What is the impact of social medias and technology updates on the prices of (some) crypto-currencies?

Damien Baldy
17-620-626

I. Foreword

As crypto-currencies are multiplying and growing in market capitalization, in the wake of Bitcoin, it is gaining traction mainly thanks to the public's attention and new investors who don't want to miss the investment that already built colossal fortunes. As an example, the creator and founder of Bitcoin, an unknown personality or group, has a fortune estimated at 980 000 Bitcoins[i], which represents (as of 04-01-2018) 14.7 billion dollars. The exponential increase has sparked many questions about the true value of crypto-currencies, the optimists claiming they represent a revolution and they will change the way we spend money and create contracts, while several bank officials[ii] called the crypto-mania a bubble. In 2017, Bitcoin went from 900$ to 15,000$ - peaking at almost 20,000$ - and young crypto-currencies like NEO had even sharper increase, rising from 0.15$ on the 1st of January 2017 to 74.54$ a year later.

Crypto-currencies are based on a technology called the blockchain, created by Satoshi Nakamoto and explained in his white paper for the Bitcoin, whose main utility is to solve the problem of trust thanks to a decentralized system secured using cryptography. Dubbed a disruptive technology, the blockchain bears the promise of extremely interesting applications, such as a payment system, a digital currency, crowdsales, smart contracts… However, due to the extreme easiness of creating a new cryptocurrency (the original code of Bitcoin is open source), the technology has attracted many entrepreneurs that launch ambitious projects based on the blockchain, most of the time emitting and selling new tokens. The money raised is used as an incentive for the development team that will work to increase the value of the tokens they have created. Every project is launched with a white-paper that describes the problem and the solution proposed by the development team, as well as a roadmap that they ambition to follow.

This complicated environment with rocket-science technologies and increasing applications trades huge amounts of money every day, with numerous speculators that day-trade on "hot" crypto-currencies, expecting incredible gains in a short time. Some crypto-currencies are deemed by the community as scams and yet increase their value in USD over-time, such as BitConnect[iii], that has been exposed as a Ponzi and yet keeps increasing in market capitalization.

II. Abstract

The problem of pricing the crypto-currencies is probably one of the hottest topics of the moments. Big merchant banks such as Goldman Sachs or Morgan Stanley have analysts dedicated to this problem and the public attacks on Bitcoin are almost monthly. The crypto-optimists claim the value of crypto-currencies is rooted in their inner-value, their virtue and the fact that they build strong communities of optimist buyers, dedicated to the projects and following closely the advancement.

Since the Ethereum blockchain has appeared, the paradigm has slightly changed in the appearance of new crypto-currencies. Most new currencies are traded on the Ethereum blockchain, which saves hashing power and is more cost efficient. However, the newly emitted crypto-currencies require a strong community and investors are highly regarding on the dedication and the seriousness of the development team. If the trust is too low, the crypto-currency never takes off and disappears in the cemetery of crypto-currencies.

The crypto-sphere has a strong assumption that Bitcoin's value is more based on its technological innovation but not so much on community building. In other words, there was no marketing plan to attract investors to Bitcoin and it increased in value due to the word of mouth and a growing number of individuals considering the technology as valuable. On the contrary, many crypto-currencies are now issued by companies that have a clear business purpose and work on building a community and a service around the blockchain.

How does one connects the dots in such a market? Should we consider the crypto-market as purely speculative, or does it increases in value by fulfilling its promises? Are technological improvements and community building factors of value creation, or does the market pay no attention on this to price a crypto-currency?

This paper will try to answer these questions econometrically on a selected set of crypto-currencies that I will present in a few words.

III. Choice in crypto-currencies and scrapper

Since crypto-currencies are extremely volatile, it is quite a challenge to understand the underlying logic when the market prices upward or downward. Bitcoin constitutes the gold standard of cryptos and most crypto-currencies are traded only against Bitcoin or Ethereum, but never directly against dollars or euros. However, since the value in Bitcoin is highly fluctuating, we will pay attention only on the price in dollars displayed by the website coinmarketcap.com. The choice not to work in Bitcoin is due to the idea, still dominant, that US dollar is a neutral instrument for measuring value, while Bitcoin remains a speculative instrument.

We will focus on the following 10 currencies:

- Bitcoin: the oldest and the most famous crypto-currency. As the first crypto, it is currently the most expensive and therefore the currency that draws the most attention.
- Dash: a crypto-currency backed by a company that uses an algorithm to make the user and the transaction completely anonymous
- Ethereum: probably the most famous crypto-currency after Bitcoin, it enables smart contract and the emission of new coins on its blockchain, making it a central player in the crypto-currency world.
- Ripple: backed by a company, it is promoted as a future instrument used by banks to settle their payments.
- NEO: has the same ambition as Ethereum, but is developed for the Chinese market, which has revealed to be a very optimistic market for crypto-currencies, despite the regulations and the hostility of the government.
- Ethereum Classic: Ethereum classic is a fork of Ethereum (the two blockchain splitted at some point in time) and is therefore a competitor of Ethereum (and NEO). A dispute among the developers of the initial project is at the origin of the fork.
- Iota: a crypto-currency that will run on the internet of things
- Litecoin: a little brother of Bitcoin that follows the tracks of the first crypto-currency, implementing the same technologies but with more supply.
- Monero: A competitor of Dash, Monero is a crypto-currency aimed at insuring complete anonymity to the users, thanks to a special algorithm. Monero, on the contrary of Dash, is an open-source project and not run by a company.
- ZCoin: Similar project using the Zero technology to secure financial privacy of its user.

Some of these coins are extremely similar, to stress the effects of the indicators measured. To help the reader visualize, we can divide the 10 coins in 3 clusters: the “regular” coins (Bitcoin, Litecoin, Ripple), the “smart-contract” coins (Ethereum, Ethereum Classic, NEO, Iota) and the “anonymity” coins (Dash, Monero, ZCoin); even though Iota is probably apart.

Between the 1st of December 2017 until the 1st of January 2018, a robot on the cloud has been executing a series of steps to redeem several data, for each of the 10 crypto-currencies described above. Every 24 hours at 23:30 (Swiss Time), the robot collected data about:

1/ Dollar variation in percent of the coin (on coinmarketcap.com)

2/ Data of the official twitter account, as registered on coinmarketcap.com, the tweet count, the followers count and the appearance in public groups.

3/ Date of the latest commit on the official github page of the coin

Since we needed coins that had an official twitter account on the 1st of December 2017 and an open-source github repository, we were constrained on the choice of coins to study. For instance, we could not study the Bitcoin Cash (a fork of Bitcoin) due to the decentralized characteristic of the coin and the absence of official twitter account.

The scrapper has been running on a cloud machine every day at 23:30 (Swiss time) in order to get proper time intervals and consistent data.

IV. Assumptions to challenge and linear regression

The Github page displays the technological advancement of the coin. The fact that the team updates the code frequently is the proof that they are working on improving their product and meeting the deadlines. If the team develops a critical update, they will also most certainly tweet about it. Twitter plays the role of the modern forum and is the place par excellence where company will make public announcements. If they improve their product, chances are high that their community managers will update the public to comfort holders that they've made a sane investment and attract new investors, which would likely increase the dollar value of the coin. This paper challenges these assumptions on a given month and a given set of crypto-currencies.

If social media activity and technological improvements are correlated with the changes in price, a linear regression should display statistically significant coefficients and reasonable error term. If, however, crypto-currencies' variations in price are dictated by exogenous factors, such as pure speculation or other crypto-currencies, our model should display strong errors and insignificant coefficients. It is however important to note that this model is built on the assumption that twitter is used by the official accounts as a public interface with the world so they wouldn't miss an important announcement, even though it was made by an outside actor.

It is possible that some crypto-currencies are closely related to their technological advancement or their social media activity and others not at all. For this reason, we will try several regressions to try and notice patterns in the variation of specific crypto-currencies:

The general regression will look like the following:

$$\log(Price) = \beta_0 + \beta_1\,DailyTweets + \beta_2\,DailyFollowers + \beta_3 DailyPublicAppearance$$

$$+\beta_4 WasCommitted + u$$

Log(Price) comes from the fact that we are dealing with the percentage change in dollar value of the coin. DailyTweets will be calculated for every day, knowing the number of tweets the day before. The same will be done with DailyFollowers, to track the attention the coin is having on social media. Again, the assumption to test is that more followers means more attention on the coin, mechanically more investors and therefore an increase in value of the coin. PublicAppearance will track the number of public groups in which the coin is located, it is another indicator of the social media attention. WasCommitted is a dummy variable that is equal to 1 if the Github repository of the project has been committed in the last 24h, it therefore tracks the technological activity of the coin.

We will apply consecutively the model to each coin. We have 32 days of data, so we can use n=31 days, since we need to calculate the daily variation for the number of tweets, the number of followers and the public appearance with the last day.

Finally, we will run the regression indifferently of the coin to see if we have a general pattern for the crypto-currencies in general.

V. Results

```
[1] "Bitcoin"

Call:
lm(formula = `24hPerVariation` ~ TweetCount + FollowersCount +
    PublicAppearance + WasCommitted, data = crypto_frame)

Residuals:
    Min      1Q  Median      3Q     Max
-15.055  -6.631  -1.390   3.430  24.427

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       5.9242219  6.9520962   0.852    0.402
TweetCount       -0.1321108  2.0364771  -0.065    0.949
FollowersCount    0.0003045  0.0012783   0.238    0.814
PublicAppearance -0.2243992  0.2317713  -0.968    0.342
WasCommittedTRUE -0.1878628  3.5546876  -0.053    0.958

Residual standard error: 8.989 on 26 degrees of freedom
Multiple R-squared:  0.04522, Adjusted R-squared:  -0.1017
F-statistic: 0.3078 on 4 and 26 DF,  p-value: 0.8701
```

Bitcoin displays only insignificant parameters (p-value > 0.10), due to extremely high standard errors. It appears obvious that Bitcoin is already a mature and famous technology that doesn't need social media activity and now has a slow development, so Github commits would be rare for the huge community to be able to follow with the changes. Also the number of actors to engage in the negotiation for a change in the code is quite high, even though the developer community of Bitcoin is quite centralized. The regression has a low R-squared and even a negative R-squared, which means the model fits worse than a horizontal line.

```
[1] "Dash"

Call:
lm(formula = `24hPerVariation` ~ TweetCount + FollowersCount +
    PublicAppearance + WasCommitted, data = crypto_frame)

Residuals:
    Min      1Q  Median      3Q     Max
-16.862  -4.264  -1.186   5.261  25.996

Coefficients: (1 not defined because of singularities)
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)      -4.1630620  5.5816652  -0.746    0.462
TweetCount        0.0928445  1.0898205   0.085    0.933
FollowersCount   -0.0005363  0.0028218  -0.190    0.851
PublicAppearance  0.3317326  0.2905624   1.142    0.264
WasCommittedTRUE         NA         NA      NA       NA

Residual standard error: 9.115 on 27 degrees of freedom
Multiple R-squared:  0.08032, Adjusted R-squared:  -0.02187
F-statistic: 0.786 on 3 and 27 DF,  p-value: 0.5122
```

Interestingly enough, Dash hasn't made any commit for the whole month of December, which shows that even though the currency is supported by a company, the development can be made public quite slowly. The other parameters are statistically insignificant with p-values above 0.10. We can make the same remark about the R-squared as Bitcoin.

```
[1] "Ethereum"

Call:
lm(formula = `24hPerVariation` ~ TweetCount + FollowersCount +
    PublicAppearance + WasCommitted, data = crypto_frame)

Residuals:
     Min       1Q   Median       3Q      Max
-17.5720  -3.7794  -0.1901   4.0064  18.0548

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      -6.294521   6.078483  -1.036    0.310
TweetCount       -0.394588   2.536899  -0.156    0.878
FollowersCount    0.004954   0.003198   1.549    0.133
PublicAppearance -0.011762   0.186233  -0.063    0.950
WasCommittedTRUE -2.251018   2.955449  -0.762    0.453

Residual standard error: 7.566 on 26 degrees of freedom
Multiple R-squared:  0.09712, Adjusted R-squared:  -0.04179
F-statistic: 0.6992 on 4 and 26 DF,  p-value: 0.5995
```

Ethereum also has only insignificant parameters, even though the followers count is a way more significant than the others. We can make the same remark about the R-squared as Bitcoin.

```
[1] "Ripple"

Call:
lm(formula = `24hPerVariation` ~ TweetCount + FollowersCount +
    PublicAppearance + WasCommitted, data = crypto_frame)

Residuals:
    Min      1Q  Median      3Q     Max
-27.673 -10.251  -3.369   3.191  60.689

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      -4.235747   8.768237  -0.483    0.633
TweetCount        0.138127   1.496589   0.092    0.927
FollowersCount    0.000485   0.001826   0.266    0.793
PublicAppearance  0.209442   0.349031   0.600    0.554
WasCommittedTRUE -5.527072  23.284975  -0.237    0.814

Residual standard error: 22.1 on 26 degrees of freedom
Multiple R-squared:  0.1213,  Adjusted R-squared:  -0.01385
F-statistic: 0.8975 on 4 and 26 DF,  p-value: 0.4796
```

Ripple displays highly insignificant parameters. We can make the same remark about the R-squared as Bitcoin.

```
[1] "NEO"

Call:
lm(formula = `24hPerVariation` ~ TweetCount + FollowersCount +
    PublicAppearance + WasCommitted, data = crypto_frame)

Residuals:
     Min       1Q   Median       3Q      Max
-14.7664  -7.3891  -0.2664   5.5509  24.6527

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)      -2.7192479  6.7720101  -0.402    0.691
TweetCount       -2.1169062  1.6848780  -1.256    0.220
FollowersCount    0.0002096  0.0035356   0.059    0.953
PublicAppearance  0.5833004  0.4101877   1.422    0.167
WasCommittedTRUE -2.6443182  3.7258840  -0.710    0.484

Residual standard error: 10.03 on 26 degrees of freedom
Multiple R-squared:  0.154, Adjusted R-squared:  0.02383
F-statistic: 1.183 on 4 and 26 DF,  p-value: 0.3414
```

NEO has better parameters, public appearance and tweet count having not so high p-values, and the R-squared are higher than the previous regressions. However, tweet count has a negative coefficient, which means the more the account tweets, the more the crypto decreases.

```
[1] "Ethereum Classic"

Call:
lm(formula = `24hPerVariation` ~ TweetCount + FollowersCount +
    PublicAppearance + WasCommitted, data = crypto_frame)

Residuals:
     Min       1Q   Median       3Q      Max
-23.2149  -3.0925  -0.2933   3.6802  25.3876

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      -6.405234   6.783579  -0.944    0.354
TweetCount       -0.421690   0.611171  -0.690    0.496
FollowersCount    0.006124   0.005486   1.116    0.274
PublicAppearance  0.100961   0.391623   0.258    0.799
WasCommittedTRUE  2.284981   4.213091   0.542    0.592

Residual standard error: 9.11 on 26 degrees of freedom
Multiple R-squared:  0.06061, Adjusted R-squared:  -0.08391
F-statistic: 0.4194 on 4 and 26 DF,  p-value: 0.7931
```

Ethereum Classic also displays insignificant parameters and a negative adjusted R-squared.

```
[1] "IOTA"

Call:
lm(formula = `24hPerVariation` ~ TweetCount + FollowersCount +
    PublicAppearance + WasCommitted, data = crypto_frame)

Residuals:
    Min      1Q  Median      3Q     Max
-28.381  -8.368   1.003   7.547  37.250

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      -9.405456   8.032351  -1.171   0.2522
TweetCount       -2.768209   2.905755  -0.953   0.3495
FollowersCount    0.015762   0.007756   2.032   0.0525 .
PublicAppearance -0.277180   0.608905  -0.455   0.6527
WasCommittedTRUE  4.914767  12.408573   0.396   0.6953
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.53 on 26 degrees of freedom
Multiple R-squared:  0.1847,  Adjusted R-squared:  0.05932
F-statistic: 1.473 on 4 and 26 DF,  p-value: 0.2391
```

IOTA has a significant follower count, since it has more than doubled its number of followers over the month of December. The value of the crypto almost tripled over the month of December (from $1.28 to $3.50), which most likely attracted a lot of investors. The R-squared is quite high as well.

```
[1] "Litecoin"

Call:
lm(formula = `24hPerVariation` ~ TweetCount + FollowersCount +
    PublicAppearance + WasCommitted, data = crypto_frame)

Residuals:
    Min      1Q  Median      3Q     Max
-24.832  -6.106   1.000   5.269  32.307

Coefficients: (2 not defined because of singularities)
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      -12.50289    7.29153  -1.715 0.097447 .
TweetCount              NA         NA      NA       NA
FollowersCount     0.15354    0.04073   3.769 0.000777 ***
PublicAppearance  -1.87023    1.20236  -1.555 0.131067
WasCommittedTRUE        NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.3 on 28 degrees of freedom
Multiple R-squared:  0.3723,  Adjusted R-squared:  0.3275
F-statistic: 8.305 on 2 and 28 DF,  p-value: 0.001473
```

The Litecoin official account made absolutely no tweet during the month of December and the official Github account also made no commit the whole month. It appears that the parameters are pretty significant with a very high factor on followers count. The R-squared is quite high, but the parameters overweight the follower count due to the absence of two of them. The Litecoin had very strong variations (up to 53% in one day), without twitting or committing on the repository.

```
[1] "Monero"

Call:
lm(formula = `24hPerVariation` ~ TweetCount + FollowersCount +
    PublicAppearance + WasCommitted, data = crypto_frame)

Residuals:
     Min       1Q   Median       3Q      Max
-20.9782  -5.7693   0.6171   5.2550  19.9490

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      -4.896838   7.599067  -0.644   0.5250
TweetCount        0.735501   0.615255   1.195   0.2427
FollowersCount   -0.004003   0.002657  -1.507   0.1439
PublicAppearance  0.679470   0.269962   2.517   0.0183 *
WasCommittedTRUE  1.177947   3.990953   0.295   0.7702
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.435 on 26 degrees of freedom
Multiple R-squared:  0.2175,  Adjusted R-squared:  0.09707
F-statistic: 1.806 on 4 and 26 DF,  p-value: 0.1579
```

Public appearance is the only significant parameter for Monero, however the follower count has a negative parameter, which is quite unexpected. The R-squared are fair enough, but it is hard to determine whether it's due to randomness.

```
[1] "ZCoin"

Call:
lm(formula = `24hPerVariation` ~ TweetCount + FollowersCount +
    PublicAppearance + WasCommitted, data = crypto_frame)

Residuals:
    Min      1Q  Median      3Q     Max
-21.006  -6.888   0.776   6.347  41.450

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      -3.41576    6.35987  -0.537  0.59578
TweetCount       -1.30406    0.68363  -1.908  0.06755 .
FollowersCount    0.05227    0.01769   2.955  0.00656 **
PublicAppearance -1.25363    1.11002  -1.129  0.26905
WasCommittedTRUE -1.44038    5.67091  -0.254  0.80150
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.1 on 26 degrees of freedom
Multiple R-squared:  0.2661,	Adjusted R-squared:  0.1532
F-statistic: 2.357 on 4 and 26 DF,  p-value: 0.0799
```

The ZCoin has a statistically significant tweet count factor and follower count factor. This means the coin's price can partly be linked with its social media activity and the fact that it has drawn attention. It gained so many followers that the tweet count has a negative parameter. The R-squared are also decent.

```
[1] "All cryptos"

Call:
lm(formula = `24hPerVariation` ~ TweetCount + FollowersCount +
    PublicAppearance + WasCommitted, data = crypto_frame)

Residuals:
    Min      1Q  Median      3Q     Max
-33.562  -6.596  -1.942   4.329  70.790

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       3.927e+00  1.295e+00   3.032  0.00263 **
TweetCount        2.394e-05  3.249e-04   0.074  0.94133
FollowersCount   -7.033e-06  2.238e-05  -0.314  0.75358
PublicAppearance  4.556e-04  1.642e-03   0.277  0.78165
WasCommittedTRUE -1.219e+00  1.771e+00  -0.688  0.49170
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.8 on 315 degrees of freedom
Multiple R-squared:  0.002635,	Adjusted R-squared:  -0.01003
F-statistic: 0.2081 on 4 and 315 DF,  p-value: 0.9339
```

Mixing all of them yields also insignificant parameters, except for the intercept that displays an overall increase of 4% daily, for any randomly selected crypto out of the 10. The R-squared is also very low.

VI. Conclusions and limitations

The results are very clear: we cannot draw a strong and consistent link between the 24-hour variation in price of a crypto-currency and Github repository update or social media activity. The statistically significant parameters that we found seem to have been quite random and were not consistent between all the models. The general model shows that there is no possible pattern.

Therefore, the assumptions on community building and technology improvement being a strong factor are quite weak, or at least are not based on pure twitter activity or Github commits. Indeed, even if the Github repository is committed, there is no certainty that the development has hit a milestone or an important improvement. Also, the official account can write useless tweets or unrelated tweets.

One obvious limitation of this study is the focus on quantitative and not qualitative. Unfortunately, it is hard to assess programmatically the quality of the tweets and their impact on the market value of the crypto.

Another limitation is the set of crypto, that has been quite arbitrary and consisted in different projects goals and different maturities. On this point of view, it was not very consistent and maybe one should focus on very early stage cryptos to monitor the "take-off" in terms of social media impact and technological development.

The model could have incorporated more data from twitter and the number of tweets addressed to the twitter account, or the number of mentions. It would have also been interesting to scrap specialized forums such as crypto subreddits to monitor the number of references to the crypto made on a daily basis and check whether attention on forum has an impact on the price of the crypto in a 24-hour timespan.

Also, many whales use private chats to communicate and coordinate to buy a specific crypto-currency, to attract smaller investors and inflate artificially the price of the crypto. Of course, it requires paying or being initiated, which is not directly "social media" and is also against orthodox financial theory of effective markets.

---

[i] https://bitslog.wordpress.com/2013/04/24/satoshi-s-fortune-a-more-accurate-figure/

[ii] http://fortune.com/2017/12/26/bitcoin-price-value-bubble-crash-says-morgan-stanley-analyst/

[iii] https://news.bitcoin.com/cracks-appear-critics-label-bitconnect-ponzi-scheme/