Adding Momentum Factors to Predict Price Change: A New Cryptocurrency Ranking Methodology

Matthew Baldree, Paul Widhalm, Brandon Hill,

Robert Viglione, Matteo Ortisi

{mbaldree, pwidhalm, bdhill}@smu.edu,

[rob@zensystem.io](mailto:rob@zensystem.io), matteo.ortisi@gmail.com

**Abstract.** The cryptocurrency market is in its infancy and highly volatile. Existing ranking sites such as [www.coinmarketcap.com](http://www.coinmarketcap.com) and [www.coincap.io](http://www.coincap.io) are insufficient for traders because they focus only on current statistics and do not provide ranking by predictive statistics. A new beta website, [www.coingecko.com](http://www.coingecko.com), adds additional factors such as developer, community, and public interest to create a new composite ranking. This ranking approach is insufficient for traders because it too does not provide ranking by predictive statistics. We propose a ranking by predictive future price statistics so investors can focus on batch investment opportunities. The ranking is created from predictive price change by an ARIMA model trained on data that includes traditional market capitalization plus new momentum factors such as period moving averages and Commodity Selection Index. The ARIMA model is developed by visualizing the time series data, stationarizing it, and plotting for optimal parameters to train it to predict future pricing. An API is also created to provide ranking of cryptocurrency based on new predictive statistics based on market momentum.

# 1 Introduction

The cryptocurrency market started in 2009 with the bitcoin network and in 2010, the first bitcoin exchange opened. As of September 7, 2017, there are 5,475 cryptocurrency exchanges according to Coin Market Cap with a total market capitalization of $164 billion for 867 currencies for a 2017 growth rate of 12-fold. By comparison, this market capitalization represents 20% of Apple’s market cap.

The market is growing exponentially. For instance, the number two cryptocurrency in market capitalization, Ethereum, grew 4,100% in eight months in 2017. The Standard and Poor’s 500 Index which is made up of 500 of the most widely traded US stocks took over 40 years to achieve the same kind of growth. The cryptocurrency market is currently in its infancy and to enable it to grow into maturity will require solid tools by which investors can rely upon.

With so much growth in an industry, many people want to get involved in this emerging market. But, the challenges for investors is navigating this young, volatile new market with limited tools for researching, trading, and transacting. Investing can be daunting for new and current investors. Only recently has mainstream financial institutions like Fidelity [1] begun to give its customers the ability to add cryptocurrencies to their portfolios. Besides continual development of the cryptocurrency products, additional marketplace tools need to be developed to support this growing marketplace.

A good place for new investors to learn about cryptocurrencies is to visit [www.coinmarketcap.com](http://www.coinmarketcap.com) website to see a ranking of cryptocurrencies by market capitalization. Market capitalization is the price of the currency times the number of currencies in circulation. This metric gives an investor a relative size of the market, see Table 1 for an example ranking of cryptocurrencies by market capitalization.

**Table 1.** Coin Market Cap top five ranking as of September 16, 2017.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rank | Name | Market Cap (million) | Price per Coin | Max Supply (million) |
| 1 | Bitcoin | $59,879 | $3,613.67 | 21,000 |
| 2 | Ethereum | $23,309 | $246.24 | - |
| 3 | Bitcoin Cash | $7,397 | $445.98 | 21,000 |
| 4 | Ripple | $6,856 | $0.18 | 100,000,000 |
| 5 | Litecoin | $2,5448 | $48.12 | 84,000 |

Investors researching currencies can only sort by backwards looking statistics. If you want to research future trends and momentum, you have to utilize a trading tool to analyze one currency at a time. In addition, traders will follow currencies on Twitter and Reddit to gauge user and developer sentiment adding this knowledge to collectively try and predict future pricing. This labor-intensive research approach makes it difficult to survey more than a handful of currencies for buy and sell opportunities.

A beta website at [www.coingecko.com](http://www.coingecko.com) incorporates additional factors to market capitalization such as liquidity, developer, community, and public interest to determine ranking. By incorporating non-financial factors and applying a custom algorithm, a different ranking results for rank three through five compared to straight market capitalization, see Table 2. But this Total Score ranking is not clear on what it means for future results. Does a high score mean the price will go up or down in the next 30 days?

**Table 2.** Coin Gecko top five ranking as of September 16, 2017.

|  |  |  |  |
| --- | --- | --- | --- |
| Rank | Name | Total Score | Max Supply (million) |
| 1 | Bitcoin | 91% | 21,000 |
| 2 | Ethereum | 83% | - |
| 3 | Litecoin | 77% | 84,000 |
| 4 | Monero | 70% | - |
| 5 | Ripple | 70% | 100,000,000 |

Current ranking of cryptocurrencies in the marketplace do not address the needs of traders who are researching buy and sell opportunities. By adding momentum factors, a time series ARIMA model can be fitted to predict future price changes. The trader may then sort a list of cryptocurrencies by relative price change for different time periods to view aggregate investment opportunities.

The approach to solving the problem of providing a new ranking system based on momentum factors is accomplished by fitting an ARIMA (autoregressive integrated moving average) model to the time series data of past performance. Using a time series to model trading data is common and analysts regularly use this approach to discover trends and other patterns.

Past performance data plus new data obtained from sites such as Coin Market Cap and Coin Cap, plus social media data from Twitter are used to calculate momentum factors. These new factors are then analyzed to determine their predictive power. Low predictive factors are removed.

The scope of the problem is first limited to a short time period with only one currency for analysis and model development. After progress is made, additional time periods are added. As outcomes improve, other currencies are added until all are incorporated. This quantitative forecasting considers historical data of a variable to forecast future values of the variable. In this case we are forecasting price variability over a period of time.

The prediction model is trained on training data and tested against held out test data to measure the mean squared errors (MSE). Adjustments to the algorithm are then made to lower the MSE until an optimal model is obtained. A ranking API is developed to provide a list of cryptocurrencies by change in momentum. The prediction model requires retraining on a regular cycle as new data is acquired and calculated.

Data science tools like Python, R, database, and machine learning toolkits are used to understand the data, developer a predictive model, and ranking API. Ranking results are then analyzed by industry veterans for value. Feedback from the industry veterans will allow fine tuning of the ranking algorithm.

## 1.1 Cryptocurrency

{“Seems like you need a section to describe the cryptocurrencies in more detail.  Maybe not all the technical details of how they work, although that is definitely needed at at least a high level, but definitely how they are stored and spent.  This would include all the vulnerabilities and risks of crypto currency.  Also, need to talk about the float (how many are available to trade) since that has a significant impact on price.  This is to say that these are not always really liquid markets.”}

# 3 Background

This section will be used to document scholarly research on the following items:

* equity price prediction,
* momentum factors predicting price movement,
* fitting ARIMA model to time series data.

# 5 Data

# 7 Results

Notes:

* Exploratory data analysis
* Algorithm development
* Application of algorithm
* Feedback from community if appropriate

# 6 Analysis

Notes:

* Analysis of results

# 8 Ethics

{What needs to go in the ethics statement?

Whoever your funder will be, it’s a good idea for your ethics statement to address the [six key principles](http://www.ethicsguidebook.ac.uk/Key-ethics-principles-15) set out in the ESRC Framework for Research Ethics. So you need to be able to explain how:

* you are ensuring quality and integrity of your research;
* you will seek informed consent;
* you will respect the confidentiality and anonymity of your research respondents;
* you will ensure that your participants will participate in your study voluntarily;
* you will avoid harm to your participants; and
* you can show that your research is independent and impartial.

<http://www.ethicsguidebook.ac.uk/what-needs-to-go-in-the-ethics-statement-20>

}

# 9 Conclusions

# 10 Future Work

Notes:

* Use new ranking methodology to determine cryptocurrencies that should be considered for various indices.

# References

1. Fidelity Labs Tests Digital Asset Wallet On Fidelity.com. August, 09, 2017. <https://www.fidelity.com/about-fidelity/corporate/fidelity-labs-tests-digital-asset-wallet-on-fidelity.com>

2. McNaly, Sean: Predicting the price of Bitcoin using Machine Learning. Naional College of Ireland (2016)

3. Soulas, Eleftherios, Shasha, Dennis: Online Machine Learning Algorithms For Currency Exchange Prediction: NYU CS Technical Report TR-2013-953 (2013)

# Appendix: Plan of Milestones

|  |  |  |
| --- | --- | --- |
| Date | Event | Goal |
| ~~31-Aug~~ | ~~Webinar~~ |  |
| ~~15-Sep~~ | ~~1630 Call with Rob~~ | ~~Initial meeting with 1st sponsor.~~ |
| ~~17-Sep~~ | ~~1900 CST group call~~ |  |
| 18-Sep | 1900 CST group call |  |
| 19-Sep | 1630 Call with Matteo | Initial meeting with 2nd sponsor. |
| *20-Sep* | *Beta Draft due\** |  |
| 24-Sep | 1900 CST group call |  |
| 26-Sep | Café Talk 1 | Receive feedback from others in the |
| 27-Sep | Café Talk 1 | class regarding the beta draft. *Must attend at* |
| 28-Sep | Café Talk 1 | *least one of the meetings* |
| Early Oct | Meet with advisors |  |
| 1-Oct | 1900 CST group call |  |
| 8-Oct | 1900 CST group call |  |
| Late Oct | Meet with advisors |  |
| 15-Oct | 1900 CST group call | Combine notes on 1st Draft |
| 22-Oct | 1900 CST call | Compile the second draft ==> submit |
| *25-Oct* | *1st Paper Draft Due* |  |
| 29-Oct | 1900 CST group call |  |
| 31-Oct | Café Talk 2 | Receive feedback from others in the |
| 1-Nov | Café Talk 3 | class regarding the beta draft. *Must attend at* |
| 2-Nov | Café Talk 4 | *least one of the meetings* |
| Early Nov | Meet with advisors |  |
| 5-Nov | 1900 CST group call |  |
| 21-Nov | 1900 CST group call |  |
| Late Nov | Meet with advisors |  |
| 19-Nov | 1900 CST group call |  |
| 26-Nov | 1900 CST group call |  |
| Early Dec | Meet with advisors |  |
| 2-Dec | 1900 CST group call | Combine notes on 2nd Draft |
| 10-Dec | 1900 CST group call | Compile the second draft ==> submit |
| *15-Dec* | *Second Paper Draft Due* | *(points count towards Capstone B [5%]) 15-20 pages* |

# History of Feedback

### September 12, 2017

Comments on your proposal - first a formatting comment: please follow the format exactly.  Do not change it.  For example, there is no blank line between paragraphs.

Question on your proposal: What is the problem you are solving? In addition, how do you measure success? Why is it not trivial to create an index? Specifically, how would it be created?  How would it be used?

In writing, do not state your opinions…eg do not use “Our belief…” Write neutral.  This is not an opinion piece.

The concept is good, but it’s not clear why this is a capstone project.  Indexes are created all the time. Why is it hard? How is this more than a simple exercise? How do you assess success?   You are missing all of the problem details and measures of success.

Please add the problem and additional details.

### September 12, 2017

Better. As papers go, there’s much to be changed (some of that in my comments below). But, the big questions I still have are “how is this different than what’s already been done? And, why is this not a simple project?” Please give me a short answer to these two questions…and you need to make the answers clear in your writeup too.

Quick comments in random order (like I’m reading the paper) -

remember to submit this (with any revisions) to 2DS

author list should be either students first, advisors last (this should reflect contribution order) or simply alphabetical by last name.

Introduction section typically is 2-4 pages in this format and does not contain figures or tables.

The introduction section is an executive summary of the work.

The abstract is a one paragraph executive summary of the work.

your future work section should be “Conclusions and Future Work” since you are missing a conclusions section.

You need results before analysis. If you don’t have any results, you don’t have anything to analyze.

Your exploratory data analysis is really Results, not Analysis ——

NOTE: terminology - Your “Analysis” section is an analysis of your results. Your “Results” section contains results that are generated by analyzing your data/systems/models/etc. Don’t be confused by the overuse of the word “analysis”

ethics section - let me point out that you have TONS of opinions in this section…in fact, it’s all opinion. When you make statements about “upside” and “downside” these are opinions. If you qualify these statements, to specify what you mean by “upside” and “downside” (note, completely qualify, not just give examples), then these are simply words. Without qualifications they are interpreted by the reader however the reader wishes to interpret them. And, what’s upside to one reader may be a downside to another. You’ve also got very broad statements eg “numerous vulnerabilities and complex issues that must be addressed…” All statements must be supported by facts provided in your paper. If you wish to say that there are “vulnerabilities” then you need to actually define the vulnerabilities (at least the major ones). If you wish to say that there are complex issues, then you must define these complex issues. Without these definitions and argued statements and analysis surrounding them, you cannot draw a conclusion that they “must be addressed before…” Without supporting evidence, you just have an unsupported opinion piece. We do not write opinion pieces.

- Kinda of an odd place to put the “questions you will address in your research” list.

ethics discussion should touch on a broad range of items including the ethics of the questions you ask.

ethics section should come just before the conclusions.

Seems like you need a section to describe the cryptocurrencies in more detail. Maybe not all the technical details of how they work, although that is definitely needed at at least a high level, but definitely how they are stored and spent. This would include all the vulnerabilities and risks of crypto currency. Also, need to talk about the float (how many are available to trade) since that has a significant impact on price. This is to say that these are not always really liquid markets.

Need a section just on the data.

Your problem seems simplistic and not quite hard enough for a Capstone. Why is what you’re proposing to do worthy of a Capstone project? Why is the solution not a simple application of an existing model (or tweak of an existing model)?

Need to talk about how traders use the rankings to trade. How do investors use this information?

Why is what you’re doing not just more of the same as what’s already been done?

Cheers!

Daniel