An Exploratory Statistical Analysis of NASDAQ-Provided Trade Data

A Thesis Presented

by

Michael Foley

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fullfillment of the Requirements
for the Degree of Master of Science
Specializing in Mathematics

October, 2014

Accepted by the Faculty of the Graduate College, The University of Vermont, in partial fulfillment of the requirements for the degree of Master of Science, specializing in Mathematics.

Thesis Examination Committee:

Advisor
Peter Sheridan Dodds, Ph.D.

Chris Danforth, Ph.D.

Chairperson
Bill Gibson, Ph.D.

Dean, Graduate College
Cynthia J. Forehand, Ph.D.

Date: August 18, 2014

Abstract

Since Benoit Mandelbrot's discovery of the fractal nature of financial price series, the quantitative analysis of financial markets has been an area of increasing interest for scientists, traders, and regulators. Further, major technological advances over this time have facilitated not only financial innovations, but also the computational ability to analyze and model markets.

The stylized facts are qualitative statistical signatures of financial market data that hold true across different stocks and over many different timescales. In pursuit of a mechanistic understanding of markets, we look to accurately quantify such statistics. With this quantification, we can test computational market models against the stylized facts and run controlled experiments. This requires both discovery of new stylized facts, and a persistent testing of old stylized facts on new data.

Using NASDAQ provided data covering the years 2008-2009, we analyze the trades of 120 stocks. An analysis of the stylized facts guides our exploration of the data, where our results corroborate other findings in the existing body of literature. In addition, we search for statistical indicators of market instability in our data sets. We find promising areas for further study, and obtain three key results.

Throughout our sample data, high frequency trading plays a larger role in rapid price changes of all sizes than would be randomly expected, but plays a smaller role than usual during rapid price changes of large magnitude ($> .0.5\%\ \Delta p$). Our analysis also yields further evidence of the long term persistence in the autocorrelations of signed order flow, as well as evidence of long range dependence in price returns.

Acknowledgements

I have been incredibly lucky. Thank you to my family, and especially my mom for putting me in a position to pursue my goals in life. And thank you to all my friends. I would also like to thank everyone in StoryLab. You are all terrific human beings, and I am grateful I have had the chance to work with you these past two years.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction and Literature Review

1.1 Introduction

Financial markets are the cornerstone insitutions of a modern economy. By channeling savings into investment, markets allow firms to accumulate capital, facilitate the transfer of risk and liquidity, as well as expedite and promote international trade. While financial markets maintain a complex relationship with the real economy, they can also exacerbate economic instability and shake consumer confidence. More recently, due to a flurry of financial innovations, financial markets have created instabilities within themselves which can end up cascading outwards and negatively impacting the real economy (Brady 1988, Kirilenko 2010).

In light of this, researchers in quantitative finance and econophysics have attempted to identify key patterns in financial markets, both at an aggregate level and at the event level, in order to better understand the social and technical mechanisms which drive markets. Somewhat surprisingly, a handful of key statistical regularities have emerged from this research, and these results have become what we now call the stylized facts.

A second area of analysis which has come to light in recent years is the impact of high frequency traders (HFT) on market behavior. HFT, or algorithmic traders, are computing platforms that trade very quickly, immediately around the spread, and attempt to minize risk as much as possible. That is, they do not generally attempt to move price, and their net positions at the end of each trading day are close to 0. Due to the abilities of these traders to place orders much more quickly than humans, the market impacts of HFT have been brought to question. Recent studies have shown that the minimum time for a human decision is on the order of one second, while HFT can place trades in under 25 milliseconds (Johnson et al. 2013). Some have argued that because HFT are so active, they decrease spread sizes and increase liquidity in markets, both indicators of a healthy market (Brogaard et al. 2013, Hendershott et al. 2011). Others have argued that because neither human traders nor regulators can follow high frequency trading in real time, it leaves markets more vulnerable to market fractures, or large changes in price that happen in minutes, seconds, or even milliseconds (Johnson et al. 2013, Wah 2013).

The Flash Crash of May 6, 2010 provides one example of a crash exacerbated by HFT. The crash began on the E-Mini S&P 500, a futures index, where the price dropped 3% in under 4 minutes. The effects quickly spilled over into equities markets, causing many HFT to halt trading, which caused liquidity to dry up, and prices to drop to drop even further. Prices eventually recovered, almost equally as quickly as they had fallen, after regulators paused trading on the E-Mini for 5 seconds. While the primary causes of the Flash Crash have been debated, there is no doubt among experts that the actions of HFT had a significant impact on the events of that day (Kirilenko 2010).

1.1.1 Literature Review

The two broad areas of study in the field of quantiative finance are empirical analysis and computational modeling. While we carry out a purely empirical analysis in this paper, we also emphasize the necessity of computational modeling as a tool in the pursuit of a mechanistic understanding of financial markets.

Over 50 years ago, Mandelbrot first identified two major stylized facts in price data which were robust across many different timescales (Mandelbrot 1963). These first two stylized facts were fat tails and clustering volatility. Fat tails describe the distribution of price returns, where the tails are "fat" because events of large magnitudes happen more often than a Gaussian distribution would allow. Clustering volatility refers to the slow decay of autocorrelations in the absolute value of price returns. Returns of similar magnitude tend to follow one another, but those returns may be positive or negative. While Mandelbrot's model of fat tails has been shown to overestimate the occurance of large returns (Lux 2009), his findings have qualitatively held across different securities and markets.

There are various ways to calculate fat tails, but the most parsimonious is with the kurtosis value of the return distribution. The kurtosis of the normal distribution is three. A distribution with a kurtosis value of greater than three is called leptokurtotic and implies that that both small and large price changes are more likely than a Gaussian distribution would predict. Conversely, moderate sized price changes occur less often in a leptokurtotic distribution. This is sometimes referred to as the distribution having a lack of shoulders.

Though many other stylized facts have been studied and described since then, controversy surrounds many of them. Thus, we focus on just one more stylized fact in price data:

the martingale property of price returns. This can be stated as

$$E[p_{t+1} \mid I_t] = p_t$$

where I_t is information at time t, and p_t is price. Thus, past returns do not yield predictive power for future returns. This martingale property manifests itself as the quick decay in the autocorrelation function of returns, and is the basis behind the efficient market hypothesis. This property is also the basis of the efficient market hypthesis, which asserts that the population of traders price expectations will be normally distributed around the "correct" market price, that is, the price reflected by the market.

The Hurst exponent is one way of measuring long-range dependence in a system. In other words, the Hurst exponent measures how the range of the data scales over the number of observations. We give the algorithm for calculating the Hurst in the Section 2.1.2. A Hurst exponent of 0.5 is a purely brownian motion process, whlie a Hurst exponent greater than 0.5 is an indication of positive correlations that persist for much longer than a true martingale process would allow.

Recently, due to its increasing availability, a new type of data has been analyzed for stylized facts: limit order book data. In order book and trade data, Toth and others have observed persistent autocorrelations in signed order flow (Toth et al. 2011). That is, buy orders tend to follow other buy orders, sell orders tend to follow sell orders, and these correlations persist over moderate time lags. Toth introduces two simple explanations for the persistence of signed order flow correlations. One possible explanation, which Toth finds strong evidence for, is order-splitting, where one trader wants to buy or sell a large amount of a security, and splits their order into many smaller orders placed closely together in time. Another explanation, which Toth finds less evidence for, is herding. Although

herding occasionally occurs in markets, especially during crisis events, it does not seem to be the leading factor causing the correlations of signed order flow.

Timestamp	Symbol	Shares	BuySell	Price	type
01-02-2008 09:31:51.924	BRE	48.0	1	40.31	HH
01-02-2008 09:36:08.129	BRE	3.0	1	40.05	NH
01-02-2008 09:36:38.054	BRE	10.0	1	39.83	NH
01-02-2008 09:40:19.737	BRE	7.0	1	40.02	HN
01-02-2008 09:40:20.016	BRE	80.0	-1	40.01	HH
01-02-2008 09:40:32.093	BRE	4.0	-1	40.0	NH
01-02-2008 09:41:36.017	BRE	4.0	-1	39.99	NH

Table 1.1: Example trade data of the stock BRE. A BuySell value of 1 indicates a buy market order executed against a sell limit order, while a value of -1 indicates a sell market order executed against a buy limit order. In the type column, H indicates an HFT, and N indicates non-HFT. The first character corresponds to whoever placed the market order.

While many have built upon Mandelbrots foundational work, Cont (2001) provides a comprehensive review of the current state of stylized facts research in financial price data. The stylized facts, described above, have been shown to be robust across the price time series of numerous markets and varying time intervals.

Together, this set of stylized facts provide empirical evidence to guide the validation testing of market models. According to Cont, "... these stylized facts are so constraining that it is not easy to exhibit even an (ad hoc) stochastic process which possesses the same set of properties and one has to go to great lengths to reproduce them with a model".

Recently, the focus in quantitative finance research has shifted towards the role of market microstructure in price movements. To fully understand financial markets, we not only need to understand how and why traders make the decisions they do, but we also need to understand the structure of the order book on which they trade (Darley et al. 2001). Computational modeling has offered some key insights into the ivmpacts of market microstructure. In particular, Farmer (2005) and Preis (2006) give examples of low intelli-

gence agent based models which are able to reproduce the stylized facts of price data, but not signed order flow. Meanwhile, LeBaron (2001) and Cont (2001) provide models with more emphasis on trader decision making which also reproduce the stylized facts.

From an empirical perspective, Easley (2012) has developed the Volume synchronized probability of informed trading (VPIN), a metric which attempts to measure informed vs. non-informed trading. Yeo (2009) offers two examples of liquidity measures which make use of the dispersion of orders in the book.

1.1.2 Data and Objectives

The analysis in this study was conducted on trade data supplied by NASDAQ from 120 hand-picked stocks. The trade data is at the tick resolution, covering all trades of the given stocks in the years 2008 and 2009. There are over 550 million trades over all stocks during this time. For each trade which was executed, the data contains a timestamp, a price, an indication of buy or sell (of the market order), and an indication of the trader type (high frequency or not) on both ends of the trade.

The objectives of this research are the following. In Section 2.1.1, we show how to statistically evaluate the clustering volatility, kurtosis, and decay in autocorrelations. In Section 2.1.2 we show how to calculate the Hurst exponent, and in Section 2.1.3 we show how to calculate signed order flow autocorrelations. We show our method of analysis of HFT in Section 2.2 in an attempt to identify patterns that correlate with market instability. Following the work of Johnson et. al. (2012), we study the effect of HFT on market fracture events, which we categorize as contributors to market instability. In Section 3.1.1, we present an overview of the stylized facts for price, and present an analysis of cross-correlations between stocks. Section 3.1.2 details our Hurst exponent results, Section 3.1.3

shows the results of the signed order flow analysis. Section 3.1.4 discusses our study on HFT and market fracture events, and our conclusions are included in Section 3.2.

Chapter 2

Methods

This research study was conducted in three phases. First we calculated stylized facts in the trade data. These are important indicators which give us a very general sense for the type of market behavior which occurred over the test period. Further, we perform a rescaled-range analysis, and calculate the Hurst exponent. Last, we analyze the role of high frequency traders in the data, and attempt to understand what market impacts these ulta-fast traders may have.

2.1 Measuring the Stylized Facts

2.1.1 Price Series Calculations

For an index set of date times T, we have that the price series $P = \{p_t \mid t \in T\}$. Note that for tick data, T is an unevenly spaced index set, since financial transactions are governed by event time, which is generally limited only by the technological capabilities of a given exchange and the participating traders. If T is evenly spaced, we let Δt be the time between

events. Due to some complexities with unevenly spaced time series data, we resample and linearly interpolate price tick data at the five minute ($\Delta t = 5$ minutes) and one hour resolution ($\Delta t = 1$ hour). We also analyze daily and weekly close time series.

From the price series, we first calculate a return series

$$R = \{r_t = ln(p_{t+1}) - ln(p_t) \mid p_t, p_{t+1} \in P\} \tag{2.1}$$

We then calculate kurtosis as $\gamma_2 = \frac{\mu_4(R)}{\sigma^4(R)}$, where μ_4 is the fourth moment about the mean, and σ is the standard deviation. Generally, in financial data over short time scales, kurtosis values are greater than 5, with no strict upper bound. One important note is that kurtosis is (quite clearly) nonlinear, thus the average kurtosis of the intraday return series over a month will not be equal to the kurtosis of the entire month's return series. In addition, as Δ_t grows very large, the distribution of returns will approach a Gaussian distribution ($\gamma_2 \approx 3$).

Next, we calculate the autocorrelations for the raw returns series as well as the absolute value return series up to a lag of 100 ticks. The efficient market hypothesis assumes that returns are uncorrelated, so we perform a Student's t-test, assuming a null hypothesis of uncorrelated returns, and rejecting our hypthosis if the condition $p < .05$ is met.

The clustering constant is calculated as

$$C = \frac{\sum_{i=1}^{50} \text{ACF}_i(|R|)}{\sum_{i=1}^{50} \text{ACF}_i(R)},$$

where ACF_i is the autocorrelation function for a lag of i ticks. For the last step in calculating our price stylized facts, we assign each return series a score of 1 or 0, where a score of 1

indicates that $\gamma_2 > 4$, $p > .05$ in the Student's t-test, and $C > 3$. A score of 0 indicates at a failure to meet at least one of these requirements. These scores serve to give us an intuition for the data, and should bring to light any unexpected behavior in our data.

2.1.2 Rescaled-Range Analysis

We perform an R/S analysis and calculate the Hurst exponent in the following manner (Mandelbrot and Hudson 2004):

1. For a price series P with n samples defined above, we first calculate the mean $m = \frac{1}{n}\sum_{i=1}^{n} p_t$, and create a mean-adjusted price series $Y_t = p_t - m$ for $t = 1, \ldots, n$.

2. We calculate a series $Z = \{Z_t = \sum_{i=1}^{t} Y_i\}$, which shows us the cumulative deviations from the mean at time t.

3. For each t, create a range-series $R_t = \max(Z_1, \ldots, Z_t) - \min(Z_1, \ldots, Z_t)$.

4. Create a standard deviation series $S_t = \frac{1}{n}\sum_{i=1}^{t}(p - \mu)^2$ where μ is the mean of the series $p_1, \ldots, p_t$

5. Now calculate $\frac{R}{S}_t = \frac{R_t}{S_t}$ for each t.

6. To calculate the Hurst exponent, plot R/S against n on a log-log plot and find the slope.

A Hurst exponent between 0.5 and 1 indicates that an increase in prices will be followed by another increase in the short term, and a decrease in prices will be followed by another decrease in a the short term. It also indicates that in the long term, prices will not generally revert to the mean. In contrast, a Hurst exponent between 0 and 0.5 indicates that in the

short term, an increase in prices will be followed by a decrease, a decrease will be followed by an increase, and price will tend to revert to its long term mean. For Hurst exponents very close to 0.5, there is no correlation between past returns and future returns, which is consistent with the efficient market hypothesis.

2.1.3 Trade Data Calculations

In the trade data, we draw from the work of Toth, Farmer and Lillo (2011), and measure autocorrelations in the series of signed orders. A signed order is simply the direction of the order, so a buy order = +1, and a sell order = -1. Thus, an autocorrelation measure on the series of signed orders tells us how likely we can predict a future buy or sell based on a past buy or sell. Farmer (2011) finds that signed orders decay as a power law with the lag time, up to 3 orders of magnitude. Thus, one would have a much better chance of predicting the sign of a future order based on past order signs than a future return based on past returns.

2.2 Analysis of High Frequency Trading

In the trade data, we look at the role of HFT in both taking and providing liquidity. We examine three types of trades: Trades with HFT taking liquidity and non-HFT providing, trades with HFT providing liquidity for a non-HFT taker, and HFT both taking and providing liquidity.

Also in the trade data, we calculate the impact of HFT on market fractures. When 10 or more trades occur consecutively in the same direction, and the average time between trades is less than 50 ms, we label the block of trades a market fracture. A large market fracture occurs when the price changes more than 0.5% during a market fracture. To measure the

impact of HFT during this phenomenon, we compare the proportion of HFT market orders over the complete time series to the proportion of HFT market orders during both types of market fractures. This measure is inspired by but not identical to work done by (Johnson et al. 2013).

We also note that the trade data has instances of multiple trades with the same timestamp. Hasbrouck (2010) interprets orders sharing the same timestamp as a single market order broken up over several smaller limit orders. For instance, if a sell market order were placed for 1000 shares, and the limit order with the best buy price is of size 900, the remaining 100 sell market orders will be executed at the next highest priced limit buy order. However, for blocks of trades with the same timestamp, NASDAQ reports all orders as having been executed at the same price. Due to uncertainty over the curation of the data, we run this analysis using the raw data provided by NASDAQ, and also by compiling all orders with the same timestamp into a single market order.

Chapter 3

Results

3.1 Overview of Results

3.1.1 Stylized Facts

First we present a naive version of the stylized facts. The Figure 3.1 shows histograms of kurtosis, clustering volatility, and autocorrelation decay for all stock over the 2008-2009 period. The plots for kurtosis and clustering volatility constants are calculated from returns series with multiple Δt values: 5-minute, hour, daily, and weekly.

For 5-minutes returns, 115 out of 117 stocks passed our stylized facts test, better than 98%. The only two failures were due to a lack of clustering volatility. For the hourly returns, the return series for all 117 stocks passed the stylized facts test. For daily returns, we find that 107 out of 117, or 91.4 % of stocks pass the stylized facts, with most failures due to a kurtosis value which was not greater than 5. However, all kurtosis values for daily returns were greater than 4, which implies they are not drawn from a Gaussian distribution.

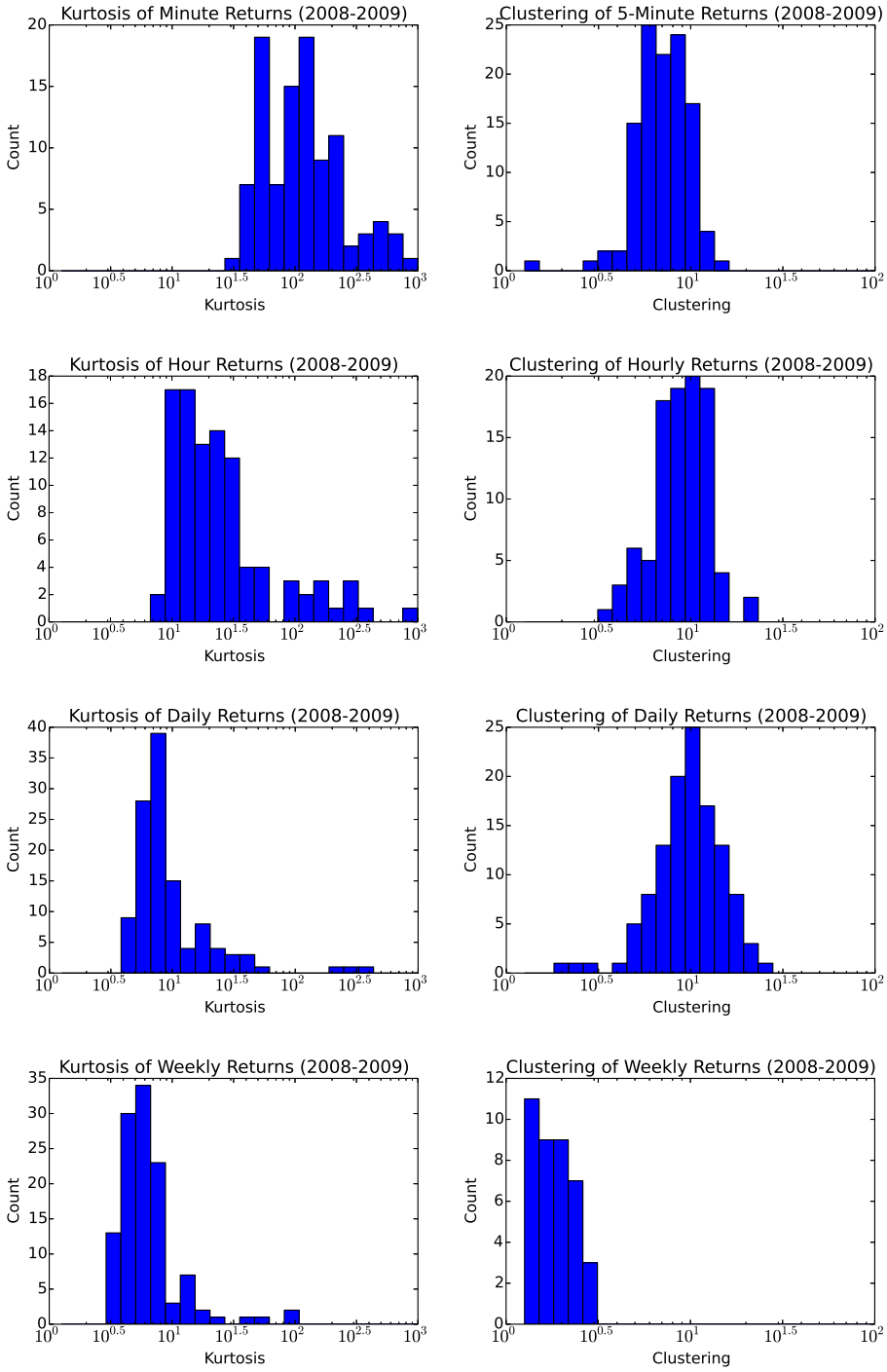
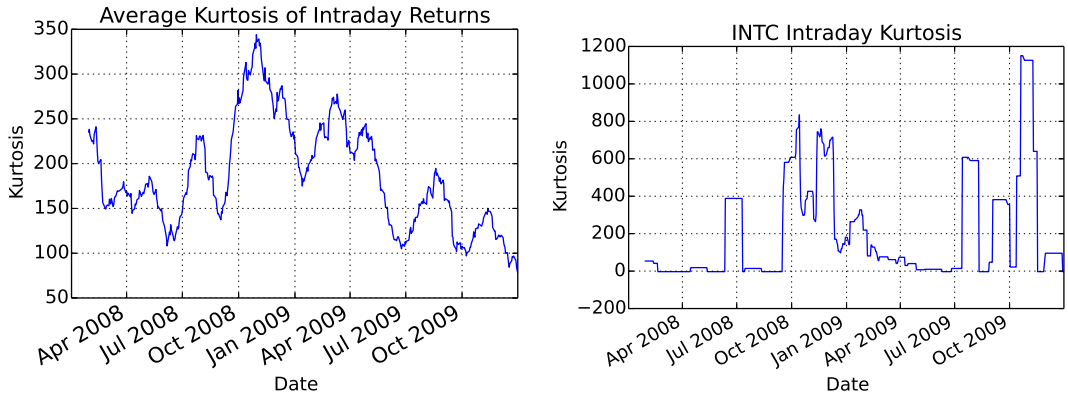


Figure 3.1: Stylized facts over different timescales



(a) (b)

Figure 3.2: As mentioned above, kurtosis is based on the 4th moment, and thus is highly nonlinear metric. Tick returns, shown in (a) and (b) will have much higher kurtosis values than minute, hour, or any lower resolution returns.

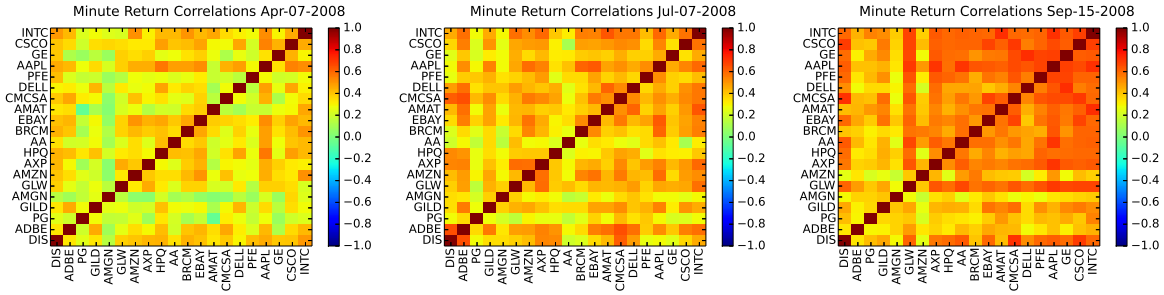
As the time between observations increases, and the number of observations decreases, both kurtosis and clustering volatility tend to decrease. This result is corroborated by other work in the field (Cont 2001), as over larger observation windows, returns tend to look Gaussian. Thus, looking at monthly returns throughout history gives very little insight into the occurance of market crashes or more generally, market volatility.

By looking at kurtosis at smaller timescales, we are able to get a sense of how volatility in price changes over time. In Figure 3.2, we see a moving average of the kurtosis of intraday returns, smoothed with a sliding window of 20 days.

Note that around October 2008, during the brunt of the crisis, mean intraday kurtosis reaches record highs for the two-year period. While it is somewhat surprising that kurtosis does not spike earlier during the crisis week of September 15, this is an average across all stocks, so we may see the behavior we expect in some of the more highly traded stocks.

INTC is the most highly traded stock during the 2008-2009 period, and below we see how the moving average of kurtosis peaks during September 2008, and maintains well above ”normal” levels for the next few months.

Kurtosis does not tell the whole picture regarding volatility of returns. We look next at correlations between returns leading up to, during, and after September 2008. The following figure details correlations of returns at the 5-minute level for the 20 most traded stocks, for the week beginning on the day specified in the plot title.



(a) April 07, 2008 (b) July 7, 2008 (c) September 15, 2008

Figure 3.3: In these plots of correlations between 5-minute returns, the week of September 15 (c) stands out emphatically.

Immediately following the bankruptcy of Lehman Brothers on September 14, lenders, which included large investment firms and banks, issued margin calls to hedge funds and other investors that they had lent money to. The investors who had to meet their margins had only a limited time to pay back the lenders, and thus had to start selling assets immediately to collect enough capital to pay back lenders. As prices fell across the financial system, more investors failed to meet margins, the cycle of rapid selling continued, and banks stopped lending, culminating in the financial crisis. This is highlighted in the cross-correlations between returns in Figure 3.3.

We run the same analysis with returns at the one hour resolution. Comparing Figure 3.3 and Figure 3.4, we note that correlations between returns at the one hour resolution are not present at the minute timescale, nor indeed at any other timescale. This could be due to random effects, and without doing further analysis on other financial data, we do not wish to make any conclusions.

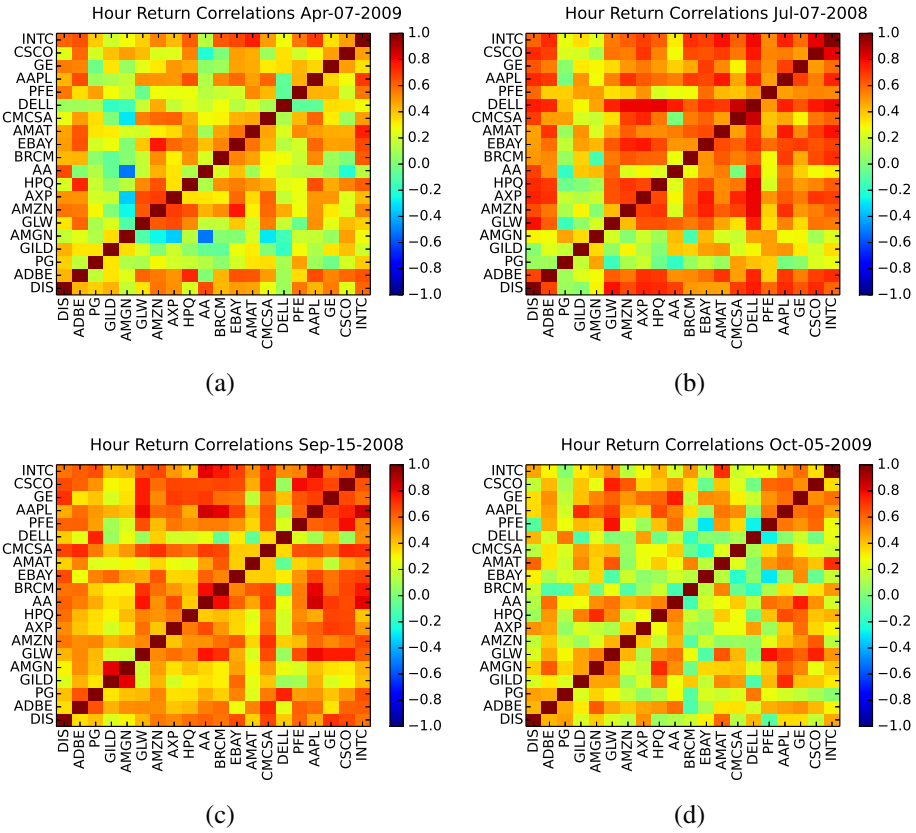


Figure 3.4: Cross-correlations of returns at the one hour resolution are much stronger than cross-correlations of 5-minute returns.

3.1.2 Hurst Exponent

Next, we calculate the Hurst exponent for returns at the 5-minute resolution level. We split the return series into 6 sets of return series: The first set consisted of the entire 2008-2009 series, which included 39815 observations at the 5-minute level. The next set consisted of two return series, including 19907 and 19908 observations. The first of these contained all of the 2008 returns, and the other contained the 2009 returns. We continued this process, splitting each series up into halves and doubling the number of series in each set. The last set of series contained 32 return series, each containing 1244 observations. We then follow the steps outlined above (Rescaled Range Analysis).

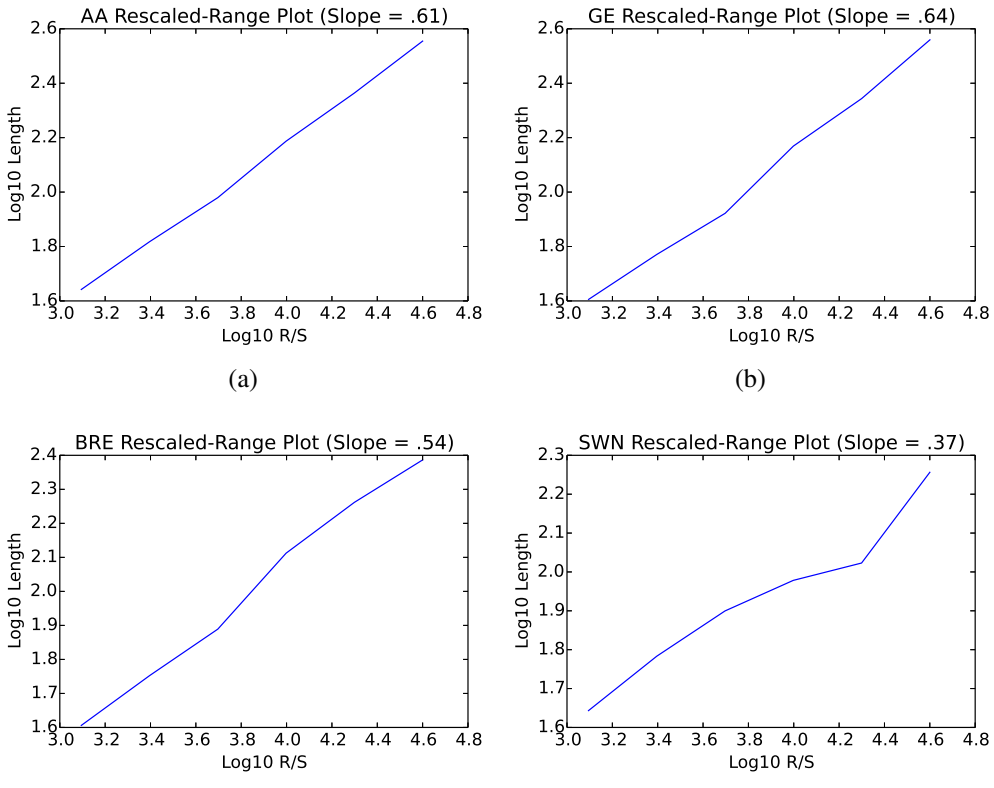
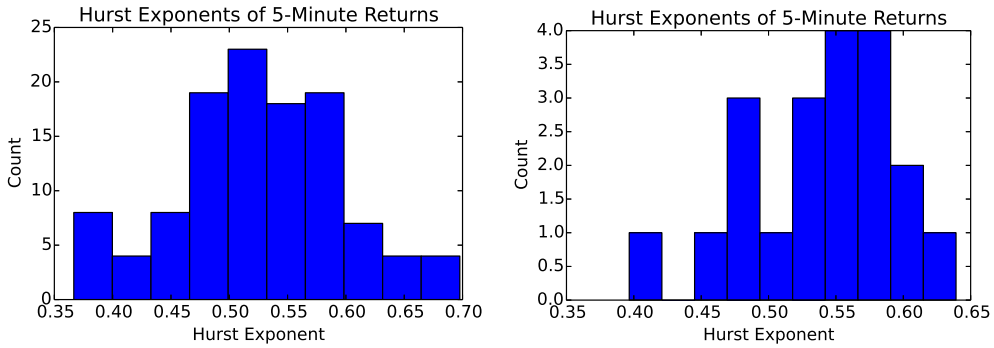


Figure 3.5: Rescaled range plots of 4 selected stocks

Some of the subplots in Figure 3.5 are noisier than others, and we see that trend in BRE and SWN. Since neither of these stocks are among the most traded, the noise is likely a result of the data interpolation which we performed to get returns at the 5-minute level. Both AA and GE are highly traded, and also have high Hurst exponents.

The average Hurst exponent across all stocks was calculated as 0.526 (median = 0.528), which is not significantly different from a Brownian motion process, indicating that on average, no long term memory was detected in stock returns. Note, however, the difference in the following plots:



(a) All Stocks (b) Top 20 Stocks

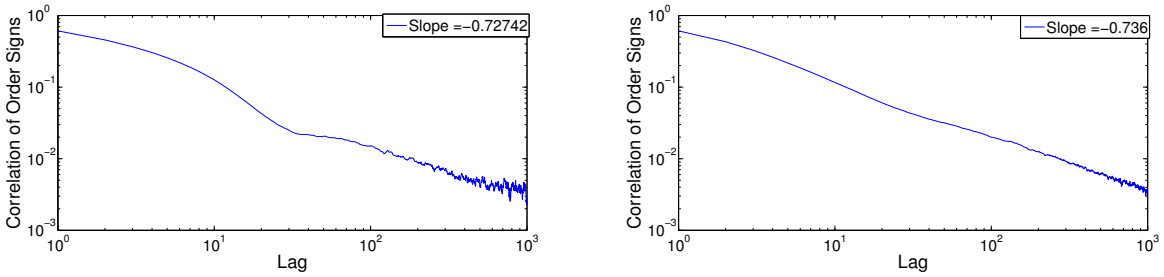
Figure 3.6: The second figure is a histogram of only the 20 most traded stocks by volume. The average Hurst exponent of these stocks is 0.538, which is slighty higher than the overall average of 0.526, and the median is 0.549.

3.1.3 Signed Order Flow

We find extreme persistence in autocorrelations over large time lags in signed order flow. While we make no claim about power law behavior of these autocorrelations, we do note that the 90% confidence interval of log-log slopes is (-0.88,-0.44), and the r^2 values for all linear fits are greater than 0.7.

Thus, the ability to predict future order flow from past order flow is much greater than the ability to predict future returns based on past returns. As noted in Section 1.1.1, Toth (2011) attributes most of this autocorrelation persistence to order splitting, the breaking down of large orders into several smaller orders to reduce market impact. We test this hypothesis against our fracture data below.

Note that these results are reported from the raw trade data (as described in Section 2.2), but the adjusted data results do not qualitatively differ. Figure 3.7 shows the fat tailed persistence of signed order flow autocorrelations.



(a) GOOG (b) AAPL

Figure 3.7: Two example signed order flow autocorrelation plots. Almost all stocks had the same break in scaling (ankle) which appears in (a) at around $10^1.1$ lags. There was no correlation between the total volume of a stock and the existence of the ankle in its autocorrelation function, implying the ankle is not related to noise in the data. The ankle shows that autocorrelations decay more quickly during the first order of magnitude than over the rest of the data.

3.1.4 HFT Analysis

During the Flash Crash, over 70% of excecuted trades were exclusively between one HFT and another HFT (Kirilenko 2010). Using our measures of fracture events in Section 2.2, we measure the proportion of HFT in three types of trades: Trades with HFT as liquidity

takers, and non-HFT providing liquidity, trades with HFT providing liquidity for non-HFT takers, and HFT both taking and providing liquidity in a trade.

Thus we have that Total HFT Taking = HFT Taking Only + Both HFT, and likewise, Total HFT Providing = HFT Providing Only + Both HFT.

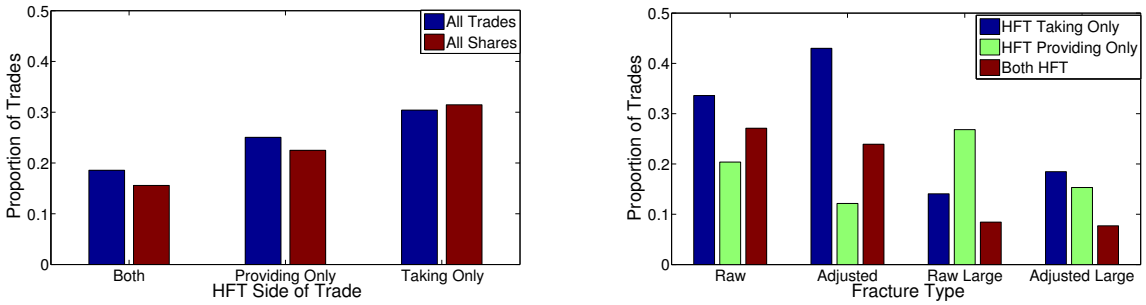


Figure 3.8: HFT makes its largest impact (as a proportion of trades) as a liquidity taker during small adjusted fracture events. HFT makes up a smaller proportion of taking, providing, and both during large adjusted fractures than it does in overall market activity.

HFT provided nearly 40% of all liquidity (in shares) over the 24 month sample period, while taking nearly 50% of the liquidity. In fact, 18% of all trades and 15% of all shares traded had HFT platforms on both sides of the trade.

In all of the blocks of raw fractures, over 37 million trades took place, or about 6% of total trades. Adjusting our data as described in Section 2.2, we are left with under 4 million uniquely timestamped trades that fit our criteria of fractures, about 0.6% of all trades made in the dataset.

Using our adjusted data, HFT plays a larger part taking liquidity during fractures than during overall trading. HFT acts as a liquidity taker in over 66% of the adjusted fractures trades, compared to 50% of overall trades. However, in the raw data, HFT plays only a slightly larger role in taking liquidity during fractures. Thus we have contradicting con-

clusions based on the curation of the data, and further work will need to be done in this area.

In contrast, looking only at large returns yields a clearer picture. HFT plays a smaller role as a taker in trades that occur during both raw and adjusted fractures. This suggests that extremely fast and moderately large price changes are not driven by HFT, but instead by traders placing large orders that are split into pieces and executed very quickly in succession, as noted by Toth (2011).

Typically, order splitting occurs to reduce market impact. If trades are being placed less than 50 ms apart, the trader who placed the order is not giving the market much time to respond, thereby rendering the order splitting useless. One hypothesis for this behavior could be that the trader is relying on HFT to provide liquidity in real time for the order, but as we see in Figure 3.8(b), HFT liquidity providing during large fractures is either equal to or less than overall HFT providing. Thus, HFT is not playing a major role in providing liquidity for these trades. Once again, further research is needed in order to determine why these blocks of trades are being placed and their impacts on market stability.

3.2 Conclusion

In order to further explore the statistical properties of financial data, we examined trade data from 2008 to 2009 of 120 stocks. We conclude that during the 2008 Financial Crisis, the mean kurtosis value across all stocks reached its maximum. Further, we find that cross-correlations in 5-minute returns have little predictive power, but cross-correlations in hourly returns may contain more information.

Our analysis of the stylized facts across different samplling periods yields results consistent with previous work. In addition, we found some evidence of long range dependence

in price returns in an examination of Hurst exponents. Last, we explored the phenomenon of persistence in autocorrelations of signed order flow, and attempt to research a connection between market fractures, signed order flow, and high frequency trading. We find that HFT plays a smaller role in taking liquidity during large market fractures than in overall trading.

We leave the mechanism driving large market fractures as an open question. We point to agent based modeling to answer these mechanistic questions, while also emphasizing the necessity of further analyzing high resolution market data to guide computational experiments.

Bibliography

(2013). *Latency Arbitrage, Market Fragmentation, and Efficiency: A Two-Market Model*.

Brady, N. F. (1988). The brady report. Technical report, U.S. Government Printing Office.

Brogaard, J., T. Hendershott, and R. Riordan (2013). High frequency trading and price discovery. *Review of Financial Studies*.

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance 1*, 233–236.

Darley, V., A. Outkin, T. Place, and F. Gao (2001). Learning, evolution and tick size effects in a simulation of the nasdaq stock market. In *Proceedings of the 5th World Multi-Conference on Systemics, Cybernetics, and Informatics*, Orlando, FL.

Easley, D., M. L. de Prado, and M. O'Hara (2012). Flow toxicity and liquidity in a high frequency world. *Review of Financial Studies*.

Farmer, J. D., P. Patelli, and I. I. Zovko (2005). The predicative power of zero intelligence in financial markets. *Proceedings of the National Academy of Sciences of the United States of America 102*(6), 2254–2259.

Hasbrouck, J. and G. Saar (2010). Low-latency trading. October 2, 2010.

Hendershott, T., C. M. Jones, and A. J. Menkveld (2011). Does algorithmic trading improve liquidity. *The Journal of Finance*.

Johnson, N., G. Zhao, E. Hunsader, H. Qi, N. Johnson, J. Meng, and B. Tivnan (2013, 09). Abrupt rise of new machine ecology beyond human response time. *Sci. Rep. 3*.

Kirilenko, A. (2010). Findings regarding the market events of may 6, 2010. Technical Report http://www.sec.gov/news/studies/2010/marketevents-report.pdf, U.S. Commodity Futures Trading Commission and U.S. Securities and Exchange Commission.

LeBaron, B. (2001). A builder's guide to agent-based financial markets. *Quantitative Finance*.

Lux, T. (2009). *Handbook of Financial Markets: Dynamics and Evolution*, Chapter Stochastic Behavioral Asset-Pricing Models and the Stylized Facts, pp. 161–215. Elsevier.

Mandelbrot, B. (1963). Variation of certain speculative prices. *Journal of Business 36*(4), 294–419.

Mandelbrot, B. B. and R. L. Hudson (2004). *The (Mis)Behavior of Markets*. Basic Books.

Preis, T., S. Golke, W. Paul, and J. J. Schneider (2006). Multi-agent–based order book model of financial markets. *Europhysics Letters*.

BIBLIOGRAPHY

Toth, B., I. Palit, F. Lillo, and J. D. Farmer (2011). Why is order flow so persistent? *Quantitative Finance*.

Yeo, W. Y. (2009). Limit order book liquidity and liquidity imbalance. Working Paper.