Kennesaw State University
Coles College of Business
Bagwell Center for the Study of Markets and Economic Opportunity

Undergraduate Research Fellowship Working Paper Series

Title:

"A Time Series Analysis of Crypto Currency Price Data"

Author(s):

Dave Hagemann

2018

A Time Series Analysis of Crypto Currency Price Data*

Dave Hagemann†

June 27, 2018

Abstract

Crypto currency markets have recently become more and more popular, but are clearly in their infancy relative to developed financial markets. Using prices series data gathered using web-scraping techniques on the more well-known coins such as Bitcoin and Ethereum, as well as an "alt" coin called Monero, I first test these time series to determine whether or not they are stationary using the Augmented Dickey-Fuller test, and as is usual with price data, find that they are not. After detrending the data, then investigate whether there are any Granger causality relationships between the different price series, and comment on whether this suggests anything about the state of the Efficient Market Hypothesis in this relatively young financial market.

JEL Classification Codes: C58; G14.

Keywords: Crypto currencies; time series; Granger causality.

*I am very grateful to my advisor, Dr. James Boudreau for all his help with this project, and emphasize that all errors are my own.

†Kennesaw State University, email: dhagema1@students.kennesaw.edu.

1 Introduction

While the notion of “crypto currencies” as actual currencies may be dubious, given the slow rate of adoption of them as a form of medium of exchange, they have undoubtedly become a form of financial asset traded in markets. As with any other asset, they therefore generate price data over time that may be analyzed with the tools of econometrics that are specialized for time series data.[1]

What makes crypto currency markets particularly interesting, as compared to more traditional financial markets, is the simple fact that they are in their infancy. In traditional financial markets, supposedly, many buyers and sellers are very informed and sophisticated, all motivated to earn profits by buying high and selling low. Any obvious arbitrage opportunities should therefore be taken advantage of immediately, thereby making price movements essentially unpredictable.

This is the notion of the Efficient Market Hypothesis, which has a long history, but was perhaps most popularized by Fama and French (1988). It suggests that arbitrage opportunities should be eliminated by sufficient competition and fully informed trading, and that previous prices should not predict future prices profitably. This may not be the case in newer, less developed markets such as those for crypto currency, however. I would like to use publicly available data to examine whether or not that is true. In particular, could it be that the price of one or more of the more dominant crypto currency assets could have been used in the past to predict the price of others? So far I am far from any definitive evidence, but this paper presents an initial econometric foray into the investigation.

[1] Though these assets may not actually be true currencies, I will use the terms currency and coins throughout the paper since that is how these assets are commonly referred to.

2 The Econometric Model

The notion of forecasting is that it is possible to predict one variable's value in the future based on currently available data. In particular, in order to profit from trading financial assets, it would be a major advantage if a person could predict the price of an asset one day (or even an hour, or even a few minutes or a few seconds) ahead of its current price with more accuracy than other traders. The field of time series econometrics can be applied to such purposes, though the the field also has much broader applications.

The approach I use for this project is one of the most fundamental in modern time series analysis, and the model I describe here is explained in more detail in the first few pages of Lütkepohl (2005, pp. 1-5). The idea is that, since time series data tends to have consistent trends or patterns, the price of financial asset X one period in the future, P^x_{t+1}, can be predicted based on the asset's current price, P^x_t, and a number of it's prices in the past. P^x_{t-1} would represent the asset's price in the previous time period, for example, P^x_{t-2} would be the asset's price two period's in the past, and so on. (Again, the size of one time period could be a day, an hour, a minute, or any interval, depending on the data available.) This is known as an autoregressive process.

Assuming the relationship between an asset's future price and it's current and past prices is linear, the relationship can be written as

$$P^x_{t+1} = \beta_0 + \beta_1 P^x_t + \beta_2 P^x_{t-1} + \ldots + \beta_k P^x_{t-k}$$

where k is the furthest date in the past considered, also known as the maximum number of lags, β_0 is a constant term, β_1 measures the impact of the price of x in the current period on its future price, and so on. Knowing the prediction will not be perfect, however, because the prices of assets are affected by random elements that can not be predicted from period to period, it is more accurate to estimate the model based on available data as

$$P^x_{t+1} = \beta_0 + \beta_1 P^x_t + \beta_2 P^x_{t-1} + \ldots + \beta_k P^x_{t-k} + \epsilon_{t+1},$$

where ϵ_{t+1} represents the influence of randomness in the future time period. Assuming the relationship is consistent, it should then also be the case that

$$P^x_t = \beta_0 + \beta_1 P^x_{t-1} + P^x_{t-2} + \ldots + P^x_{t-k-1} + \epsilon_t,$$

and that relationship can then be estimated using linear regression.

A key assumption in order to use linear regression, however, is that the random influence terms in each time period should be independent of one another. In time series data, this is often not the case, since random factors over time are often correlated with one another. That is, the data is not stationary. A statistical test known as the Augmented Dickey-Fuller test can be used to check whether or not this is the case.

If the data is found to be non-stationary, one method of transforming it to make it stationary is to look at the changes in prices from one time period to another, rather than the prices themselves. This is known as first-differencing, and is often successful in making the data stationary, since although the random influences on prices may be correlated from one time period to the next, the random influences on just how much prices change is less likely to be. Letting

$$\Delta P^x_t = P^x_t - P^x_{t-1}$$

represent the change in the price of an asset, x, from time period $t-1$ to t, the econometric model then becomes

$$\Delta P_t^x = \beta_0 + \beta_1 \Delta P_{t-1}^x + \beta_2 \Delta P_{t-2}^x + \ldots + \beta_k \Delta P_{t-k-1}^x + \xi_t,$$

where ξ_t represents the random factor impacting the change in the price of the asset from period $t-1$ to period t.

Finally, it may also the case that the current and past prices of other assets can be used as information to predict the price of x. For example, with two assets, x and y, it might be useful to estimate two equations,

$$\Delta P_t^x = \beta_0^x + \beta_1^x \Delta P_{t-1}^x + \beta_2^x \Delta P_{t-2}^x + \ldots + \beta_{k+1}^x \Delta P_{t-1}^y + \beta_{k+2}^x \Delta P_{t-2}^y + \ldots + \xi_t^x$$

$$\Delta P_t^y = \beta_0^y + \beta_1^y \Delta P_{t-1}^y + \beta_2^y \Delta P_{t-2}^y + \ldots + \beta_{k+1}^y \Delta P_{t-1}^x + \beta_{k+2}^y \Delta P_{t-2}^x + \ldots + \xi_t^y$$

Since these equations can be expressed in terms of vectors and matrices, this is known as a vector autoregression, and can be performed for any number of variables (x, y, z, etc). Using data to perform this type of analysis, it is possible to estimate the magnitudes of the β parameters, and therefore the influence of an asset's own past values on it's future price, as well as the influence of other assets' prices. We can then use that information to test for what is known as Granger causality, which suggests that the price of one asset significantly impacts the future value of another in a statistical sense.

3 Data

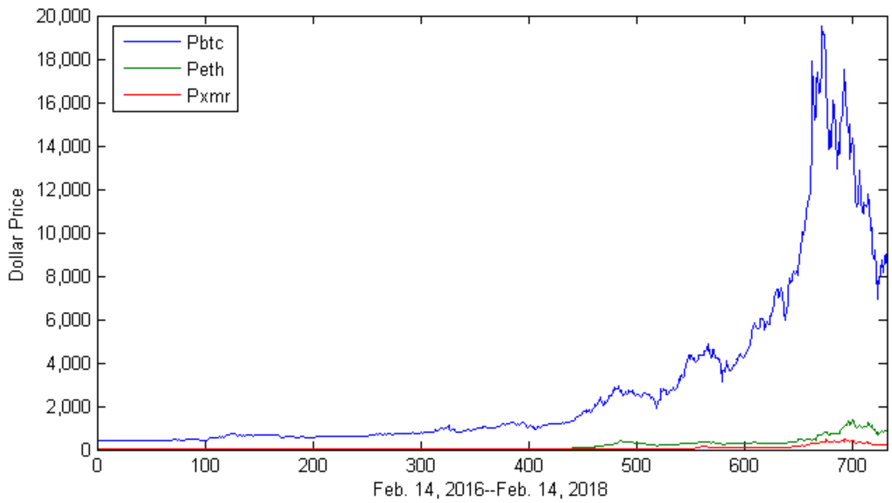
The data I use for this paper comes from a publicly available source, www.kaggle.com, which lists daily price data for a number of crypto coins. This is in contrast to traditional price data which is much more frequent, often by the second or faster, but since this is a publicly available source that covers a significant time period, I have chosen to use it for this paper.

For purposes of comparison I have chosen to examine price data on three price series: Bitcoin (Btc), Ethereum (Eth), and Monero (Xmr). I have chosen these because Bitcoin was the original crypto currency and therefore often seems to lead the market, and certainly is the most well-known. Ethereum, on the the other hand, is another dominant force in the current market and seems to be another very stable presence, though explaining its specifics is beyond the scope of this paper.

Finally, I also consider the price of Monero, an "alt" (short for alternative" coin, which is newer than the other two, and known for being more anonymous and thus perhaps of more use in dark markets. I am interested in whether the prices of the more dominant, established currencies offer patterns that may predict changes in this newer, alternative currency.

I use data on the closing market price for each coin. There is also data on daily high and low prices, but so far I have found little variation in which of the three I use for comparison. Over time I hope to gather more precise data, though for now the correlation between the trends seems to be quite close. I also have chosen to use data over the time period Feb. 14, 2016 through Feb. 14 2018, mostly as an arbitrary time window. The important inclusion

Figure 1: Prices of Bitcoin, Etereum, and Monero, 2/14/16–2/14/18



has been the most recent month and a half, during which the market experienced a sharp decline. I have experimented with partitioning the data, but so far find no major changes unless only a very narrow window of time is chosen.

4 Results

Figure 1 illustrates the three price series over the two years Feb. 14, 2016 through Feb. 14, 2018. Since the price of Bitcoin dominates the other two, I also include Figure 2 to show only the prices of Ethereum and Monero over the same time period. (Note that the ticker symbol for Monero is Xmr.) Each time period represents one day.

From the graphs it is fairly clear that the prices are closely correlated with one another over time, but before analyzing any possible causal relationships, as previously mentioned it is important to consider whether or not the series are stationary. If a series is non-

Figure 2: Prices of Etereum, and Monero only, 2/14/16–2/14/18

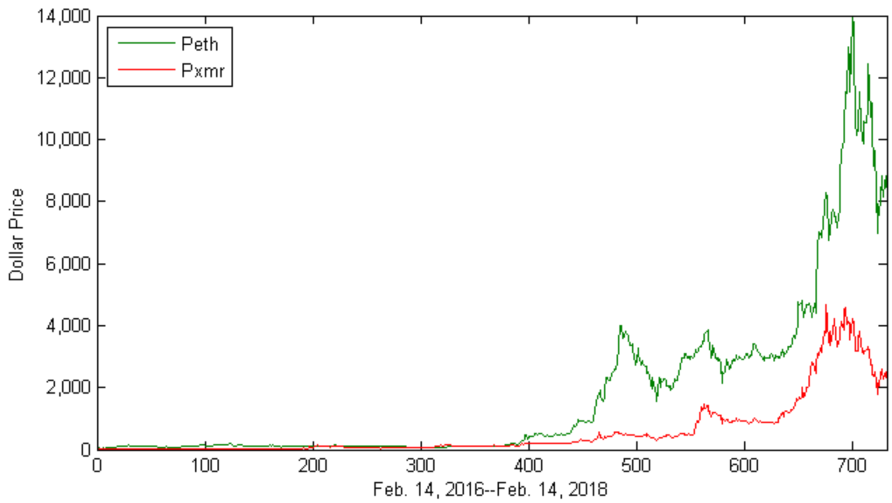


Table 1: Augmented Dickey-Fuller Statistics for unit root, 731 observations

	$Z(t)$	1% value	5% value	10% value	MacKinnnon approx. p-val. for Z(t)
Btc	-0.705	-3.430	-2.860	-2.570	0.8455
Eth	-0.168	-3.430	-2.860	-2.570	0.845
Xmr	-0.743	-3.430	-2.860	-2.570	0.8352

stationary, previous values of the price of one coin play an important role in the next day's price, and this auto-correlation can lead to faulty interpretations of statistical relationships.

To investigate whether or not the three time series are stationary or not, I use the Augmented Dickey-Fuller (ADF) Test, which is easily implemented in Stata. These results are presented in Table 1.

As expected, we reject the null hypothesis that the data are stationary at all significance levels for each series. To make the data stationary, we therefore use first differences, so each data point, rather than being simply the price at time t, P_t, is the price's change from the previous time period, $\delta P_t = P_t - P_{t-1}$. After doing this and running the ADF test

Table 2: Augmented Dickey-Fuller Statistics with differenced data, 730 observations

	$Z(t)$	1% value	5% value	10% value	MacKinnnon approx. p-val. for Z(t)
DBtc	-24.382	-3.430	-2.860	-2.570	0.0000
DEth	-24.931	-3.430	-2.860	-2.570	0.0000
DXmr	-30.344	-3.430	-2.860	-2.570	0.0000

again, the data appear to be stationary. These results are presented in Table 2.

With the data stationary, it is possible to put all three time series into a vector auto regression model (VAR) to test whether the prices and lagged prices of each variable impact one another, thereby implying a form of causality. To determine the optimal number of lags to include in the regression, the Schwarz Bayesian Information Criterion statistic can be used, and in this case indicated that four was the optimal number after running the regression once with a larger number of lags. The results of the VAR are included as a picture in the paper's appendix, since the table is quite large (note that the time period is misspecified, but the 2 years covered are in fact the most recent two; crypto currencies did not exist in the 1960s).

After running the VAR, the test for Granger Causality essentially determines whether or not there is a significant relationship when the additional time series are included in the regression. For example, when looking at the price of Btc, Btc is said to Granger cause the price of Etc, if the lags its price can improve the forecast for the price of Eth. The null hypothesis is that there is no relationship, which means that the coefficients on all of the price lags of Btc will be zero in the equation for the price of Eth. The results of the Granger causality test based on the VAR results are presented below, and suggest fairly clear evidence that the time series do impact one another. The only exception is Monero's effect on Bitcoin, which falls short of the 90% confidence interval. This is interesting,

Table 3: Granger causality Wald tests

Equation	Excluded	χ^2	Deg. of Freedom	Prob> χ^2
Dif-Btc	Dif-Eth	29.58	4	0.000
Dif-Btc	Dif-Xmr	7.7169	4	0.103
Dif-Btc	ALL	54.037	8	0.000
Dif-Eth	Dif-Btc	42.761	4	0.000
Dif-Eth	Dif-Xmr	56.435	4	0.000
Dif-Eth	ALL	78.708	8	0.000
Dif-Xmr	Dif-Btc	71.915	4	0.000
Dif-Xmr	Dif-Etc	16.897	4	0.002
Dif-Xmr	ALL	90.768	8	0.000

though perhaps unexpected, given Bitcoin's dominant status relative to Monero.

Note that Granger causality is a Wald test based on the χ^2 distribution.

5 Discussion

These results are admittedly preliminary and are intended as a beginning into a longer, deeper line of research. There are many nuances to time series research, especially when it comes to interpreting causal relationships, and whether or not they can be used to predict or project future relationships. Those intricacies are complicated even more by the fact that crypto currency markets are so new, and therefore fairly volatile.

Whether or not these markets are efficient in an informational sense, as more developed financial markets are sometimes claimed to be—though not always, for example see Shiller (2000)—is not yet clear. Examining much shorter periods of data leads to different results, and Figure 1 and 2 clearly show that there has been at least one major event, perhaps even suggesting some kind of bubble has already burst. Choosing time data selectively is dangerous, however, since it may lead one to the conclusions they are looking for rather than more objective truths. Over time I hope to investigate more, with more data and more econometric tools as these markets continue to develop.

6 References

Fama, E. and French, K. (1988). Permanent and temporary components of stock prices. *Journal of Political Economy* 96, 246?73.

Lütkepohl, H. (2005). *A New Introduction to Multiple Time Series Analysis.* Springer, Berlin.

Shiller, Robert J. (2000). *Irrational Exuberance.* Princeton University Press, NY.

7 Appendix: VAR Results

Figure 3: Stata output for VAR

Vector autoregression

Sample: 07jan1960 - 02jan1962		No. of obs	= 727
Log likelihood = -10995.57		AIC	= 30.35646
FPE	= 3.06e+09	HQIC	= 30.45145
Det(Sigma_ml)	= 2.75e+09	SBIC	= 30.60263

Equation	Parms	RMSE	R-sq	chi2	P>chi2
dbtc	13	338.981	0.0883	70.373	0.0000
deth	13	23.841	0.1195	98.67637	0.0000
dxmr	13	10.0733	0.1692	148.0665	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
dbtc						
dbtc						
L1.	.203645	.0441914	4.61	0.000	.1170314	.2902587
L2.	-.0434108	.0451552	-0.96	0.336	-.1319133	.0450917
L3.	-.0502603	.046039	-1.09	0.275	-.1404951	.0399745
L4.	.0713265	.045194	1.58	0.115	-.017252	.159905
deth						
L1.	-1.654999	.6746476	-2.45	0.014	-2.977284	-.3327143
L2.	-.172144	.6899939	-0.25	0.803	-1.524507	1.180219
L3.	2.112032	.6863957	3.08	0.002	.7667215	3.457343
L4.	-2.423403	.669561	-3.62	0.000	-3.735719	-1.111088
dxmr						
L1.	-2.301328	1.63316	-1.41	0.159	-5.502262	.8996056
L2.	-1.872547	1.657032	-1.13	0.258	-5.12027	1.375175
L3.	-2.676683	1.667464	-1.61	0.108	-5.944853	.5914869
L4.	-3.459398	1.620987	-2.13	0.033	-6.636474	-.2823212
_cons	16.44752	12.51559	1.31	0.189	-8.082589	40.97763
deth						
dbtc						
L1.	-.0044967	.003108	-1.45	0.148	-.0105884	.001595
L2.	.0194667	.0031758	6.13	0.000	.0132422	.0256912
L3.	-.0060108	.003238	-1.86	0.063	-.0123572	.0003355
L4.	.0106768	.0031786	3.36	0.001	.004447	.0169067
deth						
L1.	.1822997	.0474489	3.84	0.000	.0893016	.2752978
L2.	-.1267278	.0485282	-2.61	0.009	-.2218413	-.0316143
L3.	.1208742	.0482751	2.50	0.012	.0262567	.2154917
L4.	-.0036382	.0470911	-0.08	0.938	-.0959351	.0886587
dxmr						
L1.	-.1440694	.1148623	-1.25	0.210	-.3691954	.0810565
L2.	-.5150461	.1165412	-4.42	0.000	-.7434628	-.2866295
L3.	-.0665389	.117275	-0.57	0.570	-.2963937	.1633158
L4.	-.7123417	.1140062	-6.25	0.000	-.9357897	-.4888937
_cons	1.325641	.8802382	1.51	0.132	-.3995938	3.050877
dxmr						
dbtc						
L1.	.0039523	.0013132	3.01	0.003	.0013784	.0065261
L2.	.0080064	.0013419	5.97	0.000	.0053764	.0106364
L3.	-.0042152	.0013681	-3.08	0.002	-.0068967	-.0015338
L4.	.0067024	.001343	4.99	0.000	.0040702	.0093347
deth						
L1.	.0571565	.0200482	2.85	0.004	.0178628	.0964502
L2.	.001629	.0205042	0.08	0.937	-.0385585	.0418165
L3.	.0070243	.0203973	0.34	0.731	-.0329537	.0470022
L4.	-.0606904	.019897	-3.05	0.002	-.0996878	-.021693
dxmr						
L1.	-.2981983	.0485318	-6.14	0.000	-.3933189	-.2030778
L2.	-.2855178	.0492412	-5.80	0.000	-.3820287	-.1890068
L3.	-.0188627	.0495512	-0.38	0.703	-.1159813	.0782558
L4.	-.2318409	.0481701	-4.81	0.000	-.3262525	-.1374293
_cons	.4839669	.3719197	1.30	0.193	-.2449823	1.212916