2014

An Analysis of Bitcoin Market Efficiency Through Measures of Short-Horizon Return Predictability and Market Liquidity

William L. Brown
Claremont McKenna College

CLAREMONT MCKENNA COLLEGE

AN ANALYSIS OF BITCOIN MARKET EFFICIENCY THROUGH MEASURES OF SHORT-HORIZON RETURN PREDICTABILITY AND MARKET LIQUIDITY

SUBMITTED TO

PROFESSOR MITCH WARACHKA

AND

PROFESSOR FAN YU

AND

DEAN NICHOLAS WARNER

BY

WILLIAM LOCKE BROWN

FOR

SENIOR THESIS

SPRING 2014

APRIL 28, 2014

Acknowledgements

I would like to thank Claremont McKenna College for having provided me with four of the most rewarding and fun years of my life. With this, I want to thank everyone at CMC who has helped me to get where I am today, including all my professors and friends. I further want to acknowledge and thank my readers, Professor Fan Yu and Professor Mitch Warachka, for reading my thesis.

I also want to thank Ms. Terri Van Eaton for all her support, encouragement, and candy, not only over the course of this project, but over the entire time I've been working at the FEI. I'd further like to acknowledge Merriel Foster for her formatting skills and support while completing thesis. And finally, I would like to thank my parents, Lee Ann and Billy; and my sister, Carlyle, for their love and support throughout life. Nothing I've done would have happened without them.

Abstract

Bitcoins have the potential to fundamentally change the way value is transferred globally. Their rapid adoption over the past four years has led many to consider the possible results of such a technology. To be a viable currency, however, it is imperative that the market for trading Bitcoins is efficient. By examining the changes in availability of predictable outsized returns and market liquidity over time, this paper examines historical Bitcoin market efficiency and establishes correlations between market liquidity, price predictability, and return data. The results provide insight into the turbulent nature of Bitcoin market efficiency over the past years, but cannot definitively measure the magnitude of the change due to the limitations in efficiency analysis. The most meaningful result of this study, however, is the statistically significant short-horizon price predictability that existed over the duration of the study, which has implications for Bitcoin market efficiency as well as for continued research in short-horizon Bitcoin price forecasting models.

Table of Contents

I. Introduction

Bitcoins are decentralized digital values that since inception have been increasingly used as a form of currency. Due to the nature of Bitcoins, they have proven an effective means of storing and transferring value – allowing any two parties to transfer them almost instantly, definitely, pseudo-anonymously, and without charge. This is possible because Bitcoin is a peer to peer network, where transactions do not involve intermediaries. In a later section, I provide more information and resources regarding Bitcoins and how they operate (Qkos 2014).

In today's complex global economy, the efficient transfer of value is of huge importance. As more and more business is conducted online and over large distances, it is only becoming more crucial for monetary transactions to be fast, cheap, and of course, secure. Bitcoin demonstrates strengthening potential to be a far more efficient method than the current best method of transferring money given its speed, reliability, cost, and ease of use (Qkos 2014). However, this paper does not detail all the appealing aspects of using Bitcoin as a currency.

This paper also does not seek to explain why one would invest in Bitcoin or argue what the 'true' or 'fundamental' value of Bitcoin is, but rather it explores the efficiency of the Bitcoin market over time. To do so, I analyze price return predictability and liquidity as measures of market efficiency over four subintervals of time from July 01, 2011 to May 23, 2013, as detailed in section III. Return predictability can be measured a number of ways. Borrowing from the theoretical framework of Chordia et al. (2008), I focus on two sources of predictability of future returns in this paper: the predictability of order flow imbalances and predictability of short-horizon past returns. I then analyze the trade volume and standard deviation of Bitcoin price returns over time to determine changes in market liquidity, and finally, relate these measures of market efficiency to the holding period return of bitcoin (percent change in Bitcoin price over the time period).

The results showed that during the time frame under study, return predictability existed at significant levels. The majority of the variables used to predict short-horizon price movements in this analysis were significant above the 99% significance level with R-Square values between 0.02 and 0.07, suggesting the ability to use past data to forecast price movements in the following five-minutes (See Table 4 and Table 5). I further discuss these findings in section IV.

Bitcoins

The website https://en.bitcoin.it/wiki/Research provides a list of research papers regarding Bitcoin; however most revolve around the legitimacy of, issues with, or specific aspects of Bitcoin. While there is online discussion and some articles published on Bitcoin markets and arbitrage, there is not research published about it. Below, I draw on my own knowledge from having closely followed Bitcoins over the past few years to briefly outline some relevant aspects of Bitcoins. However, the best reference for how Bitcoin works is the paper published in 2009 by someone under the pseudonym Satoshi Nakamoto, titled “Bitcoin: A Peer-to-Peer Electronic Cash System,” which concisely explains what Bitcoin is and how it works. It is written by the founder (or founding group) of Bitcoin. Online discussion, exchanges, and the website www.Blockchain.org are also great resources for understanding how Bitcoin works.

The Bitcoin network has existed for approximately four years, having continuously grown in recognition, acceptance, and value since inception. Since its creation, there have been many exchanges created in order to facilitate the trading between Bitcoin and numerous fiat currencies. MtGox.com was the largest exchange by daily volume until mid-2013 when BTCChina overtook them. This was a paradigm shift, as BTCChina lowered their trading fees to zero and quickly thereafter, trade volume on the BTCChina exchange grew exponentially while the price of Bitcoin soared. Because Bitcoin is so new, the market has experienced turbulent times. Mtgox is now defunct. Its user base grew so quickly that the security infrastructure was not developed enough to handle the millions of dollars that passed through the exchange. Given these rapid structural shifts in the market and the rising awareness and interest, the market has experienced large fluctuations in price and volatility, making efficiency over time fluctuate and difficult to measure (Qkos 2014, Mtgox 2013).

The motivation behind this study lies in the fact that market efficiency is important for the utilization of Bitcoin as an effective currency. Bitcoin needs to be relatively efficient and stable so that traders can easily enter and exit the positions. Further, for Bitcoin to be globally accepted, the value of a Bitcoin needs to be similar enough, if not the same, everywhere, so that a transaction in Bitcoin will provide roughly equal value to the involved parties. To see this intuitively, imagine I want to send $1 to someone in China right now. The receiving party should be able to immediately convert his $1 worth of Bitcoin into $1 worth of any currency. This is possible in an efficient market, but in an inefficient market, the receiver may not be able to trade the $1 of

Bitcoin into the same value of another currency. This would be due to inefficiencies in the market.

The next section reviews some of the existing literature regarding analyzing market efficiency. Given Bitcoins' uniqueness and the fact that they are so new and have only recently gained large-scale publicity support, there is very little published research regarding them, and even less that analyzes Bitcoin markets. As a result, I draw on concepts and studies of market efficiency pertaining to other financial markets in order to analyze the efficiency of the Bitcoin market. Given the extent of research in this field, the literature review does not seek to be a comprehensive review of this field, but rather seeks to introduce and explain some of the theoretical concepts and methods of analysis used in this paper. The paper then continues to explain my data and methodology before presenting the results and concluding with a discussion of the implications of the findings.

II. Literature Review

Market Efficiency

The concept of market efficiency has really only been considered since 1969 when Eugene Fama introduced the term ‘efficient market,’ which was defined as “a market that adjusts rapidly to new information” (Fama et al. 1969). This became known as the “Efficient Market Hypothesis” and has been a recurring topic in numerous studies. It was realized, however, that there are more elements of an efficient market than just the ability to adjust rapidly to new information. In 1991, Fama provides a more modern version of the “Efficient Market Hypothesis,” claiming that “asset prices in an efficient market ‘fully reflect all available information’” (Saari 1977 via Fama 1991). The result of an efficient market is then that asset prices are always at levels consistent with ‘fundamentals’.

There have since been numerous studies and papers published regarding the Efficient Market Hypothesis and many varying versions of the hypothesis have been developed over the years. However, the general notion of efficient markets “emphasizes a lack of return predictability as the criterion for efficiency” (Chordia et al 2008). Therefore, efficiency can be measured by the relative availability or predictable outsized returns. A perfectly efficient market would not have any return predictability, and therefore investors could not expect to beat the market. For a market to become more efficient over time, it must demonstrate less return predictability over that time. Another method for measuring efficiency, as discussed further in the following section, is analyzing the market liquidity over time.

Previous work (Chordia and Subrahmanyam 2004) suggests that first order autocorrelations of returns can also effectively predict price movements when there are inefficiencies in the market. From this, the degree of return predictability can be used as a measure of market efficiency. They find that liquidity is inversely correlated with return predictability, suggesting that when markets are more liquid, they tend to have lower return predictability which can be used to gauge market efficiency (Chordia and Subrahmanyam 2004).

Liquidity

There are numerous ways to measure liquidity. One common way is to calculate the effective bid ask spreads over continuous time intervals. However, because historical market depth data is not available, I settled for two less detailed but still insightful measures of liquidity: transaction volume and return volatility (Sarr and Lybek 2002).

According to a 2002 study by the IMF prepared by Abdourahmane Sarr and Tonny Lybek, there are a number of indicators that can be used to analyze liquidity in financial markets. Different measures gauge different aspects of market liquidity, "namely tightness (costs), immediacy, depth, breadth, and resiliency." They continue to say that, "liquid markets are generally perceived as desirable because of the multiple benefits they offer, including improved allocation and information efficiency." Further, they "render financial assets more attractive to investors, who can transact them more easily" (Sarr and Lybek 2002). According to the paper, liquid assets are characterized by the following: small transaction costs, easy trading and timely settlement, and large trades have only a small impact on the market price. However, they quote another paper as saying that there “is no single unambiguous, theoretically correct or universally accepted definition of liquidity” (Baker 1996 via Sarr and Lybek 2002).

Chordia, Roll, and Subrahmanyam (2008) find that more liquid markets should exhibit less pronounced return predictability and vice versa. They further suggest that market efficiency can be measured by return predictability and measures of liquidity. Through their research and meta-analysis of other studies, they are able to conclude that greater return predictability is negatively correlated with market efficiency, while efficiency is positively correlated with liquidity. A further conclusion was that, “In an efficient market, return predictability from past information should be short-lived and minimal,” suggesting that returns would only likely be predictable short amounts of time into the future (Chordia et al 2002, 2008).

Given the availability of Bitcoin trade data, we are limited to measuring liquidity via measures of transaction volume (both number of trades and value of trades) and via the volatility of returns. This will be further explained in the following section.

III. Data and Methodology

Return predictability can be measured a number of ways. Borrowing from the theoretical framework of Chordia et al. (2008), I focus on two sources of predictability of future returns in this paper: the future return predictability of order flow imbalances and of short-horizon past returns. I then relate these to trade volume and price volatility over time as well as to Bitcoin price returns over time to establish correlations between return predictability and market liquidity.

Data

The data used in this analysis comes from a public online publishing of all historical Bitcoin to fiat currency transactions that occurred between July 1, 2011 and May 23, 2013 on Mtgox.com. During this time period, Mtgox.com was the largest exchange for Bitcoins globally and was the generally accepted and referenced source for price data. Before filing for bankruptcy on February 28, 2014, Mtgox offered streaming full trade data via their API and historical data of the trailing 2,000 trades. In order to perform the price predictability analysis, however, I need complete trade data for the entire time interval under study. Fortunately, the operators of Mtgox created a full archive of historical trade data on Google's BigQuery service which is accessible to the public (Dataset: [mtgox] 2013). This was published on May 23, 2013, and has not been updated since. All historical trades were included, however some of the data was incomplete prior to July 2011, and as such, I've limited my analysis of market efficiency to the time period of July 01, 2011 to May 23, 2013. While this is just under 23 months of data, it includes some of the largest price movements and developments in the history of Bitcoin. See Chart 1 for a historical price chart of Bitcoin.

The full trade data for this time interval includes 4,983,238 trades, and contains the variables described in Table 1. Given the size of the dataset, I had to query each month of data independently. I restricted my returned variables only to Amount, Price, Date, and Type, as given these, everything else required for the analysis can be calculated as outlined in the following sections. While Mtgox offered users the ability to trade in numerous fiat currencies, their system had only one internal order book in which all orders placed were exchanged regardless of currency. The system automatically converted the fiat currencies so that Bitcoin orders placed in one currency could be matched and transacted with orders placed in another currency (Mtgox 2014). Given this, I specified that the query included all transactions and returned the data in USD. The datasets were then stored as tables on Google Cloud Storage services before being finally downloaded locally as monthly *.CSV* files which could then be imported into Matlab for analysis.

To gain insight into how Bitcoin market efficiency has changed over time, I separate the time under focus into four distinct time periods as seen in Table 2. Note that the intervals are not all the same length of time, as interval 1H2013* only includes 41,066 five-minute subintervals.

Method of Analysis

As previously discussed, I measure market efficiency by the predictability of short-horizon returns and by measurements of liquidity. I focus on two methods of short-horizon return predictability: 1) Using order flow imbalances to predict future price movements and 2) Using short-horizon past returns to predict future returns. I then analyze the trade volume and standard deviation of returns over time to determine changes in market liquidity; and finally, relate these measures of market efficiency to the holding period return of bitcoin (percent change in Bitcoin price over the time period).

Short-horizon return predictability from order flow imbalances

I was introduced to the concept of predicting short-horizon future price movements using order imbalance ratios from Chordia et al. (2002), who explained the potential value of analyzing predictability from lagged order flow imbalances. Borrowing from their conceptual framework, I calculate the order imbalance for the Bitcoin market for small time intervals so that I can then use an Ordinary Least Squares regression to regress the Bitcoin price return in interval *t* on the order imbalance from interval *t-1* to determine the predictability of short-horizon future returns. Predictability of returns from order imbalances using daily intervals is close to zero, thus smaller intervals had to be used. Similarly to the test performed by Chordia et al. (2008), I chose to divide the data into five-minute intervals for two reasons. First, in shorter than five-minute intervals, non-trading becomes an issue. Second, with longer intervals, short-lived inefficiencies may not be as evident and, as Chordia et al. have shown, “a predictive relation between order imbalances and future returns is unlikely to last very long” (Chordia et al. 2008).

I calculate two versions of order imbalance for each five-minute interval:

1) Raw Order Imbalance (OIB): This is defined as the number of buyer-initiated less the number of seller-initiated trades divided by the total number of trades. (#Bid - #Ask)/(#Bid + #Ask).
2) Value-Weighted Order Imbalance (OIBW): This is defined as the number of dollars paid by buyer-initiated trades (#Bid$) less the dollars received by seller-initiated trades (#Ask$) divided by the total dollars traded (#Bid$+#Ask$).

To do so, I devised original code in Matlab that takes as inputs the Mtgox trade data and returns the OIB and OIBW calculations for each five-minute interval over the entire dataset. Because of the size of the dataset, I ran the code over the data broken into month long intervals.

As seen in the annotated code in Appendix A, the OIB function calculates the number of buyer initiated transactions (#Bids) and seller initiated transactions (#Asks) placed during each five minute interval and then calculates the difference between them before finally finding the OIB by dividing the difference by the sum of #Bids and #Asks.

Similarly, the OIBW function calculates the dollar value of the OIB by summing the product of Price and Quantity for each transaction in each interval and stores the #Bid$ and #Ask$ separately so that it then computes the difference between them. It then computes the OIBW by dividing the difference by the sum of #Bid$ and #Ask$.

Return Data Calculation

Another data point needed for the price predictability of the OIB and OIBW calculations is the Bitcoin price return over the five-minute intervals. As Chordia et al.

point out, there are several methods that can be used in computing such returns. Returns can be computed from mid-point price quotes, weighted interval transaction prices, or open or close transaction prices during each interval. In this study, I compute returns on each five-minute interval two ways:

1) Closing Price Return (CPR): This is calculated by dividing the final transaction price from interval *t* by the final transaction price from interval *t-1* and then subtracting 1 to generate a percent change.
2) Average Price Return (APR): This is calculated by dividing the average transaction price from interval *t* by the average transaction price from interval *t-1* and then subtracting 1 to generate a percent change.

The details of these calculations can be seen in the annotated code in Appendix A.

In my analysis, I regress both CPR and APR on the two variables OIB and OIBW each lagged by one time interval in order to find the predictability of OIB and OIBW calculations on the returns in the following time interval. The relevant OLS Regression output statistics can be seen in Table 4. Note that although t-statistics of the intercepts often imply low significance, this is negligible because the magnitude of the intercept is so small it has little effect on the predictability of the independent variables. These regression outputs will be further discussed in section IV.

Short-horizon return predictability from return auto-correlative regressions

The second test I perform to analyze short-horizon future return predictability is an auto-correlative regression which regresses the return in interval *t* on the return in the previous interval, *t-1*. I perform the test on both APR and CPR for the four distinguished time intervals under study. The resulting regression statistics can be found in Table 5. I find extreme significance of the predictability of both APR and CPR lagged returns in all four time periods under study. Further, the R-squared values greater than zero imply inefficiency in the market, as this shows some of the variation in returns is predictable by the return in the previous five-minute interval. I find R-squared values from all four time periods to be at minimum 0.029. Though seemingly small, these values are relatively large given the nature of the study. In a fully efficient market, a model of return predictability would theoretically have an R-squared of zero given that one should not be able to predict any of the variance in the returns. Further, relative to the R-squared of similar models on the United States stock market, the R-squared of these regressions are much larger, implying greater inefficiency given the higher return predictability (Chordia et al. 2008, Saari 1977).

Liquidity

One method of measuring liquidity as outlined by Chordia et al. 2008 looks at the effective market spread over time. However, given the limitations of available historical data in the Bitcoin market, we must limit our analysis of liquidity to an investigation of transaction volume and volatility. I use two measures of liquidity in this analysis: 1) Average transaction volume per five-minute subinterval in each of the four time periods and 2) Standard deviation of five-minute interval returns (on both APR and CPR). From this, the volume and standard deviation give insight into how liquid the market is in each of the four time periods. Similar to the findings in Chordia et al.'s 2004 study on the

United States stock market, I find that liquidity as measured by volatility is inversely correlated with many of the predictability measurements. This can be seen in Table 7 through the correlation of return volatility with the t-statistics and R-squared values of the predictive return regressions over the four time periods.

IV. Discussion of Results and Further Analysis

Now that I've calculated measures of price return predictability and liquidity in the Bitcoin market in each of the four time periods, I analyze the correlation of all of these measures with each other and with the holding period return of Bitcoin in each of the four time periods.

As seen in Table 6, there are fifteen variables which I correlate with each other. Table 3 lists the fifteen variables along with a description of each for reference. Note that I considered the absolute value of the t-statistics from the regressions in order to measure the magnitude of significance of the test and to allow for the degree of significance to be consistent in the correlation analysis.

The results are generally as expected and similar to the findings in the studies by Chordia et al. (2008). As they found in the US stock market, the results of this analysis show that in periods when the standard deviation of returns is higher, the significance and predictive power of both price return predictability tests generally decrease (Table 6). From 2H2011 through 2H2012, the standard deviation of returns decreases while the significance of the coefficient of the OIB variable in both CPR and APR regressions becomes much greater (as seen by the t-statistics). Further, the significance of the coefficient of return in both CPR and APR auto-regressions becomes much greater as well. Perhaps even of greater implications, the R-squared values of all four predictability regressions significantly increases in these two periods of lower volatility (1H2012 and 2H2012). The CPRAuto regression has an R-squared of 0.033 in 1H2011 and jumps to 0.0746 and 0.0779 in 1H2012 and 2H2012 respectively. This suggests that in times of lower volatility, one could predict a larger percent of the variance in price movement in the following five-minute interval using this predictive model.

The interval 1H2013* is characterized by greater standard deviation of returns than any of the previous three periods, with a five-minute interval return standard deviation of 0.64% using APR and 0.89% using CPR. Accompanied with this, however, Bitcoin prices increase 816% in the 1H2013 time period while the average number of trades per five-minute interval increase from 16 in 2H2012 to 47 in 1H2013* and the average value of those trades in each five-minute interval increases from ~$1,646 to ~22,800. The price predictability also falls greatly during this interval compared to the previous ones. During this time period, knowledge of Bitcoin began to grow and interest grew exponentially. As a result, many began to purchase, trade, and use Bitcoin leading to a significant increase in the number of trades. Interestingly, this led to high volatility but also low return predictability – which has been shown by Chodia et al. (2008) to have existed in the US stock market, but is contrary to the idea of periods of low volatility characterizing high market liquidity and low efficiency. High volatility, however, is not enough to show a market is not liquid. This volatility was likely the result of such a huge influx of Bitcoin traders and massive value increases.

It is interesting that the results suggest that in periods of lower volatility, return predictability is higher. The rapidly changing nature of the Bitcoin market over the time period made it difficult to relate trade volumes with anything but the holding period return, as both trade volumes and the price of Bitcoin increased substantially from the third period through the fourth.

Limitations and Further Studies

One major limitation in this study was that the historical data necessary for this analysis was limited to the time frame in which I focused my study. Further, this data was limited to only Mtgox trade data. This is acceptable, however, because I focused the study only on Mtgox prices and trades, and further, Mtgox was the largest exchange at the time. Measurements of liquidity were also limited in this study to only looking at volume (as measured by number of trades and the value of those trades in each interval), and volatility (as measured by the standard deviation of returns). As discussed previously in the Literature Review, there are numerous ways to measure liquidity. One way that I would consider doing so in further research would be to measure the effective bid/ask spread within the same time intervals. Lower spread levels have been shown to imply higher market liquidity, and in turn, greater market efficiency (Chordia, et al. 2008).

To further improve this study, I want to look at trade data since May 23, 2013 aggregated from all exchanges. However, there would be some limitations to this given liquidity premiums among the different exchanges. Another improvement of this research would be to utilize more robust trade data such as market depth to better study price predictability. Taking into account the size of the market depth could likely help explain some of the variation in price returns.

V. Conclusion

Bitcoin has seen turbulent times through its few years of existence. Going from being valued at pennies only four years ago to a peak of over $1,200 in November 2013, bitcoin has experienced large price fluctuations and value appreciation as more and more people have sought exposure to it (Nilz 2014). The results of this paper are exciting for a number of reasons. First, the correlations of our measures of efficiency and returns give insight into the efficiency of the Bitcoin market. The rapidly changing Bitcoin market, however, does not allow us to draw specific quantitative conclusions pertaining to the change in market efficiency over time. Further study would need to be performed to more effectively analyze the changing market efficiency.

Second, and most exciting, is that the results show that return predictability has existed throughout the entire interval under study. Given this, one could theoretically devise a trading strategy to take advantage of these inefficiencies by predicting short-horizon price movements. One would need to gather streaming data from at least one Bitcoin exchange in order to create a model to do so. An analysis of more recent data would need to be completed to determine the price return predictability of more recent returns. Given the rapidly evolving Bitcoin market, however, being able to predict prices yesterday does not ensure one could do so today. Pursuit of such studies seems to be a potentially worthwhile research topic moving forward.

VI. References

BTCChina. “Bitcoin exchange.” [online]. https://www.BTCChina.com. April 2014.

Chordia, Roll, Subrahmanyam. “Liquidity and Market Efficiency.” *Journal of Financial Economics* 87 (2008). Web. December 2013.

Chordia, Roll, Subrahmanyam. “Order imbalance, liquidity and market returns.” *Journal of Financial Economics* 65 (2002). Web. March 2014.

Chordia, Subrahmanyam. “Order imbalance and individual stock returns: Theory and evidence.” *Journal of Financial Economics* (2004). Web. March 2014.

Fama, E. “Efficient capital markets: a review of theory and empirical work.” *Journal of Finance* 25 (1970). Web. March 2014.

Fama, E. “Efficient Capital Markets: II.” *Journal of Finance* 46 (1991). Web. March 2014.

Fama, Fisher, Jensen, Roll. “The Adjustment of Stock Prices to New Information.” *International Economic Review* 10 (1969). Web. November 2013.

Mtgox. “Bitcoin exchange.” [online]. https://www.mtgox.com. December 2013.

Nakamoto, Satoshi. “Bitcoin: A Peer-to-Peer Electronic Cash System.” [online]. https://en.bitcoin.it/wiki/Research. April 2014.

Nilz. “Bitcoin Charts.” [online]. https://www.bitcoincharts.com. April 2014.

Nitrous. “New method to pull MtGox trade data.” [online forum]. https://bitcointalk.org/index.php?topic=218980.0. May 2013.

Qkos Services. “Blockchain” [online]. https://www.blockchain.info. April 2014.

Saari, Christopher Paul. “The Efficient Capital Market Hypothesis, Economic Theory and the Regulation of the Securities Industry.” *Stanford Law Review* (1977). Web. November 2013.

Sarr and Lybek. “Measuring Liquidity in Financial Markets.” *IMF Working Paper* WP/02/232 (2002). Web. April 2014.

Unknown Author. Dataset: [mt-gox:mtgox].trades. Google BigQuery; Online. March 2014.

VII. Tables and Charts

Table 1: List of data variables in Mtgox historical trade dataset, accessed via Google BigQuery.

Variable	***Description***
date	Date and time of trade
price	Price of Bitcoins in fiat currency per Bitcoin
amount	Number of Bitcoins of the order
price_currency	Specifies the fiat currency type (USD, CHY, EUR,etc…)
trade_type	Specifies either "bid" or "ask"
properties	Specifies either "market" or "limit"

Table 2: List of time period intervals. The left column assigns a name to the four intervals that are the focus of the study while the second column shows the time period of the interval. Note that interval 1H2013* has less trading days than the other intervals.

Interval	***Description***
2H2011	July 01, 2011 through December 31, 2011
1H2012	January 01, 2012 through June 31, 2012
2H2012	July 01, 2012 through December 31, 2012
1H2013*	January 01, 2013 through May 23, 2013

Table 3: List of variable names and descriptions of variables calculated to measure market efficiency. These include variables that are measures of return predictability as well as measures of market liquidity. The final variable, HPR BTC, is the Holding Period Return, which is the percent change in Bitcoin price over the time period.

	Variable	***Description***
Measures of Return Predictability	t-statCPRAuto	Absolute value of t-statistic of the coefficient of the correlative variable in CPRAutoRegression
	t-statAPRAuto	Absolute value of t-statistic of the coefficient of the correlative variable in APRAutoRegression
	rCPRAuto	R-squared of the CPRAutoRegression
	rAPRAuto	R-squared of the APRAutoRegression
	rCPR	R-squared of the CPR Regression on lagged OIB and OIBW
	rAPR	R-squared of the APR Regression on lagged OIB and OIBW
	t-statOIBCPR	Absolute value of t-statistic of OIB coefficient in CPR Regression
	t-statOIBWCPR	Absolute value of t-statistic of OIBW coefficient in CRP Regression
	t-statOIBAPR	Absolute value of t-statistic of OIB coefficient in APR Regression
	t-statOIBWAPR	Absolute value of t-statistic of OIBW coefficient in APR Regression
Measures of Liquidity	Avg#Trades	Average number of trades per five-minute interval
	Avg$VT	Average value of trades per five-minute interval
	SDAPR	Standard Deviation of APR
	SDCPR	Standard Deviation of CPR
	HPR BTC	Holding Period Return in time interval

Table 4: OIB and OIBW Regression Statistics. This table shows the relevant regression statistics of the first measures of return predictability across the four time periods. The columns hold the regression statistics for the regressions of both Closing Price Returns and Average Price Returns on the two independent variables: OIB and OIBW (both of which are lagged by one time interval from the returns, as discussed in the text) in all four time intervals. The OIB in interval *t*-1 is a stronger predictor of earnings in interval *t* than is OIBW, as suggested by the t-statistics of the coefficient.

	2H2011		1H2012		2H2012		1H2013*	
	Lagged on CPR	**Lagged on APR**	**Lagged on CPR**	**Lagged on APR**	**Lagged on CPR**	**Lagged on APR**	**Lagged on CPR**	**Lagged on APR**
Intercept coeff	-2.79E-05	-1.33E-05	6.19E-06	5.48E-06	8.02E-05	9.84E-05	7.08E-05	-1.16E-05
SE	3.34E-05	2.61E-05	2.05E-05	1.59E-05	1.98E-05	1.54E-05	4.55E-05	3.26E-05
t-statistic	-0.835	-0.509	0.302	0.344	4.041	6.384	1.556	-0.356
OIB coeff	-1.04E-03	-6.18E-04	-8.83E-04	-6.03E-04	-1.25E-03	-1.06E-03	-6.62E-04	-5.34E-05
SE	9.23E-05	7.22E-05	5.73E-05	4.45E-05	5.40E-05	4.19E-05	1.35E-04	9.71E-05
t-statistic	-11.248	-8.562	-15.416	-13.564	-23.254	-25.258	-4.892	-0.550
OIBW coeff	3.98E-04	3.78E-04	1.21E-04	3.57E-05	2.10E-04	3.66E-06	4.78E-04	5.17E-04
SE	7.83E-05	6.12E-05	4.63E-05	3.59E-05	4.25E-05	3.30E-05	9.74E-05	6.99E-05
t-statistic	5.087	6.166	2.610	0.995	4.942	0.111	4.912	7.405
R-Squared	0.003386	0.001426	0.010006	0.009121	0.019875	0.032661	0.000671	0.002656
Observations	52989	52989	52409	52409	52985	52985	41065	41065

Table 5: CPR and APR Auto-correlative Regression Statistics. This table shows the relevant regression statistics of the second measure of return predictability across the four time periods. Each of the four intervals includes regression statistics for both the Closing Price Return Auto-correlative regression as well as the Average Price Return Auto-correlative regression. Closing Price Returns in interval *t* -1 seem to be a better predictor of returns in interval *t* than do Average Price Returns, as seen by the generally greater absolute value of the t-statistic and higher regression R-squared values across the four intervals.

	2H2011		1H2012		2H2012		1H2013*	
	CPRAuto	**APRAuto**	**CPRAuto**	**APRAuto**	**CPRAuto**	**APRAuto**	**CPRAuto**	**APRAuto**
Intercept coeff	6.84E-06	-5.21E-06	2.28E-05	1.48E-05	2.98E-05	2.24E-05	1.10E-04	6.02E-05
SE	3.27E-05	2.60E-05	1.98E-05	1.59E-05	1.86E-05	1.50E-05	4.34E-05	3.10E-05
t-statistic	0.209	-0.201	1.150	0.931	1.599	1.492	2.534	1.941
Auto-correlative coeff	-0.183	0.023	-0.273	-0.102	-0.279	-0.149	-0.172	0.194
SE	0.004	0.004	0.004	0.004	0.004	0.004	0.005	0.005
t-statistic	-42.768	5.237	-65.001	-23.365	-66.924	-34.578	-35.392	40.099
R-Squared	0.033	0.001	0.075	0.010	0.078	0.022	0.030	0.038
Observations	52988	52988	52408	52408	52984	52984	41064	41064

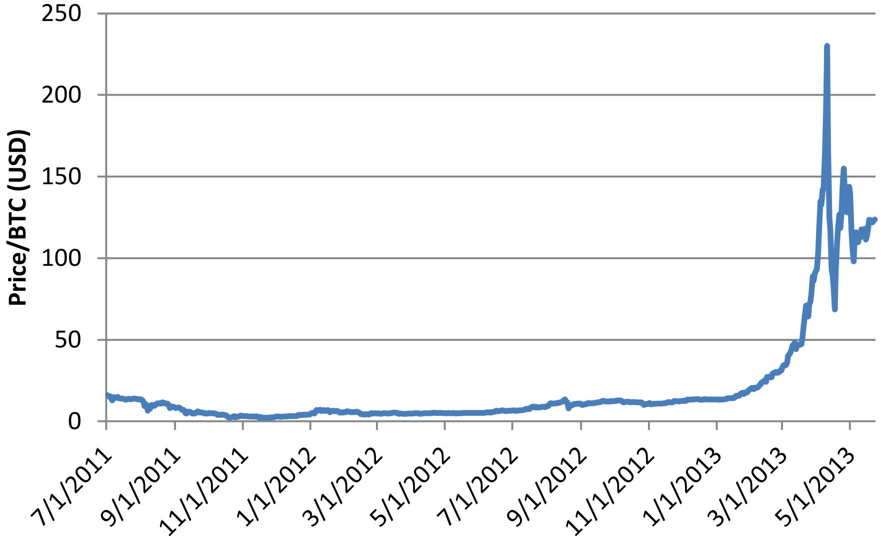
Table 6: Values of calculated variables of return predictability and liquidity measures as well as the holding period return across each of the four intervals under study. The description of each of these variables can be found in Table 3. The correlations of each of these variables with each other are found in Table 7.

Variable	***Time Interval***			
	2H2011	1H2012	2H2012	1H2013*
t-statCPRAuto	-42.768	-65.001	-66.924	-35.392
t-statAPRAuto	5.237	-23.365	-34.578	40.099
rCPRAuto	0.033	0.075	0.078	0.030
rAPRAuto	0.001	0.010	0.022	0.038
HPR BTC	-0.708	0.423	1.019	8.164
rCPR	0.003	0.010	0.020	0.001
rAPR	0.001	0.009	0.033	0.003
t-statOIBCPR	-11.248	-15.416	-23.254	-4.892
t-statOIBWCPR	5.087	2.610	4.942	4.912
t-statOIBAPR	-8.562	-13.564	-25.258	-0.550
t-statOIBWAPR	6.166	0.995	0.111	7.405
Avg#Trades	20	22	16	47
Avg$VT	997	1247	1647	22800
SDAPR	0.00598	0.00366	0.00349	0.00641
SDCPR	0.00765	0.00472	0.00447	0.00892

Table 7: Paired correlations of measures of efficiency and liquidity as well as holding period return each with one another. See Table 3 for variable descriptions and Table 6 for variable values in each of the four time periods under study. This analysis is limited by the number of observations of each variable. Because I had only four distinct time intervals under study, the correlations are measured only on four observations. Future study could include more time intervals of smaller time length so that correlations in the change over time could be better established.

Variable	*t-statCPRAuto*	*-statAPRAuto*	*rCPRAuto*	*rAPRAuto*	*HPR BTC*	*rCPR*	*rAPR*	*t-statOIBCPR*	*-statOIBWCPR*	*t-statOIBAPR*	*-statOIBWAPR*	*Avg#Trades*	*Avg$VT*	*SDAPR*	*SDCPR*
t-statCPRAuto	1.000														
t-statAPRAuto	0.071	1.000													
rCPRAuto	0.991	0.197	1.000												
rAPRAuto	-0.268	0.941	-0.145	1.000											
HPR BTC	-0.591	0.749	-0.480	0.917	1.000										
rCPR	0.903	0.227	0.902	-0.073	-0.465	1.000									
rAPR	0.761	0.406	0.779	0.153	-0.252	0.961	1.000								
t-statOIBCPR	0.913	0.012	0.884	-0.284	-0.643	0.977	0.896	1.000							
t-statOIBWCPR	-0.532	0.042	-0.547	0.246	0.251	-0.133	0.083	-0.149	1.000						
t-statOIBAPR	0.896	0.041	0.869	-0.248	-0.614	0.982	0.915	0.999	-0.104	1.000					
t-statOIBWAPR	-0.997	-0.134	-0.996	0.205	0.544	-0.924	-0.799	-0.920	0.499	-0.906	1.000				
Avg#Trades	-0.750	0.523	-0.659	0.748	0.950	-0.711	-0.542	-0.846	0.174	-0.828	0.721	1.000			
Avg$VT	-0.707	0.637	-0.608	0.848	0.988	-0.590	-0.382	-0.747	0.296	-0.720	0.666	0.980	1.000		
SDAPR	-0.997	-0.141	-0.998	0.200	0.529	-0.901	-0.768	-0.895	0.548	-0.879	0.998	0.698	0.651	1.000	
SDCPR	-0.999	-0.026	-0.984	0.311	0.628	-0.899	-0.750	-0.918	0.527	-0.900	0.994	0.778	0.738	0.993	1.000

Chart 1: Historical Bitcoin price graph in United States dollars. Data is from Mtgox and covers the time interval under focus in this study (July 01, 2011 through May 23, 2013). As seen, the price rose significantly between January and May of 2013.



VIII. Appendix

This appendix includes the original Matlab code written for this project. There are four distinct files, three of which manipulate the data and calculate the needed variables as detailed in the paper, and one which executes the first three along with other necessary steps for each of the inputted data sets. Percent signs (%) designate comments which explain what the code is doing.

1) OIB.m

```
%This function finds and returns ORDER IMBALANCES in each interval
%Typecount returns 4 columns: #bids,#asks,(#b-#a), OIB
function  TypeCount = OIB(Date, IntervalTimeSize, Type)
%Called in OIB Test
%   Counts the number of bids and asks in each time interval by saving
the
%   number of both buy and sell initiated trades in each interval

n = IntervalTimeSize; %simplify IntervalTime
A = Type; %simplify Type

%initial conditions:

%number of 5 minute intervals in the data
numIntervals = (Date(length(Date)) - Date(1))/n;

%Declare array of Integers to store results where
% Col1 = #bids and Col2 = #asks in the interval (represented by the
row)
TypeCount = zeros(numIntervals,4); %this is by default a double, may
need to make an int

i = 1; %initialize i for the following loop

for k=1:numIntervals
    while (Date(i)-Date(1)) < n*k
        if A(i) =='b'
            TypeCount(k,1) = TypeCount(k,1)+1;
        else TypeCount(k,2) = TypeCount(k,2)+1;
        end
        i=i+1;
    end
end

%Now let Col3 of TypeCount be the imbalance (#Buys-#Asks) and let Col4
be
%the OIB calculation (#Buys-#Asks)/(#Buys+#Asks)
```

```
for c=1:length(TypeCount)   %for each row, make col3 = col1-col2
    TypeCount(c,3) = (TypeCount(c,1)-TypeCount(c,2));

    %Also make col4 = Col3/(col1+col2) if (col1+col2)>0, otherwise 0
    if (TypeCount(c,1)+TypeCount(c,2))> 0
        TypeCount(c,4) =
(TypeCount(c,3)/(TypeCount(c,1)+TypeCount(c,2)));
    else
        TypeCount(c,4) = 0;
    end

end
end


%This function returns the average price in each interval for SHORT
HORIZON
%RETURN AUTOCORRELATION TEST
function AvgPrice = AP(Date, IntervalTimeSize, Price)

%simplify IntervalTimeSize
n = IntervalTimeSize;

%number of intervals
numIntervals = (Date(length(Date)) - Date(1))/n;

%The resulting matrix
AvgPrice = zeros(numIntervals,2);

i = 1; %initialize i for the following loop

for k=1:numIntervals

    %declare a variable x to hold the average
    x = 0.00;
    %declare variable to represent the number of observations being
    %averaged
    t = 0;

    while (Date(i)-Date(1)) < n*k
        x = x + Price(i);
        t = t + 1;
        i=i+1;
    end

    %calculate average price
    AvgPrice(k,1) = x/t;

end
end
```

2) OIBW.m

```
%This function finds and returns Dollar Weighted ORDER IMBALANCES in
each interval
%Typecount returns 4 columns: #bids, #asks, difference, and OIBW
function  TypeCount = OIBW(Date, IntervalTimeSize, Type, Price, Amount)
%Called in OIB Test
%   Counts the number of bids and asks in each time interval by saving
the
%   number of both buy and sell initiated trades in each interval

n = IntervalTimeSize; %simplify IntervalTime
A = Type; %simplify Type
Price = Price/100000; %Bring price into correct dollar size
Amount = Amount/100000000; %Bring amount into correct number

%initial conditions:

%number of 5 minute intervals in the data
numIntervals = (Date(length(Date)) - Date(1))/n;

%Declare array of Integers to store results where
% Col1 = #bids and Col2 = #asks in the interval (represented by the
row)
TypeCount = zeros(numIntervals,4); %this is by default a double, may
need to make an int

i = 1; %initialize i for the following loop

for k=1:numIntervals

    while (Date(i)-Date(1)) < n*k
        if A(i) == 'b'

            TypeCount(k,1) = TypeCount(k,1)+ (Price(i)*Amount(i));
        else TypeCount(k,2) = TypeCount(k,2)+ (Price(i)*Amount(i));
        end
        i=i+1;
    end
end

%Now let Col3 of TypeCount be the imbalance (#Buys-#Asks) and let Col4
be
%the OIB calculation (#Buys-#Asks)/(#Buys+#Asks)
for c=1:length(TypeCount)   %for each row, make col3 = col1-col2
    TypeCount(c,3) = (TypeCount(c,1)-TypeCount(c,2));

    %Also make col4 = Col3/(col1+col2) if (col1+col2)>0, otherwise 0
    if (TypeCount(c,1)+TypeCount(c,2))> 0
        TypeCount(c,4) =
(TypeCount(c,3)/(TypeCount(c,1)+TypeCount(c,2)));
    else
        TypeCount(c,4) = 0;
```

```
    end
end
end
```

3) AP.m

```
%This function returns the average price in each interval for SHORT
HORIZON
%RETURN AUTOCORRELATION TEST
%AP returns 3 columns: Average price, Closing Price, Interval Start
Date
function AvgPrice = AP(Date, IntervalTimeSize, Price)

%simplify IntervalTimeSize
n = IntervalTimeSize;

%Bring price into correct Dollar size
Price = Price/100000;

%number of intervals
numIntervals = (Date(length(Date)) - Date(1))/n;

%The resulting matrix
AvgPrice = zeros(numIntervals,3);


i = 1; %initialize i for the following loop

for k=1:numIntervals

    %declare a variable x to hold the average
    x = 0.00;
    %declare variable y to hold the closing price in each interval
    y = 0.00;
    %declare variable to represent the number of observations being
    %averaged
    t = 0;

    %set 3rd col to the start date of the interval
    AvgPrice(k,3) = Date(1)+ n*(k-1);

    while (Date(i)-Date(1)) < n*k
        %sum the prices and the number of them summed
        x = x + Price(i);
        t = t + 1;
        %let y be the price of the last trade in the interval
        y = Price(i);

        i=i+1;
    end
```

```
    %if statement to catch error when no transactions occur in the
    %interval
    if t > 0
        %calculate average price
        AvgPrice(k,1) = x/t;
        %assign the closing price
        AvgPrice(k,2) = y;
    else
        AvgPrice(k,1) = AvgPrice(k-1,1);
        AvgPrice(k,2) = AvgPrice(k-1,2);
    end
end
end
```

4) RUNITALL.m

```
%This file runs operations needed in sequential order for each of the
%inputted monthly raw trade data sets.
A = cell2mat(Type);
load('1st_interval.mat', 'n');
AP = AP(Date,n,Price);
OIB = OIB(Date,n,A);
OIBW = OIBW(Date,n,A,Price,Amount);
OUTPUT = horzcat(AP,OIB,OIBW);
```