[redacted]

A Time Series Analysis of Crypto Currency Price Data

Dave Hagemann[y]

June 27, 2018

Abstract

Crypto currency markets have recently become more and more popular, but are clearly in their infancy relative to developed nancial 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 rst 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, nd that they are not. After detrending the data, then investigate whether there are any Granger causality relationships between the di erent price series, and comment on whether this suggests anything about the state of the E cient Market Hypothesis in this relatively young nancial market.

JEL CsaCsd C58; G14.

Kyw 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.

[y] 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 nancial 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 nancial markets, is the simple fact that they are in their infancy. In traditional nancial markets, supposedly, many buyers and sellers are very informed and sophisticated, all motivated to earn pro ts 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 E cient 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 su cient competition and fully informed trading, and that previous prices should not predict future prices pro tably. 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 de nitive 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 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

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

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

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

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

$$P_t^x = \ _0 + \ _1 \ P_{t\ 1}^x + \ _2 \ P_{t\ 2}^x + \ldots + \ _k \ P_{t\ k\ 1}^x + \ _t;$$

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

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

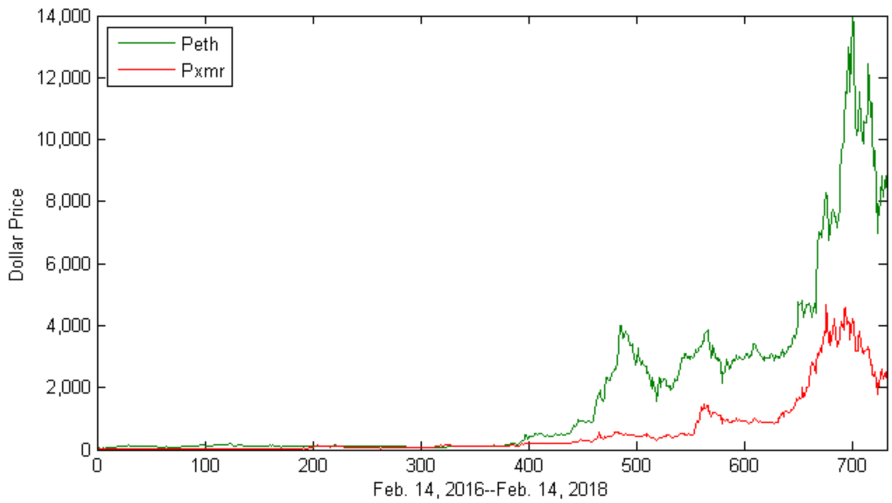

Table 1: Augmented Dickey-Fuller Statistics for unit root, 731 observations

	$Z(t)$	

(t)

Table 2: Augmented Dickey-Fuller Statistics with di erenced 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 misspeci ed, 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 signi cant 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

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

[redacted]