,600%. Since the start of Jan. 2018, Bitcoin price plummeted by 60% from the historical high.

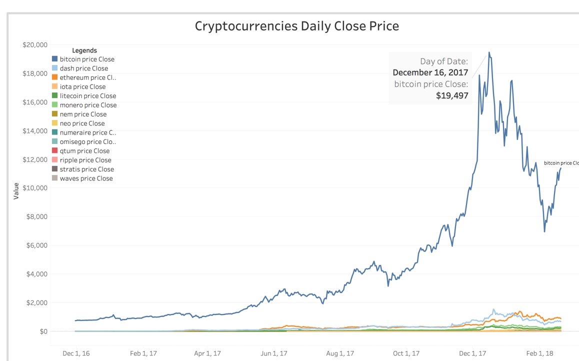

Figure 4 Cryptocurrencies Daily Close Price

2.6. Cryptocurrency Volatility

Different from stock and future, the risk and volatility of cryptocurrency are much higher. There are several reasons for such abnormal volatility. First because of the decentralized characteristic of cryptocurrency, one of the most important characteristics of Bitcoin.

Decentralization is the major innovation which makes Bitcoin safe and popular. However, Bitcoin, different from traditional currencies such as U.S dollar, are not controlled by any government. Thus, the stationarity of cryptocurrency cannot be guaranteed.

Additionally, Bitcoin is vulnerable to external clashes also because of decentral effect. Incidents such as hacker attacks, coding mistake, fall of major traders, changes in regulation and transaction size will dramatically affect the value of Bitcoin [7]. For instance, the birth of "WannaCry", a Bitcoin extortion virus outbreak during Summer 2017 resulted in the doubled price of Bitcoin. Similarly, the ban of cryptocurrencies in countries such as China also fluctuated the price of cryptocurrency.

2.7. Cryptocurrency Forecast

Like stock and future, the forecast of bitcoin price and volatility is always a popular topic. There are some existing approaches for forecasting the baseline price of cryptocurrency. For instance, nonlinear methods, such as kernel regression model, exponential autoregressive models, artificial neural network, Bayesian neural networks, predict price and volatility [8] or VAR model for dataset incorporated gold price and U.S dollar exchange rate to predict volatility [9]. Experiment on testing different GARCH model to gain insights of heteroscedasticity [10].

3. Methods

3.1. Data Sources

Cryptocurrency Historical Prices

Acquired from Kaggle, the dataset contains the daily open, high, low, close prices, transaction volume, the market capitalization of top cryptocurrency including Bitcoin, Ethereum, Ripple, Bitcoin Cash, etc. Different from major market securities/commodities where weekend data is not available due to market-close, cryptocurrencies are traded on 24/7.

Oil Price (West Texas Intermediate)

West Texas Intermediate (WTI) crude oil is the underlying commodity of the New York Mercantile Exchange (NYMEX) oil futures contracts. The price of WTI is usually referenced as the benchmark of oil pricing.

Economic Policy Uncertainty Index for the United States

Collected from Economic Research, Federal Reserve Bank of ST. Louis, this daily index measures economic policy uncertainty for the United States. A higher index value represents a higher uncertainty related to economic policy. The index is published daily by a joint research team from Northwestern University, Stanford University and University of Chicago [11].

CBOE Volatility Index (VIX)

Acquired from Chicago Board Options Exchange (CBOE), VIX index is a popular measure of the stock market's expectation of volatility implied by S&P 500 index options. It is often referred to as the fear gauge as it formulates a theoretical expectation of stock volatility in the near future [12].

Gold Price

Gold has been used throughout history as money and been a relative standard for currency equivalents [13]. Acquired from World Gold Council, the data contains daily gold price in units of US dollar, Euro, Japanese Yen, Pound Sterling, Canadian dollar, Swiss franc.

Don Jones Industrial Average (Dow)

Dow is a price-weighted stock market index of 30 largest public companies (blue chip stocks) based in the United States and traded on New York Stock Exchange (NYSE) and Nasdaq. It is designed to serve as a proxy for U.S. economic outlook. The companies included in Dow are well-known names like Apple, Home Depot and Walmart [14].

NASDAQ Composite (Nasdaq)

Nasdaq composite is a market capitalization-weighted index for about 3,000 common equities listed on Nasdaq Stock Exchange. It does not limit its scope for U.S. based companies and has a broader evaluation of the market condition. The technology sector has the heaviest weights in this composite index [14].

3.2. Data Wrangling

We initially mined daily price data for all ~20 cryptocurrencies, starting May 2013; but since we were primarily interested in analyses of recent trends, we truncated the data from Jun. 2017. We also zeroed down our interests on top 5 cryptocurrencies by market capitalization – Bitcoin, Ethereum, Ripple, Litecoin, and Neo.
Since our eventual objective was to predict the currency prices accurately for a day in advance, we focused our attention on the closed-price time series data for cryptocurrencies as well as some conventional stocks/financial indices like Gold, Oil, VIX, and Dow-Jones.

For predictions of any day P_N, we use data for the past 30 days $\{V_{N-1}, V_{N-2}, \dots, V_{N-30}\}$ (here, Vi is the value at lag i) as we did not want to use very large windows (say 90 days) to ensure that we capture only the recent behavior of currencies, while at the same time we did not use a very small window (say 7 days) to be able to capture enough trend and volatility patterns. This method was implemented in R, allowing for flexibility of the training period and number of days for which the forecast is to be made that are passed as parameters in the function (set at 30 days and 1 day for our case).

This also allows us to analyze the distribution of performance measures across prediction time-frame and conclude if our model does well.

3.3. Exploratory Data Analysis

By Market capitalization as of Feb. 20th, 2018, top 5 cryptocurrencies in the sequence of market capitalization are Bitcoin, Ethereum, Ripple, Litecoin and Neo. These 5 currencies contribute more than 90% of the total capitalization of cryptocurrency market.

All cryptocurrencies have ridden the tides of market optimism on bitcoin and blockchain technology since summer 2017. Interestingly, though in the long-run, cryptocurrency prices move with coherent directions. In the short trading period, however, the prices on different cryptocurrencies could possibly move in opposite directions.

For example, with the falling of bitcoin price from the historical high of $19,000 since the late Dec. 2017, the prices of Ethereum, Ripple, and Neo kept the upward trends for additional weeks and reached their historical high in the mid-Jan. 2018. This observation may suggest some lead-lag relationship between cryptocurrencies. If investors are pessimistic about a specific cryptocurrency and concerned about the potential loss, they may move the assets to other more stable cryptocurrencies and drive their prices.

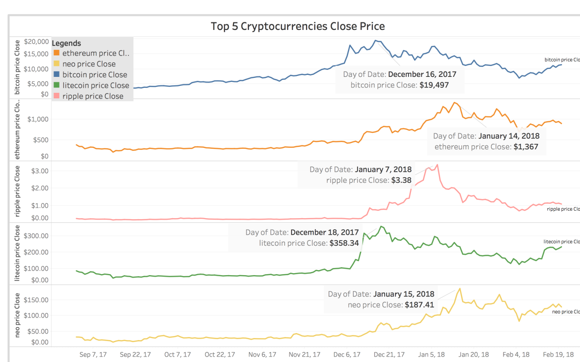

Figure 5 Top 5 Cryptocurrencies Close Price

The price volatility seems correlates with the daily transactional volume for cryptocurrencies.

This could be explained that when prices experience large variations, investors tend to be more active in trading activities of buying and selling to capture short-term opportunities or step further loss, which may further impact market sentiments and prices.

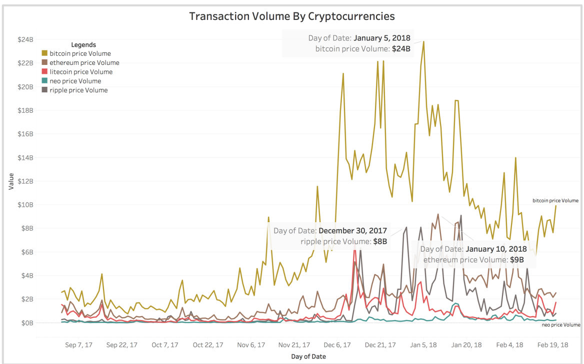

Figure 6 Daily Transaction Volume of Cryptocurrencies

External factors may drive or at least help predict the movement of Bitcoin and other cryptocurrencies. The following graph illustrates when Bitcoin price was plunging, and investors were suffering from huge potential loss, the VIX index and political uncertainty (usepunidex) moved in the opposite directions, showing a stronger concern about the overall risk and policy of the market. Such observations could suggest using external factors to predict the price of cryptocurrencies.

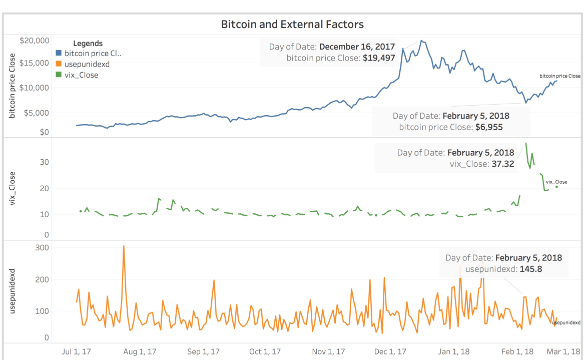

Figure 7 Bitcoin Price and External Factors

3.4. Hypothesized Model

From extensive literature study and exploratory data analysis, the team proposes the following approaches in predicting cryptocurrency pricing and volatilities using Time Series Analysis:

1. **ARIMA**: Future cryptocurrency's price depends on past lags (AR) and past random shocks (MA)
2. **ARIMAX**: An extended version of ARIMA which incorporates exogenous factors to improve prediction of the cryptocurrency's price.
3. **VAR**: Price of one Future cryptocurrency depends on past lags of other cryptocurrencies and related exogenous factors.
4. **GARCH** & **ARMA-GARCH**: Future cryptocurrency's conditional volatility depends on realized volatility and past shocks. Joint GARCH estimates the conditional mean in conjunction with conditional variance.

3.5. ARIMA

ARIMA is the most basic, yet powerful autoregressive time-series modeling approach that used past values of both the baseline values and error estimates to estimate the forecast value. In our case, let "i" be the index of currency, and let "t" be the index of time (in our case, it is daily index). The ARIMA (p, d, q) model can then be defined as:

$$Y_{i,t} = \emptyset_{i,0} + \emptyset_{i,1} Y_{i,t-1} + \dots + \emptyset_{i,p} Y_{i,t-p} + Z_{i,t} + \theta_{i,1} Z_{i,t-1} + \dots + \theta_{i,q} Z_{i,t-q} + \epsilon_{i,t}$$

Here, $Y_{i,t}$ can be time series of closed-price, or any differenced form $Y_{i,t} = \Delta P_{i,t} = P_{i,t} - P_{i,t-1}$ of i^{th} currency at time "t". In the above equation, the lagged part of $Y_{i,t}$ is the autoregressive part and lagged part of $Z_{i,t}$ is the moving average part. The model parameters $\{\emptyset\}$ and $\{\theta\}$ are estimated using MLE estimates, while the optimal order (p, q) and the optimal differencing-operator are estimated by minimizing the AIC values for model fits by doing a grid-search.

As a part of our initial analysis with ARIMA, we tried to use it directly on the cryptocurrency prices. The plots for these clearly showed a trend (discussed above) which depicted a need of differencing. We also saw sudden spikes in the prices, where log-transformations would be helpful. It can be seen from Figures 8 and 9.

Accordingly, we took the log of prices, and fit ARIMA to moving training windows of 30 days to forecast baseline price of currencies for the subsequent day. Below are the trends of the forecast values of Bitcoin and Dow-Jones Index against the actual values from July 2017 to Feb 2018. We clearly see that for most of the cases, our forecasts are close to the actual values, and they follow the values with a lag of one/two days. We also notice that the actual values usually fall within the 95% confidence band of forecasts. Also, the plots for other cryptocurrencies look similar.

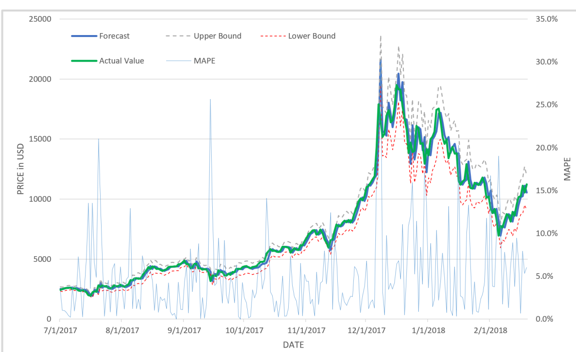

Figure 8 ARIMA predictions vs actual and MAPE for BTC

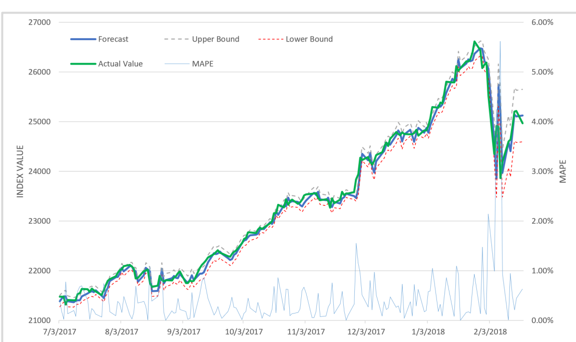

Figure 9 ARIMA predictions vs actual and MAPE for DJI

Results: We evaluate the performance measures, specifically the MAPE to gauge how good the fit of models is. From the MAPE distribution, we see that for ~87% of 235 forecasts were within 10% error range, while for DJI it was ~99%. Clearly, DJI being more stable and showing lesser fluctuations was more predictable.

We also looked at the (p, d, q) values for different currencies in different timeframes. In Figure 10, we see the distribution of the values that we get for 235 models that we fit for Bitcoin. We see that ~80% of the time, it is either the past value (one-lag) of currency that dictates the future – either directly (p = 1 and d, q = 0) or indirectly (d = 1 and p, q = 0). We also looked at the coefficients and p-values related to different models, and ~97% of them were statistically significant.

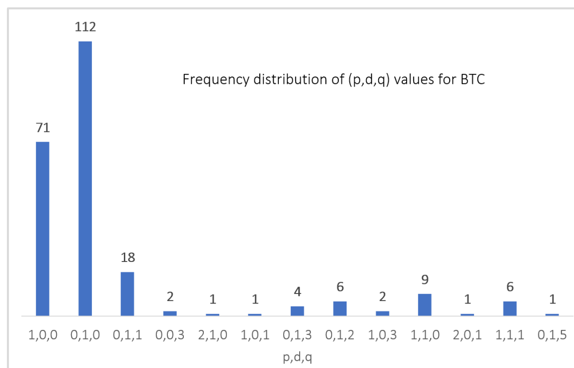

Figure 10 Distribution of Optimal (p,d,q) for BTC

In figure 11, is the output of a sample BTC-ARIMA model, which has a t-value of 5.62, showing that it is statistically significant. Also, we see the ACF/PACF of the sample model in figure 12, which clearly indicate that the time series is stationary, and the model abides by the assumptions. The normality and ARCH tests also supported the results.

```
Call:
arima(x = time_series, order = c(1, 0, 0), method = "ML")

Coefficients:
         ar1  intercept
      0.7296     7.8643
s.e.  0.1300     0.0243

sigma^2 estimated as 0.001427:  log likelihood = 55.33,  aic = -104.66
```

Figure 11 Sample output of ARIMA model for BTC

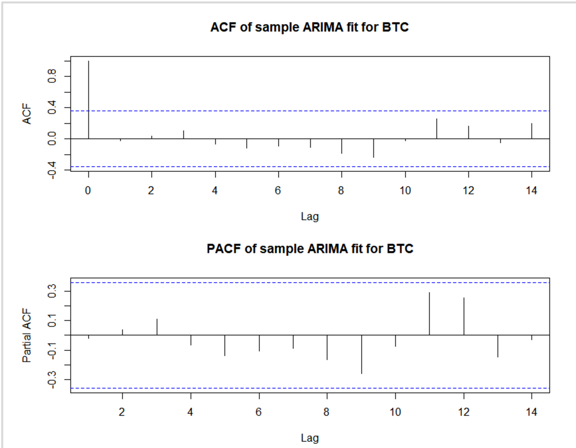

Figure 12 ACF and PACF for sample ARIMA - BTC

We will be using these predictions for simulating a trading scenario to gauge the profits we can make by investing in Bitcoins/DJIS (a virtual stock that just follows the Dow-Jones index) under certain constraints.

3.6. ARIMAX

ARIMAX extends ARIMA models through the inclusion of exogenous variable X. The ARIMA (p, d, q) model can then be defines as:

$$Y_{i,t} = \emptyset_{i,0} + \emptyset_{i,1}Y_{i,t-1} + \dots + \emptyset_{i,p}Y_{i,t-p} + Z_{i,t} + \theta_{i,1}Z_{i,t-1} + \dots + \theta_{i,q}Z_{i,t-q} + \beta X_{i,t} + \epsilon_{i,t}$$

Here, $X_{i,t}$ can be time series data of external variables that the team investigated earlier which includes: VIX, Gold price, Oil price, Dow index, Nasdaq index and economic policy uncertainty index. Bitcoin

The implementation is similar to ARIMA, and done by selecting the order of (p,d,q) based on minimal AIC values. The model parameters $\{\emptyset\}$, $\{\theta\}$ and $\{\beta\}$ are estimated using conditional-sum-of-squares to find starting values, then maximizing the likelihood function (MLE).

As the log transformation stabilizes the variation in the Bitcoin price and addresses the scale issue between different exogenous predictors, we fit the ARIMAX model similarly with a moving training window of 30 days on Bitcoin price plus exogenous factors as a data matrix in the same time period to make 1-step ahead forecast on Bitcoin price.

Results: Examining the performance metrics across 235 predictions, the model performs reasonably well: actual observations of Bitcoin price stay within the 95% confidence band of the prediction, the mean MAPE is ~ 5.5% and ~84% of the predictions are within 10% error range.

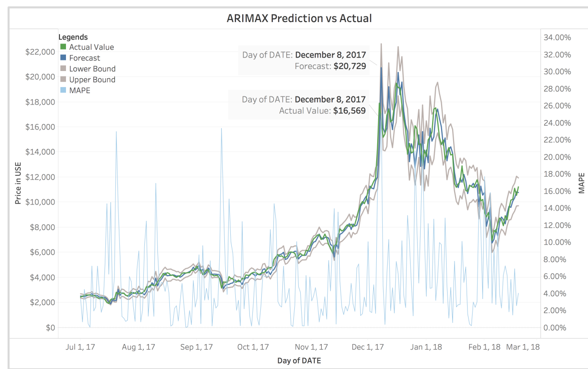

Figure 13 ARIMAX predictions vs actual & MAPE for BTC

Compared with ARIMA approach, exogenous factors do not seem to offer significant improvement over ARIMA. Instead, additional noise may be introduced in the fitting process and resulting a slightly worse prediction accuracy.

3.7. VAR

VAR model is the extension of the univariate autoregressive model on multivariate time series and can be useful to capture the dynamic behavior of financial time series between instruments. The formulation of VAR can be defined as:

$$Y_t = c + \pi_1 Y_{t-1} + \dots + \pi_p Y_{t-p} + \epsilon_t$$

Here, Y_t is a $k \times 1$ vector of endogenous variables that could include the close price of cryptocurrencies and other identified market indices like VIX and political uncertainty at a time *"t"*. The coefficient π_i is the coefficient matrix of i-th lag with the dimension of $k \times k$.

The implementation of VAR requires stable multivariate time series. The close price of top 5 cryptocurrencies and external factors including VIX, gold price, oil price, Dow index, Nasdaq index and economic policy uncertainty are applied with differenced log transformation to fit the VAR model.

With the same moving window of training, the VAR model uses OLS to estimate the coefficients of each equation. Similar to earlier implementation, the fitted model then makes 1-step ahead forecast on Bitcoin price.

Results: Examining the model performance, the model gives acceptable predictions on Bitcoin price except for the period where cryptocurrencies experience very large variations. Hence the VAR model may amplify the effects of such volatility. For example, almost all cryptocurrencies start to surge in price from early Dec. 2017. The prediction of Bitcoin during that period tends to follow the optimistic outlook and over-estimate the pricing. This VAR model has mean MAPE of ~12.61% which is worse than ARIMA.

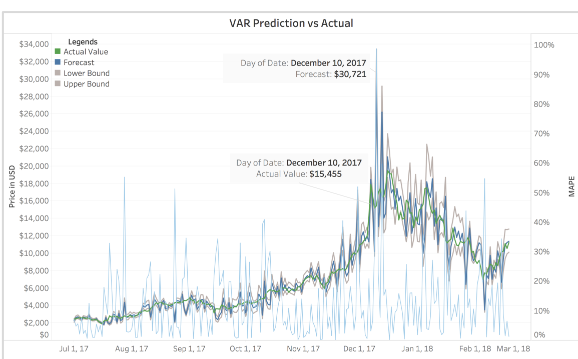

Figure 14 VAR predictions vs actual and MAPE for BTC

VAR model has the limitation of only taking into the consideration of autoregressive (AR) components for multivariate time series. For Bitcoin, the interactions between endogenous factors may not be significant. To force the model fit these additional factors could possibly introduce unnecessary noise in prediction.

3.8. GARCH

Till now, we have only looked at baseline predictions of prices and indices. If we can effectively forecast the volatility of cryptocurrency in addition to the baseline price, then we could predict the risk of investing cryptocurrency and confirm the uncertainty in decision making. Therefore, we tried to investigate whether the price of cryptocurrency has heteroscedasticity. If heteroscedasticity exists, GARCH model will be more likely to capture the volatility behavior than models assuming constant variance [15]. In the following analysis, we will take Bitcoin for example.

After fitting the ARIMA model for log return of cryptocurrency, we first analyzed the residual of ARIMA model to see whether heteroscedasticity effect exist. Since we started our training window at Jun. 1st, 2017, the total number of training window will be 234. And according to the Ljung-Box test, we focused on the training window with the p-value less than 0.05. And after fitting the model and testing Ljung-Box on all of those fitted models, we get 28 training windows with the p-value less than 0.05, rejecting the null hypothesis. The following are the ACF plots of squared residual for those training windows with the p-value less than 0.05, which means rejecting the null hypothesis of uncorrelated.

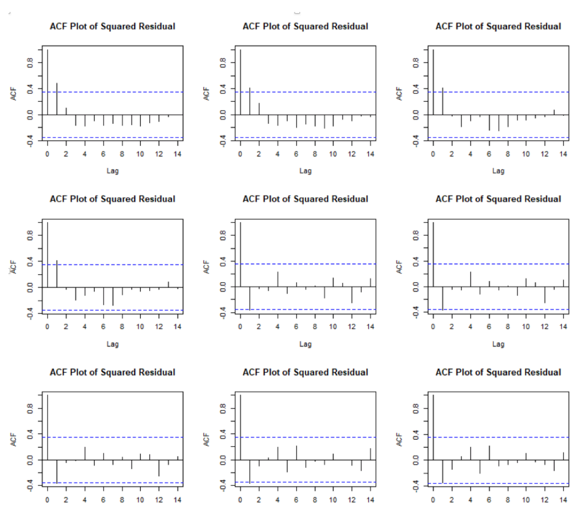

Figure 15 ACF of ARIMA Squared Residuals

According to the ACF plots of squared residual, we can see that even though those training windows are associated with the p-value less than 0.05 for Ljung-Box test, there are not so many significant spikes in the ACF plot, only the spike of lag 1 is out of the interval.

Although the analysis above suggesting that GARCH model might not be appropriate, we still tried to find an ARMA-GARCH model with optimal BIC.

The joint GARCH model are very useful to model the conditional expectation along with conditional variance. And ARMA-GARCH model is one of the most common used joint models currently. The ARMA(p,q)-GARCH(m,n) model is defined as:

$$Y_t = \mu + \emptyset_1 Y_{t-1} + \dots + \emptyset_p Y_{t-p} + Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$$

$$Z_t | F_{t-1} \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \gamma_0 + \gamma_1 Z_{t-1}^2 + \dots + \gamma_m Z_{t-m}^2 + \beta_1 \sigma_{t-1}^2 + \beta_n \sigma_{t-n}^2$$

Since the experiment on individual train window is both computation and time consuming, we fitted an ARMA-GARCH model on the whole dataset. The optimal ARIMA model is ARIMA(0,0,0) and the ACP plot of both residual and squared residual represent a white noise process.

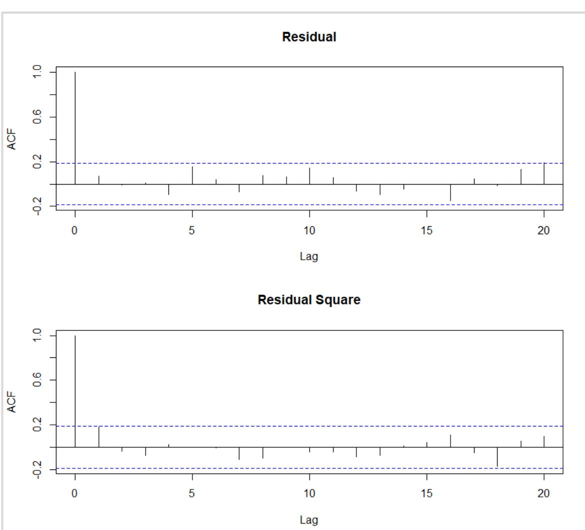

Figure 16 ACF of Residual and Squared Residual

After refining GARCH and ARMA orders for joint model, the final model we got is ARMA (0,0)-GARCH (1,1). But the performance of ARMA-GARCH model is very poor.

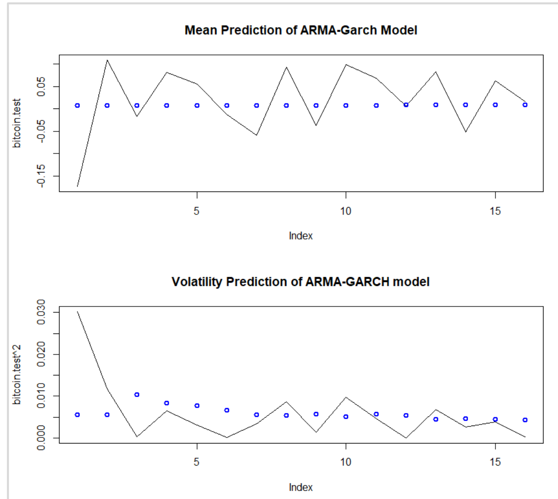

Figure 17 Conditional Mean & Variance Prediction

Since the performance of ARMA-GARCH model is not satisfactory, we think that the heteroscedasticity might not be able to be model by GARCH model appropriately. Then we turned to the second method to use a non-parametric fit for the variance. For this part, we tried both local polynomial (blue) and GAM (red). But neither of them can provide insights for volatility clusters.

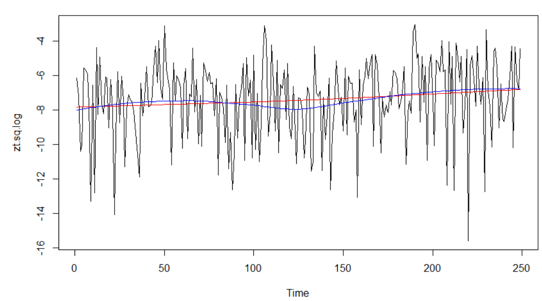

Figure 18 Local Polynomial Regression and GAM

According to the result we get, we conclude that log return of bitcoin is quite likely to be a white noise process without heteroscedasticity. Similarly, analysis has been done for Ethereum and Ripple, the log return of price is very likely to be a white noise process without heteroscedasticity as well. Therefore, we decide to not use any heteroscedasticity model, assuming the log return of cryptocurrency price has constant mean. So, we will use the standard deviation from ARIMA model to estimate the volatility.

Future recommendations

For some time series, there existing leverage effect or residuals follow asymmetric distribution instead of Gaussian distribution, which cannot be modeled appropriately by GARCH model. Other heteroscedasticity models such as EGARCH and APAGARCH could also be tested on fitting conditional variance in such cases.

4. Applications

4.1. Dynamic Trading Strategy Using ARIMA Forecasts

Based on the forecast values and error estimates that we get from ARIMA (since predictions assume a normal distribution with mean as the baseline prediction and standard error in R reported as s.e), we device a trading strategy making following assumptions to keep the problem simplified:

a. We can only trade a single unit of a currency on any day
b. We can always buy and sell whenever we want. Essentially, we assume no restrictions on cryptocurrency
c. If we buy a unit on one day, we sell it the next day. This assumption is taken to sync with only a single-day forecast
d. If we sell a unit some day and the price goes down the next day, we realize a "virtual profit" of the amount by which the currency goes down

Though these assumptions are simplistic in nature and don't take into consideration the long-term strategies, these make our problem tractable.

Given these assumptions, our trading strategy uses the "risk-adjusted-returns R_{t+1}" defined as:

$$R_{i,t+1} = \frac{E[P_{i,t+1}] - A_{i,t}}{SE_{i,t+1}}$$

to trade. Here, for any currency *"i"*, $R_{i,t+1}$ is the estimate of risk adjusted return at time *"t+1"*, $E[P_{i,t+1}]$ is the expected value of the forecast by ARIMA model, $A_{i,t}$ is the actual value at time *"t"*, and $SE_{i,t+1}$ is the standard error of the forecast by ARIMA. **Note:** we use standard error from the ARIMA model as variance estimate because GARCH doesn't give good estimates for volatility.

With the above definition of risk adjusted returns, we devised the **trading strategy** in which we use grid search on past data to identify with the objective to maximize the profits. We fine-tune the threshold parameter "T_U" above which the currency is bought, and a threshold "T_L" below which the currency is sold. Also, let's define return for a day *"t"* be defined as "Pr_t". Our strategy can then be summarized as:

```
If Ri,t+1 >= TUi:
      Decision = Buy
      Pri,t+1 = Ai,t+1 - Ai,t
Else If Ri,t+1 <= TLi:
      Decision = Sell
      Pri,t+1 = Ai,t - Ai,t+1
Else
      Decision = Stay Put
      Pri,t+1 = 0
```

Following the above strategy, below are the cumulative profit trends for Bitcoin, DJI (trading on Saturday/Sunday not allowed) and both combined.

From figure 19, we clearly observe that overall net-profit that could have been made with this algorithm is **$10,909**.

We also observe the following trends related to profits:

- The profits, as well as losses for BTC, are of higher magnitude as compared to DJI. This is primarily because of higher variance in the BTC values, which ARIMA is not able to capture.
- The losses are higher on dates where we observe a higher MAPE for BTC.

- The profits on DJI are highly stable as the forecasts are accurate due to lesser variance.

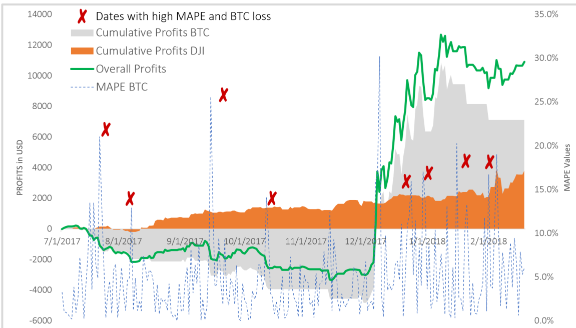

Figure 19 Cumulative Profit between BTC and DJI

Some additional pointers pertaining to the trading strategy:

- Clearly, the strategy works better and is more consistent with the currencies/stocks are that more stable.
- The initial profits of Bitcoin are significantly negative because ARIMA is not able to capture the steep increasing pattern with confidence, and mostly under-predicts.
- Net units bought for BTC are 9, while for DJI they're -15 (means the strategy leads to selling more units than buying).

4.2. Dynamic Trading Strategy Using XG-Boost Model

For this part, we use open price and close price to treat this as a classification problem on whether the close price for a day is greater than open price or not. The trading strategy is like what we saw in ARIMA (able to trade only a single unit of currency in a day) with following changes:

a. Instead of using only the close-prices, we use both open and close.
b. Instead of making the buy/sell decision one day in advance, we trade at the start of day.
c. Instead of daily continuous forecasts of prices, we forecast the probabilities of close-price being higher than open-price for a 235-day window using the earlier data – hence a better approach (helps us decide on long terms)

The features used for model are **momentum variables** which dictate how the open/close price trends have behaved in the past 2-5 days: their average (for baseline estimate) and standard deviations (to account for variability) were used as predictors, with response being a binary variable to depict whether close > open or otherwise.

Among the models we tried for our classification problem [Logistic Regression, SVM, Random Forest, XGB], since XGB gave the best results, we went ahead with that.

The probability values are used to estimate the confidence of predictions, and 10-fold cross-validation is used to fine-tune the threshold parameter for both buy/sell decision.

Let $Prob_{i,\,t}$ be the probability as predicted by model for Close > Open for currency *"i"* and time *"t"*. Let TU_Pr_i and TL_Pr_i be the threshold of currency *"i"* for buy and sell respectively. Also, Let $C_{i,t}$ and $O_{i,t}$ be the close and open price of currency *"i"* at time *"t"*. Then our logic used to trade can be summarized as below:

```
If Prob_{i,t+1} >= TU_Pr_i:
      Decision = Buy
      Pr_{i,t+1} = C_{i,t+1} - O_{i,t+1}
Else If Prob_{i,t+1} <=
TL_Pr_i:
      Decision = Sell
      Pr_{i,t+1} = C_{i,t+1} - O_{i,t+1}
Else
      Decision = Stay Put
      Pr_{i,t+1} = 0
```

In figure 20, we see a comparison for profits generated using XG-Boost and ARIMA predictions. We see that the profits made using XG-Boost are certainly better and more robust than ARIMA as they don't fluctuate too much.

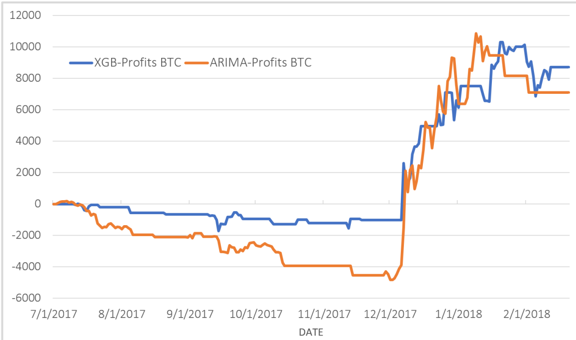

Figure 20 Profit Comparison: ARIMA & XG-Boost

Trading results:

A comparison of XG-Boost and ARIMA shows that using machine learning techniques to capture trends for prices is certainly better for currencies that are more volatile. Also, for our strategy, we have made assumptions that are naïve, and do not reflect the real-life trading scenarios. Hence, we suggest that long term forecasts would be better by accounting for appropriate trends and volatility, as they would help us to see longer in future and accordingly decide on the long/short positions.

5. Results & Discussions

5.1. Model Performance on Price Forecasting

The boxplot illustrates the forecasting performance of ARIMA, ARIMAX and VAR measured in MAPE. In terms of mean MAPE value, ARIMA performs best with the lowest of 0.049 followed by ARIMAX with 0.054. VAR does not generate satisfying results with 0.126.

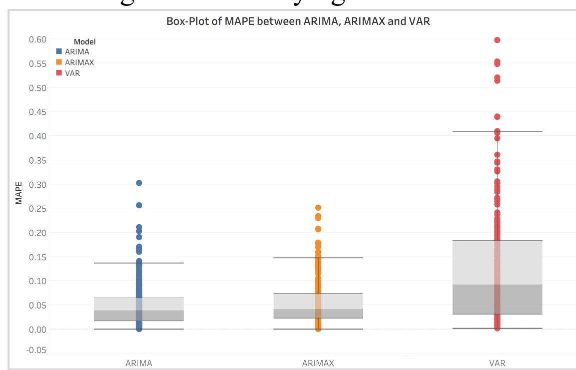

Figure 21 Boxplot of MAPE between Forecasting Model

ARIMA model is offering satisfying results for being able to capture autocorrelation information from most recent lags and shocks. The introduction of several external factors as an exogenous matrix in fitting ARIMA model does not significantly improve model performance. The team also experimented ARIMA coupled with individual exogenous factor. The result is on par with ARIMA.

The errors of VAR model could be introduced from fitting multivariate time series. The prices of other cryptocurrencies and external factors may have limited real lead-lag interactions in predicting Bitcoin price. Instead, Bitcoin may lead other cryptocurrencies for its market dominance and being a leading market indicator. Due to the large number of possible outcomes from such combinations, the changing market dynamics and our focus of the Bitcoin, the investigation of predicting other cryptocurrencies using VAR is recommended for future studies.

5.2. Model Performance on Trading Strategy

With the basic assumptions that we took for trading, we realized our forecast-models along with basic strategies performed reasonably well, both in the scenarios of highly volatile and relatively stable instruments. Given that the assumptions that we took are rather far from the real-world scenario, some assumptions that would be better to relax:

a. Allowing to trade more than one unit and put a cap on the amount we start with.
b. Use better predictors for risk and use combination of commodities to form a hedged-portfolio.
c. Use models/features that perform better for long term forecasts, to hold “Short/Long” positions for a larger duration.
d. Trading cryptocurrencies daily might not be possible due to supply-restrictions.

Incorporating these suggestions would probably enhance the performance of trading strategies and can be deployed to trade in markets.

6. Conclusion

We find that ARIMA model for cryptocurrency can provide a very satisfactory result in terms of prediction accuracy. The ARIMAX model incorporated with exogenous factors such as VIX, Gold price, Oil price, Dow index, Nasdaq index and economic policy uncertainty index does not necessarily increase the performance of prediction. In the meanwhile, by introducing exogenous factors, the complexity of model increases, which lead the model to capture additional noise instead of true variations.

Similarly, though VAR model could capture some statistically significant lead-lag effects within cryptocurrency and between cryptocurrency with exogenous factors; such lead-lag effects are short-term effects as the selected orders are like 1 or 2. Also, the MAPE of VAR model is much higher than ARIMA model. Thus, for the reason of simplicity and accuracy, we stayed with ARIMA model.

As for the heteroscedasticity, after testing Ljung-Box test, analyzing the ACF plot of squared residual, fitting heteroscedasticity models, we believe that the log-return of cryptocurrencies are very likely to have a constant variance. Thus, instead of modeling conditional variance, we choose to use standard deviation from ARIMA model as predicted volatility.

In the end, based on our time series analysis on cryptocurrency, we come up ways to devise trading strategies using ARIMA forecast and XG-Boost. Although both of applications are based on naïve assumptions, such application could be considered as prototype and can be developed in the future.

Reference

[1] Narayanan, A., Bonneau, J., Felten, E., Miller, A. and Goldfeder, S. *Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction*. Princeton University Press, 2016.
[2] Swan, M. *Blockchain: Blueprint for a New Economy*. O'Reilly Media, Inc., 2015.
[3] McNamara, G. *Making sense of bitcoin, cryptocurrency and blockchain*. PricewaterhouseCoopers, City, 2018.
[4] Hayes, A. S. Cryptocurrency value formation: An empirical study leading to a cost of production model for valuing bitcoin. *Telematics and Informatics*, 34, 7 (2017/11/01/ 2017), 1308-1321.
[5] Yermack, D. *Chapter 2 - Is Bitcoin a Real Currency? An Economic Appraisal A2 - Chuen, David Lee Kuo*. Academic Press, City, 2015.
[6] Andrews, A. *Bitcoin: The Complete Guide to Understanding Bitcoin Cryptocurrency and Bitcoin Mining*. CreateSpace Independent Publishing Platform, 2018.
[7] Kai, J. *Bitcoin and Central Banks: a Monetary Revolution*. Paris Innovation Review, City, 2016.
[8] Jang, H. and Lee, J. An Empirical Study on Modeling and Prediction of Bitcoin Prices With Bayesian Neural Networks Based on Blockchain Information. *IEEE Access*, 6 (2018), 5427-5437.
[9] Dyhrberg, A. H. Bitcoin, gold and the dollar – A GARCH volatility analysis. *Finance Research Letters*, 16 (2016/02/01/ 2016), 85-92.
[10] Katsiampa, P. Volatility estimation for Bitcoin: A comparison of GARCH models. *Economics Letters*, 158, C (2017), 3-6.
[11] Baker, S. R., Bloom, N. and Davis, S. J. Measuring Economic Policy Uncertainty*. *The Quarterly Journal of Economics*, 131, 4 (2016), 1593-1636.
[12] Whaley, R. *Understanding VIX*, 2008.
[13] Diebold, F. X., Husted, S. and Rush, M. Real Exchange Rates under the Gold Standard. *Journal of Political Economy*, 99, 6 (1991), 1252-1271.
[14] Duarte, F., Tenreiro Machado, J. and Monteiro Duarte, G. *Dynamics of the Dow Jones and the NASDAQ stock indexes*, 2010.
[15] Chu, J., Chan, S., Nadarajah, S. and Osterrieder, J. *GARCH Modelling of Cryptocurrencies*, 2017.

Appendix

GitHub Repo: https://github.com/mxu007/ISYE6402-Project-Cryptocurrency

mxu007 clean files		Latest commit 1c78e5c 5 minutes ago
data	clean files	5 minutes ago
graphs	clean files	5 minutes ago
literature review	clean files	5 minutes ago
README.md	Update README.md	17 days ago
arima_code.R	final arima code with net profits and all files required for plots	3 days ago
arimax_var.R	clean files	5 minutes ago
data_prep.R	arima and data-prep codes	16 days ago
garch.R	no change	13 days ago

README.md

ISYE6402-Project-Cryptocurrency

ISYE6402 Time Series Analysis Project -- Cryptocurrency Pricing Forecast

Team Member : Nirmit Chetwani, Tianyi Liu, Minghan Xu