Cryptocurrency Safety Optimization Using Statistical Neural Network

Zhiyu Lin, Isabelle Kemajou-Brown, Evelyn Sander

zl281@georgetown.edu, elisabeth.brown@morgan.edu, esander@gmu.edu

INTRODUCTION

Different research papers have come to opposite conclusions about whether cryptocurrencies can act as financial safe havens. Some argue that cryptocurrencies such as Bitcoin have safe-haven qualities due to their decentralized and scarce nature, especially during recessions. [1, 3, 6] While others find that even having a small portfolio allocation to Bitcoin increase downside risks [2, 4, 5, 7]. Although both sides have merit, most statistical/machine learning research on this topic are limited to examining one or two popular cryptocurrencies.

This research evaluates these arguments and finds that eight types of popular cryptocurrencies show a jump in correlation to gold prices during the COVID-19 economic recession, in support of the safe haven argument. It then utilizes statistical and machine learning feature selection frameworks to maximize a cryptocurrency portfolio's safety through three steps: static feature selection with statistical models, dynamic selection with recurrent neural networks, and dynamic selection with uncertainty measures, to answer these questions: "which cryptocurrencies are safer than others", "what combination of cryptocurrencies should a trader maintain on a daily basis", and "what combination should a trader maintain daily and how confident they should be".

In this last portion, we apply a robust rank-based Double-Generalized Linear Model (rrDGLM) [8] as the network minimizer. This is especially helpful for using neural networks to model cryptocurrency prices because rrDGLM updates the mean and the variance recursively to control for the diverging number of parameters in the network. This technique can become useful for monitoring AI safety and providing network uncertainty measure through statistical methods.

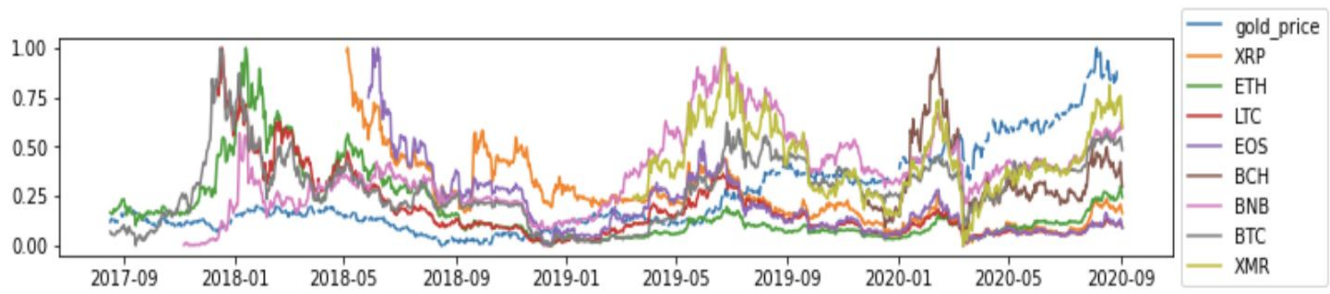
DATA

Data is obtained from Binance, a cryptocurrency exchange company with a Python facing API. Prices and Volume for eight type of cryptocurrencies are collected: Ripple (XRP), Ethereum (ETH), LTC, Ethereum on Steroids (EOS), Bitcoin Cash (BCH), Binance Coin (BNB), Bitcoin (BTC), Monero (XMR) from August of 2017 till September of 2020 on a daily basis. We also obtained daily gold prices for the same time period from the Federal Reserve Bank of St. Louis (FRED Economics Research) and removed days with missing values.

This dataset is first imputed and smoothed before model fitting. Imputation is done for each individual currency using linear interpolation, and smoothing is done with a 14-day rolling average. After these steps, the data is then reshaped with 30-day overlaps for dynamic feature selection. Prior to model fitting, It is also reshaped to three dimensions each representing the number of time stamps, the number of features or cryptocurrency assets, and the number of samples for neural network models.

PART I -- SAFE HAVEN

According to Investopedia, a safe haven is an asset that retains or gains in value during financial turbulences. Safe haven assets are expected to be less correlated of negatively correlated to the market trend. A few examples that can be utilized to benchmark asset safety include: gold, treasury bills, U.S. Dollars and defensive stocks. Baur and Hoang argue that Bitcoin is designed to closely resemble gold, thus exhibiting gold-like features such as safe haven. [6] Several other cryptocurrencies such as Tether and Ethereum are also designed to be backed by gold. Due to these close relationships, we decide to utilize gold as the safe haven measure in this study.



gold_price	2067.15
XRP	0.9028
ETH	1388.02
LTC	352
EOS	14.7207
BCH	493.77
BNB	38.6766
BTC	19102.7
XMR	117.86

Figure 1. Cryptocurrency prices from 2017 to 2020 after max-min scaling. Table on the right: original scale for each currency. Unit: U.S. Dollars

From the scale table, we see that out of the eight types of cryptocurrencies, Bitcoin and Ethereum have much higher values than the others. After normalization, we observe a similar trend between these prices, especially starting from early 2019.

From the plot, we observe a similar trend between these currencies, especially during the peak of 2019. From 2020 January onwards, we see that these cryptocurrencies exhibit a similar pattern compared to gold (blue line).

To further examine the relationship between cryptocurrencies and gold, especially the changes that happen after COVID-19, we compare the corrrelation of each cryptocurrency's with respect to gold price.

We observe that four out of the eight cryptocurrency prices are negatively correlated to gold price before COVID-19. The other four currencies have weak correlation (max ~0.5) with gold before the economic recession posed by the pandemic. However, we see that all eight types of currency exhibit positive correlation with gold price during the COVID-19 recession, which provides evidence in support of the argument that cryptocurrencies can act as a safe haven during recession.

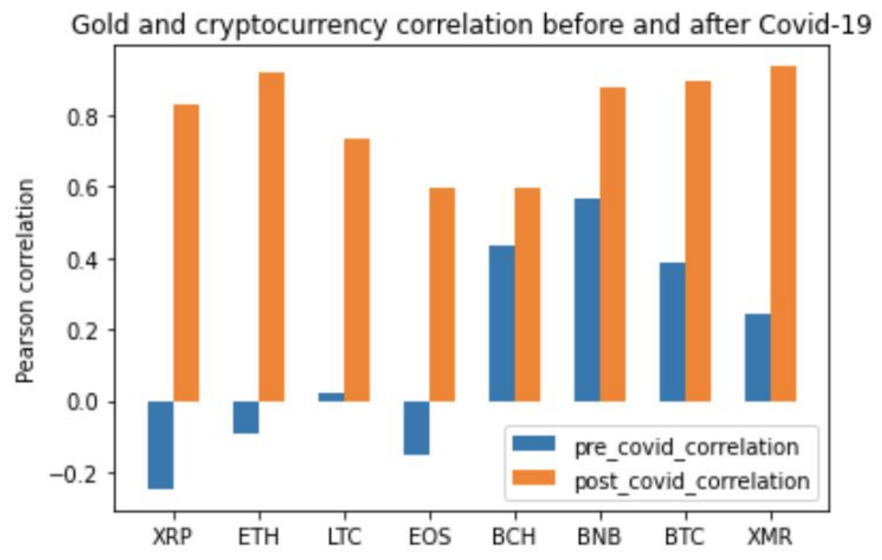


Figure 2. Gold and Cryptocurrency Correlations before and after COVID-19 Recession

PART II -- PORTFOLIO OPTIMIZATION

STATIC SELECTION

Applying a few classical feature Selection methods, we observe that some cryptocurrencies such as XRP BNB and BTC are selected by multiple methods which means that these two currencies are important predictors for gold price.

We see that the methods in the Lasso family have similar opinions, so if we were to rank these features by votes, we would give high ranks to XRP, followed by BNB, BTC, Decision Tree, ETH and the rest over a static period of time. However, having a ranking for each cryptocurrency is too general and in the next section, we will utilize recurrent neural networks to get a daily portfolio importance combinations.

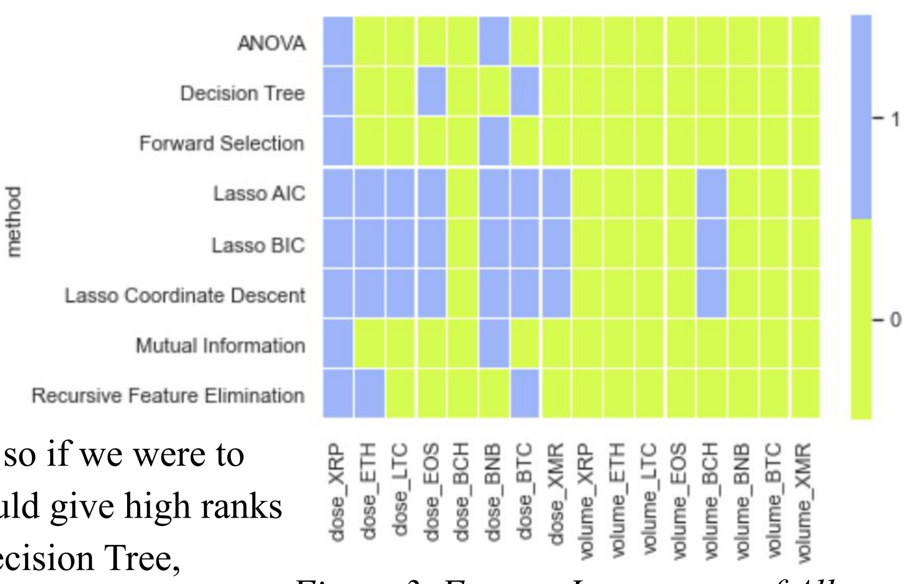


Figure 3. Feature Importance of All Times By Classical Selection Methods

DYNAMIC SELECTION

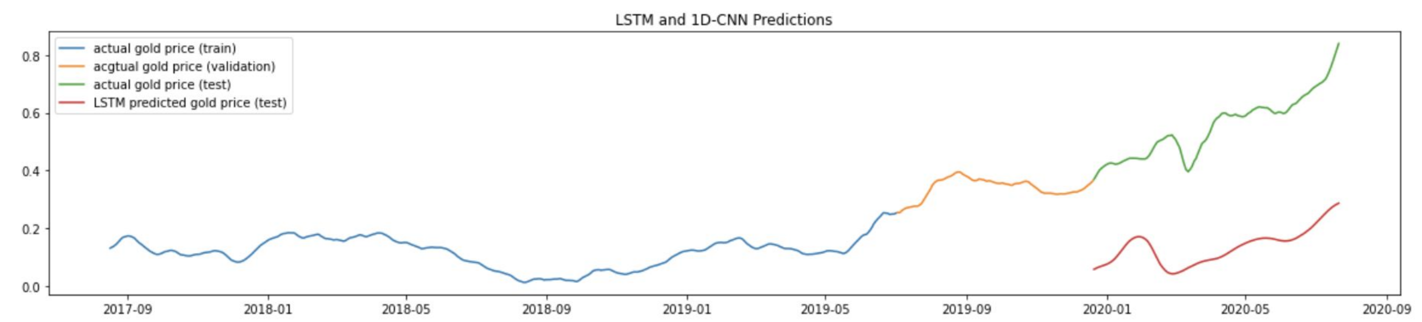


Figure 4. LSTM predictions

From the plots above we see that LSTM can closely follow the trend of the gold price, even though the intercept is off. This does not matter for our purpose because in order to come up with a portfolio that resembles gold prices, we only care about the proportion between each type of cryptocurrency.

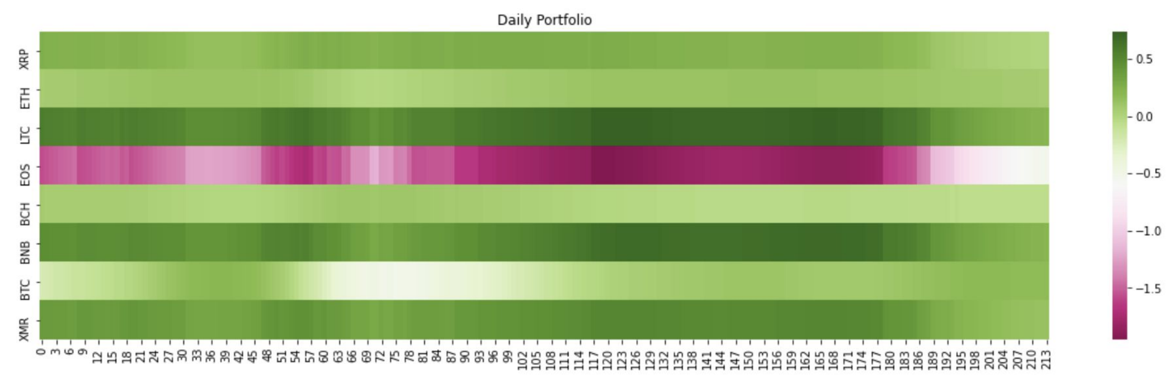


Figure 5. Cryptocurrency daily portfolio for 30 days after the training period

By reshaping the weights, we have a daily cryptocurrency portfolio for how much each cryptocurrency should be held to most similarly have the portfolio resemble gold prices. We see that EOH is considered to be an important feature most of the days during the testing period, but it contribute negatively to the portfolio. On the other hand, BNB, XMR and LTC are important features and contribute positively to the daily portfolio. Moreover, we also see that except for occasional white spaces (gradient close to zero) where the feature importance reduces, most currencies maintain a rather consistent rank across a period of time.

DYNAMIC SELECTION WITH UNCERTAINTY MEASURE

DGLM

A Generalized Linear Model (GLM) measures linear regression models covariates against the response variable via a link function. Thus, it explains the feature importances through the magnitude of the covariates. However, a GLM makes the assumption that the dispersion of the covariates are constant. A DGLM breaks this assumption by monitoring the mean and the dispersion parameters at the same time via two separate link functions.

Nguelifack and Kemajou-Brown developed a feature selection method using rank-based DGLM with adaptive Lasso. In particular, this method achieves good performance when the data has diverging number of parameters. Below is a summary of the key definitions for rrDGLM.

$$\|\mathbf{v}\|_{\varphi} = \sum_{\mathbf{i}=1}^{\mathbf{n}} \mathbf{a}[\mathbf{R}(\mathbf{v_i})]\mathbf{v_i},$$
The rank pseudo norm for against high leverage outliers

$$Q_n(\boldsymbol{\theta}, \boldsymbol{\beta}) = \frac{1}{n}\sum_{i}^{n} w(\mathbf{x}_i)\left[\frac{R(e_i)}{n+1} - \frac{1}{2}\right] e_i,$$
Rank-based Wilcoxon dispersion function

$$e_i = \frac{\mathbf{y}_i - \mu_i}{\sqrt{\phi_i V(\mu_i)}},$$
The Pearson residual

$$D_n(\boldsymbol{\theta}, \boldsymbol{\beta}) = Q_n(\boldsymbol{\theta}, \boldsymbol{\beta}) - n\sum_{j=1}^{p_n} p_{\lambda_j^{(1)}}(|\theta_j|) - n\sum_{k=1}^{q_n} p_{\lambda_k^{(2)}}(|\beta_k|),$$
The objective function with L1 & L2 penalty

Neural Network with Uncertainty Measures

In recent years, lots of research focus has been devoted to neural network uncertainty estimation and AI safety. [10, 11, 12] The idea is to understand not only what the prediction is, but also how certain the model is about the prediction to avoid the situation where the model is over confident and lead to harmful decisions. One explanation, as Detlefsen et al point out, is that predicting the variance is a different task than predicting the mean. Some of the common methods that try to measure NN uncertainty include: MC Dropout, uncertainty-aware NN, Bayesian NN, Gaussian Processes, ensemble methods, etc.

In this study, we apply the rank-based DGLM objective function to the LSTM network to update the mean and the variance using Majorize-Minorize iterative updates.

Neural Network with DGLM Loss with Mean-Variance Iterative Update

Pseudo-code:

Initialize *Residual* with random numbers

While min_tolerance < tol **or** max_iteration > iter:

Variance = LinkV(*X*, *Residual*)

Mean = LinkM(*X*, *y*, *Variance*)

Residual = (*y* - LinkM(*X*, *Mean*))^2

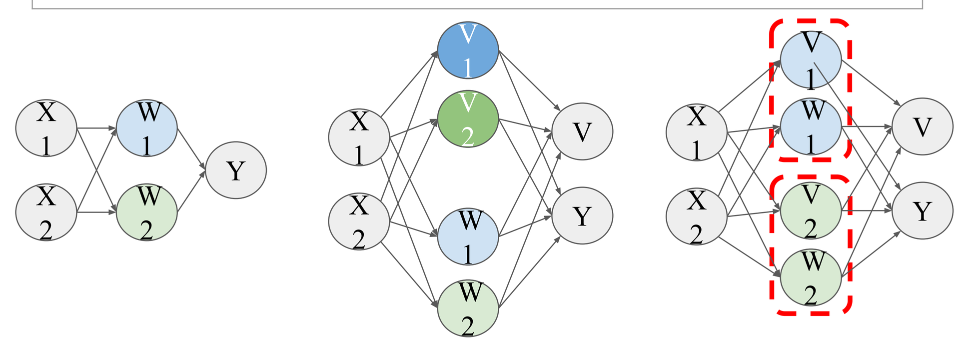


Figure 6. Illustration of DGLM mean-variance estimation. Left: regular network, middle: mean-variance separate estimation, right: DGLM mean-variance network.

Kendall et al and Loquercio et al proposed something similar where a variance variable is predicted during training. However, these methods generally estimate variance separate from the mean. Our approach updates the mean and the variance iteratively, creating a dependent loss function in each layer update. Another advantage is that our method inherently control for high leverage points and diverging number of parameters in a NN by using rank pseudo-mean in the loss function.

Figure 6 demonstrates the difference between a regular network, a mean-variance estimation network, and our proposal.

CONCLUSION

This study investigated the claim that cryptocurrencies can act as financial safe havens. We observe a jump in correlation between eight types of popular cryptocurrencies and gold price during COVID-19 recession. Taking gold price as the measurement for safe haven, we then applied statistical and machine learning methods comparing the asset importances with respect gold. Network gradients are utilized to compute the daily portfolio of cryptocurrencies to optimize their safe haven characteristic. Finally, a neural network uncertainty estimation methodology is developed by applying Nguelifack and Brown's rrDGLM feature selection methodology. Future effort should be put into testing this methodology on different datasets.

ACKNOWLEDGEMENTS

Special thanks to Dr. Isabelle Kemajou-Brown (Morgan State University), Dr. Evelyn Sander (George Mason University) for their consistent guidance and resource inputs. Special thanks to Dr. Brice Nguelifack (US Navy) for developing the DGLM network uncertain estimation. Special thanks to Dr. Nathaniel Strawn (Georgetown University) for initial brainstorms and consistent encouragement.

REFERENCES

1. Chapman, Kyle R. *How Will Cryptocurrencies Fare in a Recession?*, Barrons, 11 Mar. 2019, www.barrons.com/articles/cryptocurrencies-recession-bitcoin-ethereum-ripple-51552071925.
2. Conlon, Thomas, and Richard McGee. "Safe Haven or Risky Hazard? Bitcoin during the Covid-19 Bear Market." *SSRN Electronic Journal*, 2020, doi:10.2139/ssrn.3560361.
3. Conlon, Thomas, et al. "Are Cryptocurrencies a Safe Haven for Equity Markets? An International Perspective from the COVID-19 Pandemic." *SSRN Electronic Journal*, 2020, doi:10.2139/ssrn.3601045.
4. Dwita Mariana, Christy, et al. "Are Bitcoin and Ethereum Safe-Havens for Stocks during the COVID-19 Pandemic?" *Finance Research Letters*, vol. 38, 2021, p. 101798., doi:10.1016/j.frl.2020.101798.
5. Maniff, Jesse Leigh, et al. "Safe-Haven Performance in the Age of Bitcoin." *Federal Reserve Bank of Kansas City*, 2020, www.kansascityfed.org/publications/research/eb/articles/2020/safe-haven-performance-age-of-bitcoin.

6. Moore, Simon. "The Risk Of Treating Bitcoin As A Safe Haven Asset." *Forbes*, Forbes Magazine, 22 Jan. 2021, www.forbes.com/sites/simonmoore/2021/01/21/the-risk-of-treating-bitcoin-as-a-safe-haven-asset/?sh=1ff0ce932db5.
7. Baur, Dirk & Hoang, Lai. (2019). A Crypto Safe Haven Against Bitcoin. SSRN Electronic Journal. 10.2139/ssrn.3457688. https://www.researchgate.net/publication/336204433_A_Crypto_Safe_Haven_Against_Bitcoin
8. Smales, L.A.. (2018). Bitcoin as a safe haven: Is it even worth considering?. Finance Research Letters. 30. 10.1016/j.frl.2018.11.002. https://www.researchgate.net/publication/328977093_Bitcoin_as_a_safe_haven_Is_it_even_worth_considering

9. Nguelifack, Brice M., Kemajou-Brown, Isabelle (2019) Robust rank-based variable selection in double generalized linear models with diverging number of parameters under adaptive Lasso, Journal of Statistical Computation and Simulation, 89:11, 2051-2072, DOI: 10.1080/00949655.2019.1607346, https://www.tandfonline.com/doi/abs/10.1080/00949655.2019.1607346
10. "Gold Fixing Price 10:30 A.M. (London Time) in London Bullion Market, Based in U.S. Dollars." *FRED*, 13 Feb. 2021, fred.stlouisfed.org/series/GOLDAMGBD228NLBM.
11. Detlefsen, N. et al. "Reliable training and estimation of variance networks." *NeurIPS* (2019).
12. Michelmore, Rhiannon et al. "Evaluating Uncertainty Quantification in End-to-End Autonomous Driving Control." *ArXiv* abs/1811.06817 (2018): n. Pag.
13. Loquercio, Antonio et al. "A General Framework for Uncertainty Estimation in Deep Learning." *IEEE Robotics and Automation Letters* 5 (2020): 3153-3160.