Cryptocurrencies - Google TRENDS

Cryptocurrencies have become an extremely relevant topic in national and international media channels, especially after the price surges of the most well-known one, Bitcoin, in 2017. As this topic receives attention and is intrinsically digital, a question arises: What is the relationship between investor attention through internet searches and the characteristics related to the price of these assets?

Therefore, this study aims to evaluate the econometric/statistical characteristics of the price relationships of three cryptocurrencies, Bitcoin, Ethereum, and Litecoin, with the level of searches for these terms on Google. For this purpose, weekly data were collected from the Google Trends search platform and the CoinBasePro website, in the period of x. The analyses were conducted using Python code, with the assistance of the statsmodels library, analyzing the following characteristics between the time series: Correlation, Stationarity, Cointegration, Causality, and the performance of ARIMA and GARCH models, with and without the use of search data.

Introdução

Cryptocurrencies have been a topic discussed in scientific circles for some time now [https://www.sciencedirect.com/science/article/pii/S1572308914001259?via%3Dihub#bib0060] , but it is noted that they have become a very popular topic in media channels in recent years, following the 2017 price hike of the first and main cryptocurrency, Bitcoin [reference to search growth]. The system launched by the person or organization, whose pseudonym is Satoshi Nakamoto, solved the problem of double spending in online transactions by creating a secure and independent system from financial institutions. Transactions occur directly from person to person, secured by a chain of encrypted codes [https://bitcoin.org/bitcoin.pdf].

Thus, due to the fact that cryptocurrencies are intrinsically linked to the virtual world, it is evident that fluctuations/volatility in the price of this asset are more than ever linked to speculations and behavior associated with the internet. As an example of this, one can note the recent devaluation of Bitcoin and the rapid appreciation of Dogecoin after comments by Elon Musk, CEO of the automotive industry Tesla, on his personal Twitter profile [https://www.lemonde.fr/economie/article/2021/06/14/tesla-rouvre-la-porte-a-un-paiement-en-bitcoin-sous-reserve-d-une-devise-moins-polluante_6084025_3234.html?utm_term=Autofeed&utm_medium=Social&utm_source=Twitter#Echobox=1623649311].

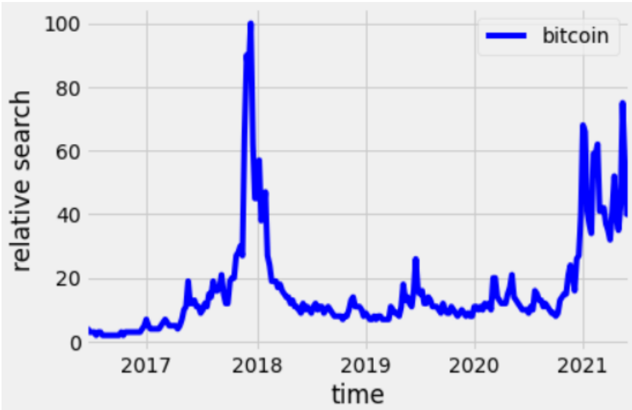
Therefore, as all emissions and transactions involve the internet, the use of search and social media data can be interesting predictors for the price of cryptocurrencies. In this sense, the decision was made to take into account the frequency data per search term on Google, made available through the Google Trends platform. The choice of this tool was made due to its widespread use in various areas such as: healthcare [https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0109583], real estate, financial market [https://www.nature.com/articles/srep01684?__hstc=113740504.2a1e835c34ab7bf88e972fdd7a7debc8.1424476800061.1424476800062.1424476800063.1&__hssc=113740504.1.1424476800064&__hsfp=3972014050], car sales [https://onlinelibrary.wiley.com/doi/abs/10.1002/for.1252], and even for bitcoin itself [https://www.nature.com/articles/srep03415].

In this sense, this study aims to contribute to the understanding of the relationship between the attention of investors through internet searches and the characteristics related to the price of cryptocurrencies, specifically Bitcoin, Ethereum, and Litecoin, by evaluating their econometric/statistical characteristics using Python code with the aid of the statsmodels library. The following characteristics were analyzed between the time series: correlation, stationarity, cointegration, causality, and the performance of ARIMA and GARCH models, with and without the use of search data. Data was collected from the Google Trends platform and the CoinBasePro website on a weekly basis in the period of x.

Therefore, this study aims to contribute to the predictive possibility of cryptocurrency prices with the help of Google Trends.

Method/data

For this purpose, Google Trends search data was collected while cryptocurrency prices were collected from the CoinBase Pro website, using Python code. Since Google Trends does not differentiate between upper and lower case characters, the collected time series involve the following terms: 'bitcoin', 'ethereum', 'litecoin'. From this, the returns were grouped into weekly averages, as Google Trends only provides data on this time scale in the long term.



Correlação

To estimate the correlation between price time series and their respective Google search trends, Pearson and Spearman correlation coefficients were used. The choice of these two coefficients is necessary because the Pearson coefficient measures the linear relationship between two variables, while the Spearman coefficient assesses the monotonic relationship between two variables, where the rate of change does not necessarily need to be constant [https://support.minitab.com/pt-br/minitab/18/help-and-how-to/statistics/basic-statistics/supporting-topics/correlation-and-covariance/a-comparison-of-the-pearson-and-spearman-correlation-methods/]. Thus, a comparison between the results obtained through both methods allows for evaluating prediction models. The correlations of the price time series with their respective Google Trends searches can be seen in the table below:

SÉRIE TEMPORAL	PEARSON	SPEARMAN
Bitcoin-close	0.675211	0.749537
Bitcoion-volume	0.650855	0.794395
Ethereum-close	0.885527	0.847960
Ethereum- volume	0.634139	0.703158
Litecoin- close	0.489426	0.763695
Litecoin- volume	0.694352	0.830188

The obtained data show that the Spearman correlation coefficient is higher for most variables, which may provide an interesting perspective for the possibility of choosing non-linear forecasting models. The best performance of non-linear models to predict cryptocurrency prices with Google Trends has already been demonstrated [https://link.springer.com/article/10.1007/s42786-021-00027-4#:~:text=There%20is%20no%20much%20improvement,although%20not%20consistently%20through%20time].

pearson

$$Cor(X,\ Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

spearman

$$Spearman\ =\ 1\ -\ \frac{6\sum_i d_i^2}{n(n^2-1)}$$

Stationarity

There are many definitions for the stationarity of time series, but in general, it is said that a time series is stationary when it fluctuates randomly around a constant mean. A series is said to be strictly stationary when the statistical characteristics (mean, variance, and autocorrelation) of its joint distribution do not vary with time.

To prove the stationarity or non-stationarity of time series, there are several tests, and this study chose the Augmented Dickey-Fuller Test (ADF) [Bronlee, Introduction to time series]. This test applies an autoregressive model, evaluating whether the slope coefficient tends to zero, which would make the process stochastic.

$$\Delta Y_t\ =\ \Delta Y_{t-1}\ +\ \epsilon_t$$

If the test result has some time dependency, that is, if the test value is statistically significant, the null hypothesis, that the series is stationary, is rejected. In addition, it is noted that sometimes the ADF fails to reject the null hypothesis. Therefore, we opted to use the KPSS test [http://www.portalaction.com.br/sites/default/files/SeriesTemporais/planilhas/kpps_test.pdf] to confirm the non-stationarity of most of the time series present in this study. Unlike the ADF, the null hypothesis for the KPSS test is that the series is non-stationary. Both tests were implemented using the statsmodel library [https://www.statsmodels.org/stable/examples/notebooks/generated/stationarity_detrending_adf_kpss.html].

	KPSS		ADF	
SÉRIE TEMPORAL	P-VALUE	TEST STATISTIC	P-VALUE	TEST STATISTIC
Bitcoin-close	0.010000	0.784085	0.039991	-2,948504
Bitcoin-volume	0.100000	0.342779	0.,000075	-4,725669
Bitcoin-GT	0.100000	0.270875	0.305790	-1,956887
Ethereum-close	0.042008	0.498485	0.852875	-0,675972
Ethereum-volume	0.061773	0.435686	0.000000	-6,69398
Ethereum- GT	0.090714	0.368543	0,927855	-0,283164
Litecoin- close	0.100000	0.318790	0,154134	-2,357362
Litecoin- volume	0.100000	0.146241	0,242244	-2,105663
Litecoin-GT	0.100000	0.105307	0.000000	-5,934196

As expected, the time series are not stationary. Following the convention in the financial market [https://epge.fgv.br/we/MD/EconometriaAplicadaFinancas/2009?action=AttachFile&do=get&target=Financial_Econometrics_1.pdf], we opted for logarithmic returns to solve the problem. These returns, in turn, proved to be stationary according to the tests applied, so we decided to use them.

$$retorno = ln(\frac{P_t}{P_{t-1}})$$

Cointegration

Cointegration tests demonstrate whether there is any long-term relationship between two time series. In this study, we opted to use the Engel-Granger test, implemented through the stattools library [https://www.statsmodels.org/stable/generated/statsmodels.tsa.stattools.coint.html], which works as follows: Considering two stationary time series, a linear relationship between the

two series is discovered through regression. From this, the residuals of this regression are considered, and it is verified whether they are stationary through the ADF. If the residuals form a stationary process, then the two series are cointegrated, that is, they present a relationship.
[https://www.econ.queensu.ca/sites/econ.queensu.ca/files/wpaper/qed_wp_1227.pdf][econometrics textbook]. The results for the variables present in this study are presented in the table below:

COINTEGRAÇÃO-PESQUISAS GT

RETORNOS	P-VALUE	TEST STATISTIC
Bitcoin-close	2,009E-21	-12,12635
Bitcoin-volume	1,43E-15	-12,657048
Ethereum-close	7,17E-21	-11,88008
Ethereum-volume	1,21E-15	-12,22544

As can be observed in the tables, the values appear to be cointegrated (except for Litecoin, which has a problem with returns).

Granger Causality

If past values of a time series (X) help predict future values of another time series (Y), then it is said that series X Granger causes Y. Mathematically, this can be understood as a bivariate regression for lagged values of X. The null hypothesis of the Granger causality test is that X does not cause Y. Therefore, the p-values shown in the table below indicate that we cannot reject the hypothesis that GT cannot cause cryptocurrency prices, meaning that it is possible for GT to cause them.

$$Y_t = \epsilon_t + \sum_{j=1}^{m} a_t Y_{t-j} + \sum_{j=1}^{m} b_t X_{t-j}$$

GRANGER CAUSALITY

Relações de causalidade	lag-1	lag-2	lag-3
Bitcoin(close)->GT	0,011	0,0264	0,02
GT->Bitcoin(close)	0,053	0,0159	0,038
Ethereum(close)->GT	0,619	0,248	0,261
GT->Ethereum(close)	4E-07	2E-10	9E-11

ARIMA:

I plan to compare univariate forecasting with forecasting using the exogenous variable, GT.

GARCH:

I plan to compare univariate forecasting with forecasting using the exogenous variable (GT).

Resultados

Discussão

Referências