UNIVERSITY OF CALIFORNIA
SANTA CRUZ

HIGH FREQUENCY TRADE DIRECTION PREDICTION

A thesis submitted in partial satisfaction
of the requirements for the degree of

MASTER OF SCIENCE

in

PHYSICS

by

Augustine Stav

June 2015

The Thesis of Augustine Stav
is approved by:

_____
Professor Gregory Laughlin, Chair

_____
Professor Anthony Aguirre

_____
Professor Eric Aldrich

_____
Professor Tyrus Miller
Vice Provost and Dean of Graduate Studies

Contents

List of Figures

High Frequency Trade Direction Prediction

by

Augustine Stav

Abstract

High frequency trading involves large volumes and rapid price changes. The Volume Synchronized Probability of Informed Trading (VPIN) metric characterizes order flow toxicity. This toxicity is the unbalance of order flow between informed traders who possess knowledge of future price directions and market makers who do not have this information. VPIN requires trades to be classified as buys or sells and works with volume as a proxy for information arrival. As an alternative, trade volatility is determined from the trade direction. Subsequent trades are either at the same price, reversive, or trending. The virtual price takes continuous values between the bid-ask spread and exhibits Brownian motion. The realized price is the virtual price rounded to the nearest tick. Changes in the actual price occur when the realized price crosses the spread. The volatility parameter of the Brownian probability density function is determined so that the model has the greatest correlation to the observed trade directions. According to the Chicago Board Options Exchange Market Volatility Index, VIX, August 5th – 7th and August 20th – 22th, 2014 were periods of relatively high and low volatility for the S&P 500 respectively. The volatilities obtained for the high and low periods are (4.80 ± 2.02) dollars^2/trade and (4.21 ± 0.71) dollars^2/trade respectively.

Acknowledgments

I want to thank my parents, John and Sokhan Stav, and my friend, Monica Tutor, for their support. I would also like to thank Professor Gregory Laughlin for his guidance and invaluable feedback.

1. Introduction

High frequency (HF) traders act as market makers, quoting bid and ask prices. They attempt to earn small margins on a large volume of trades. HF traders turn over their inventories five or more times per day, leading to the rapid growth of HF trading to 73% of U.S. equity trading volume (Sussman et al. 2009). Easley et al. (2011) have proposed using order flow, or signed trades, to determine short-term price movement. Aggressive buys (sells) are market orders to buy (sell) at the ask price (bid price). Informed traders possess knowledge of future price directions that the market makers do not. Trading against informed traders leads to losses. From a market makers' perspective, order flow is toxic when the other parties in a trade are informed. (Easley et al. 2011).

HF trading data may involve rapid price movements and unsigned trades. If trades are unsigned, order flow classification schemes are necessary. Easley et al. (2012) developed a bulk classification method involving the price changes between "volume buckets" described in Section 2. The Volume Synchronized Probability of Informed Trading (VPIN) is given by the sum of the differences between sell and buy volume as a fraction of the total volume. High order flow toxicity, represented by VPIN, may force market makers to liquidate or abandon market making activities. VPIN, then, may predict short-term toxicity-induced volatility. Easley et al. (2012) divide the VPIN distribution into twenty 5 percentile bins. 51.84% of the times that VPIN enters a bin in the upper quartile of its distribution, there is at least one absolute

return between volume buckets greater than 0.75% before VPIN leaves that 5%-tile.

Volatility characterizes uncertainty about future returns. It can be defined as the standard deviation of the return provided in one year when the return is expressed using continuous compounding. Volatility can be estimated empirically from historical data or implied volatilites can be obtained from option pricing models. The Black–Scholes–Merton formulas for the prices of European call and put options, for example, are solutions to a differential equation characterizing the dependence of a derivative price on the underlying stock. Implied volatilities are calculated in such a way that, when input into the pricing model, they return the option prices observed in the market. Contrasting with historical volatilities, implied volatilities characterize the market's opinion about the volatility of a stock. VIX is an index of the implied volatility of 30-day options on the S&P 500. (Hull 2012).

The purpose of this paper is to derive a metric from millisecond data, based on trade direction, as an alternative to order flow toxicity. Section 2 describes the VPIN model, Section 3 presents the trade direction view of trade data, and Section 4 matches a random walk to the trade direction data.

2.1 The Volume Synchronized Probability of Informed Trading Model

The Probability of Informed Trading (PIN) developed by Easley et al. (1996) is the fraction of informed trading relative to total order flow. Traders are either informed or uniformed. Information arrives between trading days with probability α. This information is either bad, with probability δ, or good, with probability 1-δ. Informed traders sell after bad news and buy after good news following a Poisson process characterized by a daily arrival rate μ. Uninformed traders submit sell or buy orders according to independent Poisson processes with arrival rates ϵ_s and ϵ_b respectively.

The likelihood of observing B buys and S sells on a trading day is:

$$\begin{aligned} L((B,S)|(\alpha,\delta,\mu,\epsilon_b,\epsilon_s)) &= (1-\alpha)e^{-\epsilon_b}\frac{(\epsilon_b)^B}{B!}e^{-\epsilon_s}\frac{(\epsilon_s)^S}{S!} \\ &+ \alpha\delta e^{-\epsilon_b}\frac{(\epsilon_b)^B}{B!}e^{-(\epsilon_s+\mu)}\frac{(\epsilon_s+\mu)^S}{S!} \\ &+ \alpha(1-\delta)e^{-\epsilon_s}\frac{(\epsilon_s)^S}{S!}e^{-(\epsilon_b+\mu)}\frac{(\epsilon_b+\mu)^B}{B!} \end{aligned} \quad (1)$$

Using (1), maximum likelihood estimates of the parameters (α, δ, μ, ϵ_b, ϵ_s) are obtained.

PIN is given by:

$$PIN = \frac{\alpha\mu}{\alpha\mu + \epsilon_b + \epsilon_s} \quad (2)$$

Easley et al. (2011) have developed a high frequency estimation of PIN: Volume Synchronized Probability of Informed Trading. Whereas PIN models information arrival with clock time, VPIN uses trade volume as a proxy for

information arrival. The average daily trade volume $\overline{V}$ is divided by a number n into volume buckets of equal size $V = V_\tau^B + V_\tau^S$. V_τ^B and V_τ^S are the buy and trade volume in volume bucket τ, and

$$\frac{1}{n}\sum_{\tau=1}^{n}(V_\tau^B + V_\tau^S) = V = E[V_\tau^B + V_\tau^S] = \alpha\mu + \epsilon_b + \epsilon_s \tag{3}$$

$E[V_\tau^B + V_\tau^S]$ is the expectation value of the total number of trades in volume bucket τ. The expected trade imbalance approximates αμ (the fraction of orders from informed traders):

$$\frac{1}{n}\sum_{\tau=1}^{n}\left|V_\tau^B - V_\tau^S\right| = E\left|V_\tau^B - V_\tau^S\right| \approx \alpha\mu \tag{4}$$

With unsigned data, calculation of the order imbalance requires a volume classification scheme. The bulk volume classification method classifies a fraction of the τ^{th} volume bucket as buys or sells. Trades are rolled over into 1 minute timebars, indexed by t and having volume V_i, to mitigate order splitting effects. Let t(τ) be the index of the last timebar included in bucket τ, and let t(τ-1) + 1 be the timebar immediately after the last timebar of bucket τ-1. A fraction of each timebar volume V_i is assigned as buys or sells assuming a normally distributed change in price between timebars ($P_i - P_{i-1}$):

$$V_\tau^B = \sum_{i=t(\tau-1)+1}^{t(\tau)} V_i Z\left(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}}\right) \quad \text{and} \quad V_\tau^S = V - V_\tau^B$$

$$\text{where} \sum_{i=t(\tau-1)+1}^{t(\tau)} V_i = V \tag{5}$$

where Z is the cumulative distribution function of the standard normal distribution and $\sigma_{\Delta P}$ is the estimated standard deviation of the price changes between time bars. The order flow toxicity metric is:

$$\begin{aligned} VPIN &= \frac{\frac{1}{n}\sum_{\tau=1}^{n}\left|V_{\tau}^{B}-V_{\tau}^{S}\right|}{\frac{1}{n}\sum_{\tau=1}^{n}(V_{\tau}^{B}+V_{\tau}^{S})} = \frac{\sum_{\tau=1}^{n}\left|V_{\tau}^{B}-V_{\tau}^{S}\right|}{nV} \\ &\approx \frac{\alpha\mu}{\alpha\mu+\epsilon_{b}+\epsilon_{s}} = PIN \end{aligned} \tag{6}$$

(Easley et al. 2012) (Abad and Yagüe, 2012).

2.2 The Application of VPIN to the E-mini S&P 500

Our primary source of information for this analysis is market depth data for the E-Mini S&P 500 Futures contract purchased from the Chicago Mercantile Exchange (CME) (CME DataMine 2015). These data are recorded and timestamped at the Globex matching engine, currently located in Aurora, Illinois (longitude -88.24º W, latitude 41.80° N). At the time of the Flash Crash on May 6th, 2010, the matching engine was located at 350 E. Cermak Road in Chicago (longitude -88.13° W, latitude 41.73° N). Session data is written to ASCII files using the FIX specification. Level-2 order book activity to a price depth of 10 on both the bid and the offer side of the order book is captured, along with trade records and other information relevant to recreating the trading session. All order book events are time-stamped to millisecond precision, with time signals propagated from GPS receivers. Events that occur within a single millisecond are time ordered.

US Equity futures prices formed in Chicago are generally understood to lead

cash prices formed in the US equity markets themselves (see, e.g. Hasbrouck 2003). Specifically, equity prices are determined by the price of the near-month E-Mini S&P 500 futures contract, which is traded on the CME's Globex platform and valued at 50 times the numerical value of the S&P 500 Stock Index on the contract expiration date. The E-mini contract trades on the March quarterly cycle (March, June, September, and December) and expires on the third Friday of the contract month. On the so-called "roll date," eight days prior to expiry, both liquidity and price formation shift to the contract with the next-closest expiry date. Several million E-mini contracts are traded each day, corresponding to dollar volumes that frequently exceed $200 billion.

A second, higher-level, source of tick data has been obtained from the NYSE Technologies Market Data Custom Trade and Quote Services (NYSE Custom TAQ 2014), and consists of trades in the symbol SPY aggregated across the Consolidated Market System, with price and trade size information time-stamped to the millisecond.

We obtain Figure 1 using the VPIN calculated from Equation (6). The figure shows VPIN in red and SPY price in blue vs. time on May 6, 2010, the day of the Flash Crash. This calculation uses average daily trade volume $\overline{V}$ = 100,000,000 and number of volume buckets n = 50.

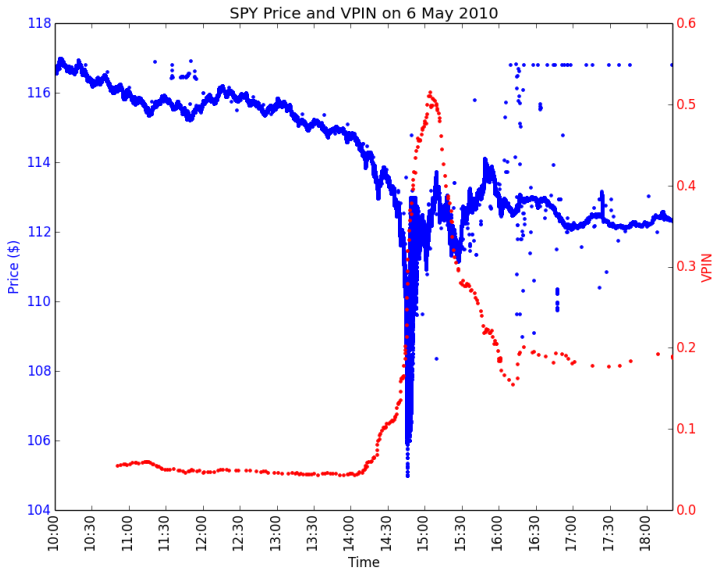

Figure 1: SPY Price and VPIN metric on 6 May 2010.

Figure 1 shows the SPY price begin to drop precipitously at approximately 2:30pm. VPIN begins its ascent 20 minutes prior and could serve as an indicator of toxic order flow. Easley et al. (2011) determine an empirical distribution of VPIN for the E-mini S&P 500. They find that CDF(VPIN) exceeds 95% at 1:08pm, compared to the 2:32pm accepted 'start' of the crash.

3. Predicting Trade Direction

The following analysis looks at consecutive past trades to predict the next trade type of the E-mini S&P 500 Futures. Using the sign of previous price changes, this metric attempts to predict the direction of future price movements. Current trades are either at the same price, trending, or reversive relative to the previous trade. Let trades be organized chronologically and have index i. The price change is $\Delta P_i = P_i - P_{i-1}$. Trending trades continue the direction of price movement, so that $\text{sign}(\Delta P_i) = \text{sign}(\Delta P_{i-1})$. Reversive trades change the direction of price movement, so that $\text{sign}(\Delta P_i) = -\text{sign}(\Delta P_{i-1})$.

Each trade i has n^i_{xy} prior consecutive trades of a certain type. The subscript x denotes the type of trade i (reversive, trending, or same price), while y represents the type of prior consecutive trades. Let n_{xy}, without the superscript i, represent the number of y trades immediately before any x trade. N is the number of instances where $n^i_{xy} = n_{xy}$. The relative frequency of n_{xy} for given trade types x and y is:

$$F(n_{xy}) = \left(\frac{N}{T_{day}} \right) \tag{7}$$

where T_{day} is the total daily number of trades. To mitigate the effects of order splitting, trades with the same millisecond timestamp, price, and sign (aggressive buy or sell) are aggregated. As an example, Figure 2 plots the frequency according to Equation (7) for the different types of trades prior to same price trades on August 5-7, 2014 for the E-mini S&P 500 Futures. The "Reversive Preceding Same Price" subplot (middle of the figure) shows the relative frequency of the number of consecutive reversive

trades immediately prior to each same price trade. The other plots are organized similarly.

The percentage of the times that n consecutive y trades leads to an x trade is:

$$G(n_x) = 100\% \times \frac{F(n_{xy})}{\sum_{allx} F(n_{xy})} \tag{8}$$

As an example, let $F(n_{rs})$, $F(n_{ss})$, and $F(n_{ts})$ be the $F(n_{xy})$ for consecutive same price trades immediately prior to reversive, same price, and trending trades respectively. Then the percentage of the times that n consecutive same price trades leads to a reversive trade is:

$$G(n_r) = 100\% \times \frac{F(n_{rs})}{F(n_{rs}) + F(n_{ss}) + F(n_{ts})} \tag{9}$$

Figure 3 provides a graphical example of $G(n_x)$ for n consecutive same price, reversive, and trending trades on August 5-7, 2014. The subscript x signifies that the trade after the consecutive string is either same price (blue), reversive (green), or trending (red). For example, the "% Same Price Leading to Same, Reversive, or Trending" subplot shows the percentage of the times that n consecutive same price trades leads to a same price, reversive, or trending trade.

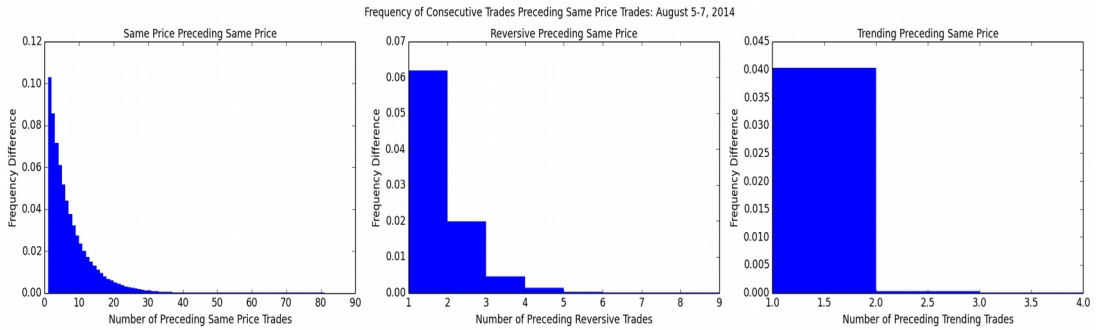

Figure 2: Frequency of trade types preceding same price trades, E-Mini S&P 500 on August 5-7, 2014.

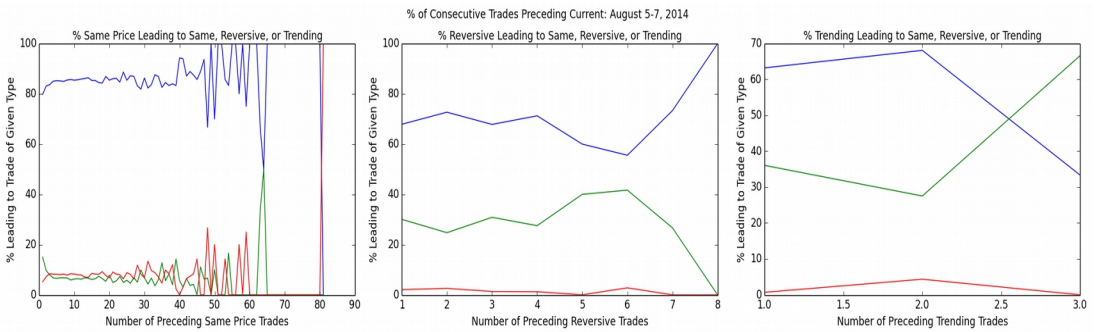

Figure 3: Percentage of n consecutive given trade types leading to same price (blue), reversive (green), or trending (red) trades on August 5-7, 2014.

In order to test the effect of volatility on the analysis, the relative frequency $F(n_{xy})$ is computed for three days of a local VIX maximum: August 5th, 6th, and 7th 2014. These days had a VIX of 16.87, 16.37, and 16.66 respectively. This is repeated for three days of a nearby local VIX minimum: August 20th, 21th, and 22th 2014. These days had a VIX of 11.78, 11.76, and 11.47 respectively. For comparison, the 90-day VIX simple average ending on September 30, 2014 was 12.63 (Yahoo! Finance 2014). Let the difference in the relative frequency be:

$$\Delta F(n_{xy}) = \sum_{VIX_{max}} F(n_{xy}) - \sum_{VIX_{min}} F(n_{xy}) \tag{10}$$

where the summation is over each three day period. Figures 4 through 6 plot $\Delta F(n_{xy})$ in August 2014 for trades prior to reversive, same price, and trending trades respectively. These Figures are repeated with uncertainty estimates in the Appendix in Figures 14 through 16. The uncertainty is obtained from bootstrap resampling. The bootstrapping procedure involves random resampling of the data with replacement. Each resample is equal to the size of the original data set. The low and high bounds of the error bars are obtained from the 15^{th} and 85^{th} percentiles of 100 bootstrap resamples. Examples of single realizations of the bootstrap resampling procedure are shown in Figures 8 through 10, which can be compared to the corresponding original data represented in Figures 4 through 6. Figure 7 shows $G(n_x)$ vs. n consecutive trades leading to same price (blue), reversive (green), or trending (red) trades for the high and low VIX periods in August 2014. Negative percentages are possible because $\Delta F(n_{xy})$ can take negative values.

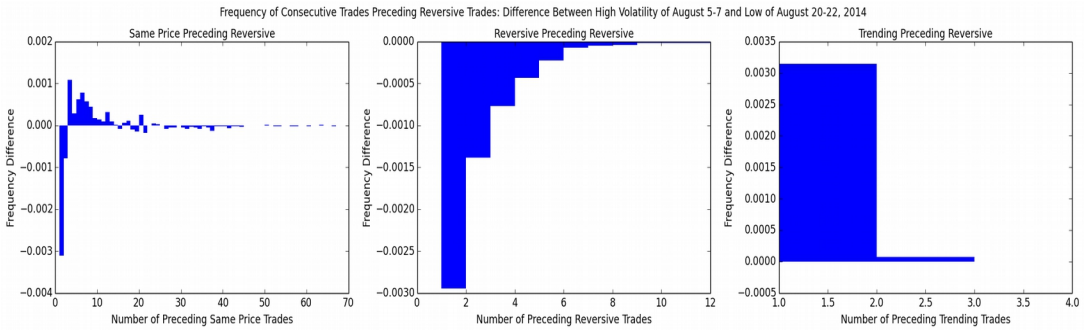

Figure 4: Frequency of trade types preceding reversive trades, difference between high VIX and low VIX for E-Mini S&P 500 in August 2014.

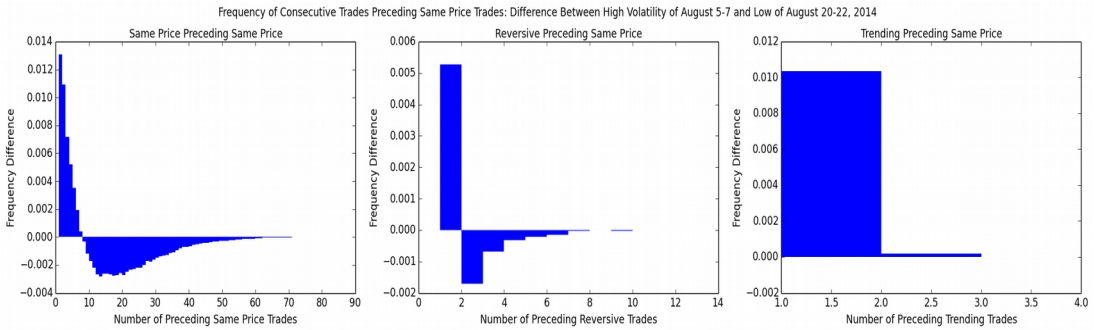

Figure 5: Frequency of trade types preceding same price trades, difference between high VIX and low VIX for E-Mini S&P 500 in August 2014.

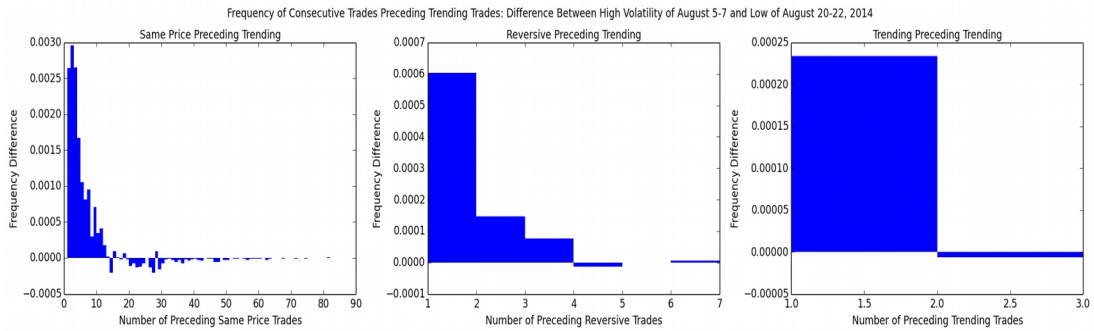

Figure 6: Frequency of trade types preceding trending trades, difference between high VIX and low VIX for E-Mini S&P 500 in August 2014.

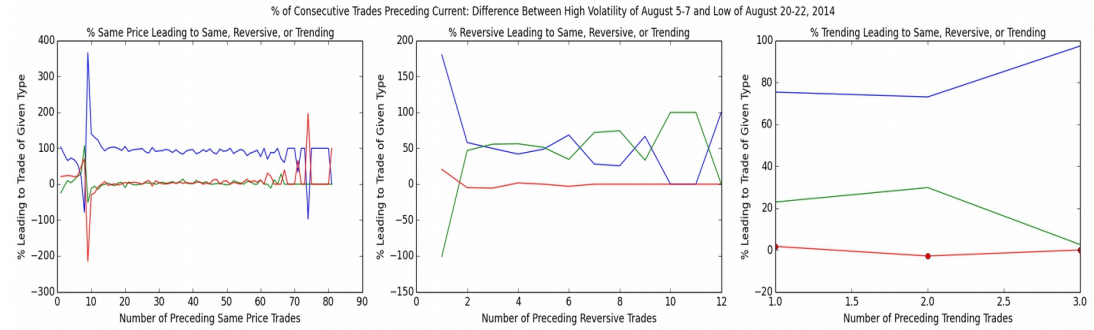

Figure 7: Percentage of n consecutive given trade types leading to same price (blue), reversive (green), or trending (red) trades, difference between high VIX and low VIX for E-Mini S&P 500 in August 2014.

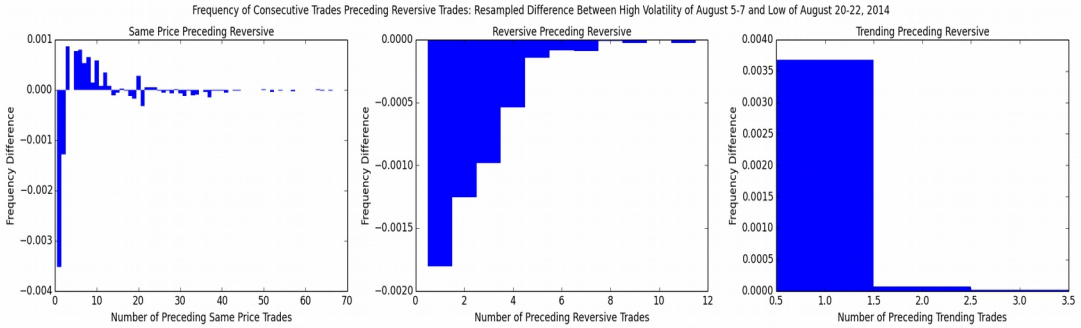

Figure 8: Bootstrap resample of frequency of trade types preceding reversive trades, difference between high VIX and low VIX for E-Mini S&P 500 in August 2014.

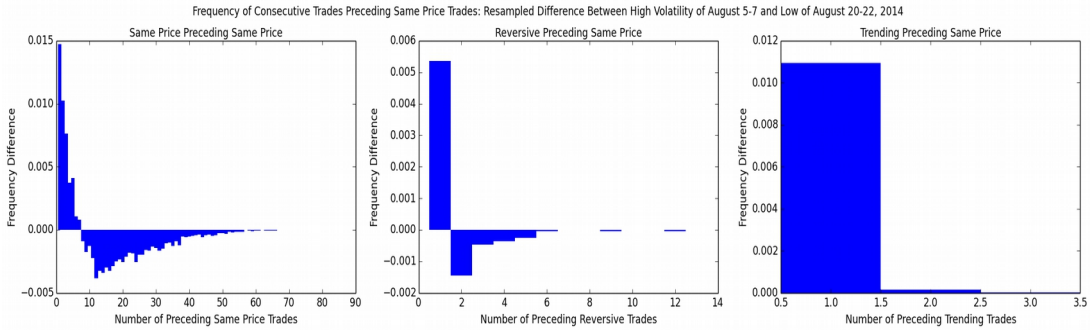

Figure 9: Bootstrap resample of frequency of trade types preceding same price trades, difference between high VIX and low VIX for E-Mini S&P 500 in August 2014.

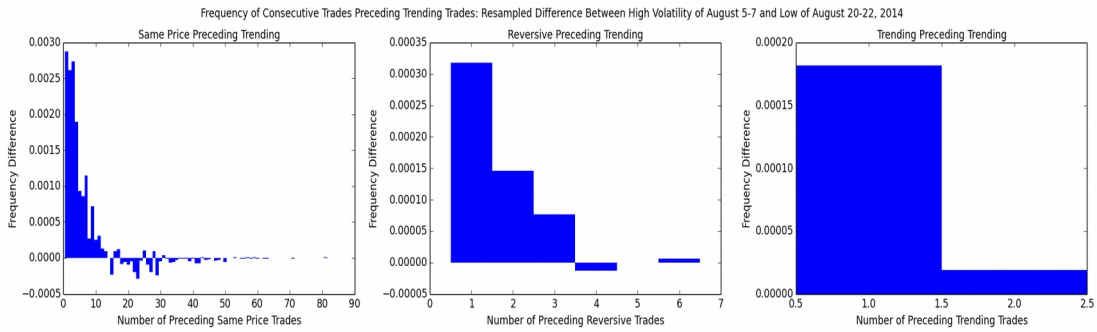

Figure 10: Bootstrap resample of frequency of trade types preceding trending trades, difference between high VIX and low VIX for E-Mini S&P 500 in August 2014.

See the Appendix on page 21 for Figures 14 through 16 with error bars obtained from

the 15th and 85th percentiles of 100 bootstrap resamples.

Same price trades dominate. Figures 4 through 6 show that larger numbers, greater than $n \approx 10$, of consecutive same price trades leads to reversive, same price, or trending trades more often on low VIX days, as $G(n) < 0$. Similarly, $G(n) < 0$ for reversive trades preceding reversive or same price trades. These indicate that low VIX days correspond to less variation in the price (larger numbers of consecutive same price trades). In addition, low VIX days can be characterized by a larger number of consecutive reversive trades than on high VIX days, indicating a tendency to undo previous price changes. $G(n)$ takes positive values in the "Same Price Preceding Same Price" subplot of Figure 5 only for a number of consecutive trades fewer than $n \approx 10$. This shows that on high VIX days the price fails to remain constant for more than a few trades.

4. Parameterizing A Random Price Walk

The actual price of E-mini futures has a minimum price change of one tick or $12.50. Let a virtual price take continuous values between the bid-ask spread and exhibit Brownian motion. The realized price is the virtual price rounded to the nearest tick. Changes in the actual price occur when the realized price crosses the spread. Brownian motion is characterized by the probability density function

$$f(p,t) = \frac{1}{4\pi Dt} e^{\frac{-p^2}{4Dt}} \tag{11}$$

where p is the price. The value t is a proxy for time elapsed: the trade number (trade time). The parameter D is expected to follow

$$\overline{p^2} = 2Dt \tag{12}$$

D is determined with an iterative search. For each trial value of D, the virtual price evolves as follows: the price changes by increments drawn from a normal distribution with standard deviation $(2D)^{1/2}$, equation (11). The total number of trade time steps is set equal to the number of trades for each period of high and low volatility: August 5th – 7th and August 20th – 22th, 2014. D is chosen so that the evolution of the actual prices produced have the greatest correlation to the data obtained from equation (7) in Section 3.

For the high volatility days, D = (4.80 ± 2.02) dollars^2/trade. For the low volatility days, D = (4.21 ± 0.71) dollars^2/trade. To a first approximation, the spread = one tick = $12.50. Due to rounding, for the realized price to cross the spread, the virtual price must change by half the spread. According to Equation (12), D ≈

(spread/2)^2/2t_0 where t_0 is the average number of trades before a price change. For the high volatility days, t_0 = 3.45 trades, so that D ≈ 5.66 dollars^2/trade. Figures 11 through 13 plot $\Delta F(n_{xy})$ for trades prior to reversive, same price, and trending trades respectively, simulated using the D values obtained from the iterative search. Figures 17 through 19 in the Appendix are the same plots, with error bars obtained from the upper and lower D bounds.

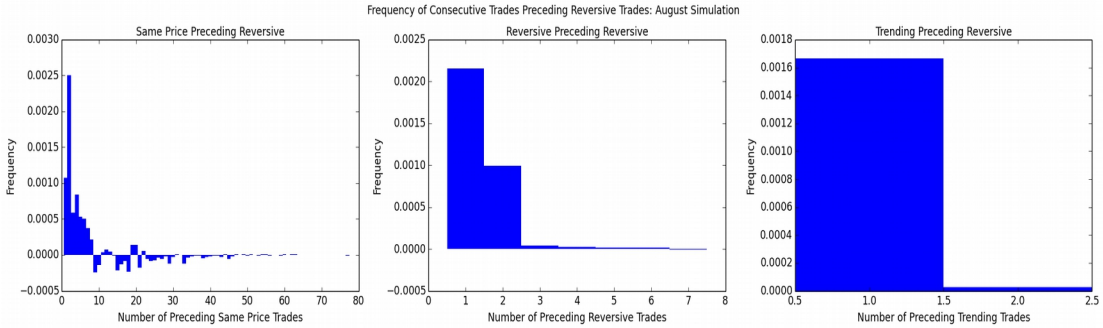

Figure 11: Brownian motion simulation: frequency of trade types preceding reversive trades, difference between high volatility and low volatility.

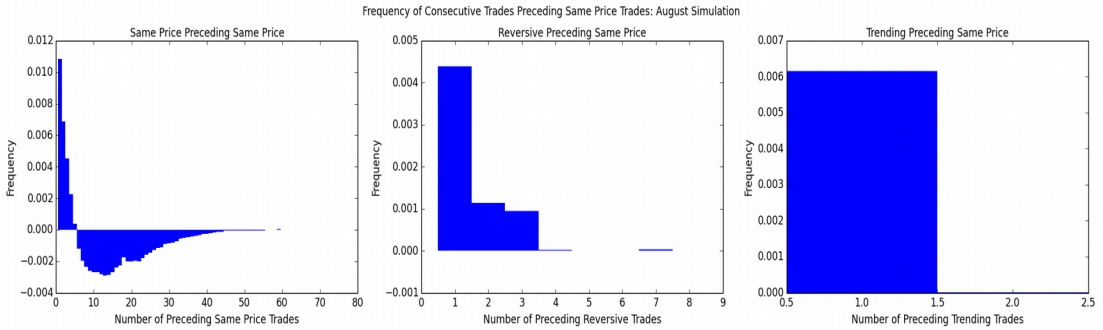

Figure 12: Brownian motion simulation: frequency of trade types preceding same price trades, difference between high volatility and low volatility.

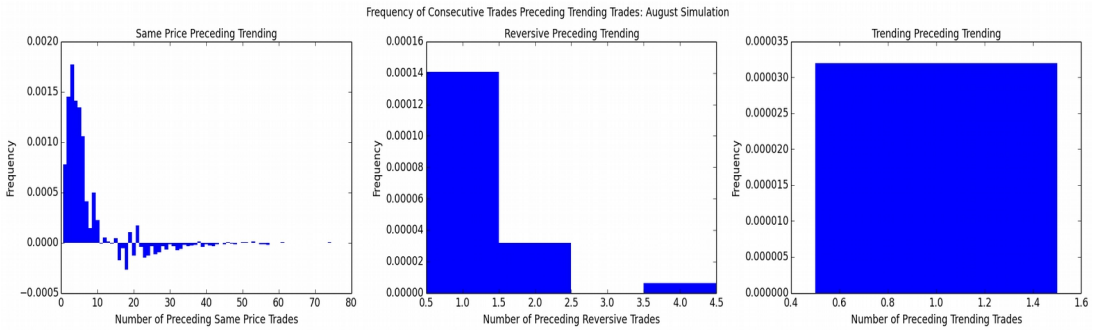

Figure 13: Brownian motion simulation: frequency of trade types preceding trending trades, difference between high volatility and low volatility.

5. Conclusion

The trade direction analysis looks at consecutive past trades to predict the next trade type of the E-mini S&P 500 Futures. Current trades are either at the same price, trending, or reversive relative to the previous trade. The millisecond trade data used in the analysis represents periods of relatively high and low volatility: August 5^{th} – 7^{th} and August 20^{th} – 22^{th}, 2014 respectively. Let $\Delta F(n_{xy})$ be the difference between the relative frequencies with which n consecutive trades of type y precede a trade of type x in the high and low volatility periods. Figures 4 through 6 plot $\Delta F(n_{xy})$ vs. n for each trade type.

The volatility is determined from the direction of price movement. Trade time is used as a proxy for clock time. A virtual price, simulated with Brownian motion, takes continuous values between the bid-ask spread. The realized price is the virtual price rounded to the nearest tick. Changes in the actual price occur when the realized price crosses the spread. The volatility parameter of the Brownian probability density function is determined so that the model has the greatest correlation to the observed directions of price movement. The volatilities obtained for the high and low volatility periods are (4.80 ± 2.02) dollars^2/trade and (4.21 ± 0.71) dollars^2/trade respectively. As a check, equation (12) gives an estimate for the volatility of 5.66 dollars^2/trade.

Figures 12 and 13 plot the modeled $\Delta F(n_{xy})$ vs. n for trades prior to same price and trending trades respectively. Qualitatively, these subplots are similar to their real-data counterparts in Figures 5 and 6 in both magnitude of $\Delta F(n_{xy})$ and the "flip"

behavior at n ≈ 5 in the "Same Price Preceding Same Price" and n ≈ 10 in the "Same Price Preceding Trending" subplots. This flip is due to $F(n_{xy})$ on the low volatility days exceeding $F(n_{xy})$ on the high volatility days at higher n.

This simple Brownian model has limitations. The "Same Price Preceding Reversive" subplot of Figure 11 fails to reproduce Figure 4 because the model does not include bid-ask bounce. Additionally, in Figure 11, and the subplots involving reversive trades in Figures 12 and 13, reversive behavior at low n is not captured. It is evident from the middle subplot of Figures 4 through 6 that reversive trades preceded each trade type more frequently in the low volatility period than the high volatility period. The Brownian motion simulation failed to replicate this.

The Commodity Futures Trading Commission (CFTC) recently filed a complaint against Navinder Sarao and his company Nav Saro Futures limited, alleging price manipulation and spoofing (CFTC Press Release 2015). Spoofing involves a trader putting in larger orders without the intention to fulfill to move the price in a way that is advantageous for the spoofer (Hope et al. 2015).

The CFTC complaint alleges that, in June 2009, the defendants used a modified trading platform to layer four to six large sell orders into the E-mini S&P central limit order book. This "Layering Algorithm" modified the price of the orders to stay three or four levels away from the best asking price, and eventually most orders were canceled. The complaint goes on to state that on May 6, 2010, the day of the Flash Crash, the defendants used the Layer Algorithm for over two hours before

the crash. According to the complaint, the defendants' activities applied close to $200 million worth of downward pressure on the E-mini S&P price and contributed to an extreme order book imbalance which led to the Flash Crash (CFTC Press Release 2015).

The E-mini S&P order book imbalance caused by the spoofing for over two hours prior to the crash was not picked up by the VPIN metric on May 6, 2014. As shown in Figure 1, VPIN begins its ascent about 20 minutes prior to the crash at 2:32pm. Easley et al. (2011) find that the CDF(VPIN) exceeds 95% at 1:08pm. Because it is alleged that the spoofing began half an hour or more prior to this, it is not clear that Sarao's activities were a factor in the flash crash.

Appendix

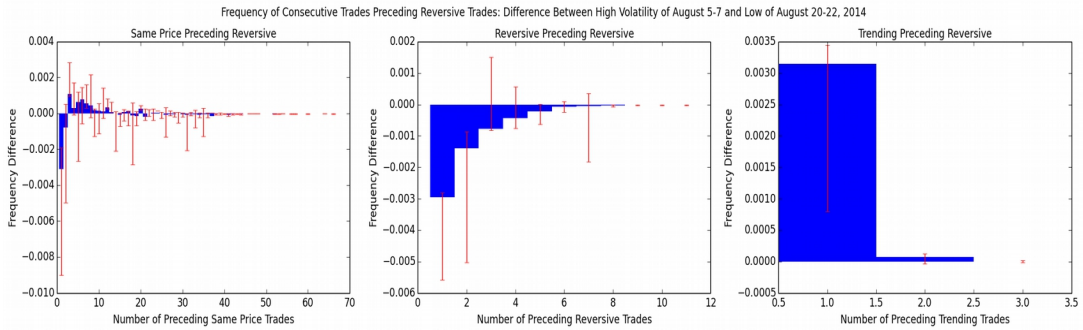

Figure 14: Frequency of trade types preceding reversive trades, difference between high VIX and low VIX for E-Mini S&P 500 in August 2014. Error bars in Figures 14 to 16 are obtained from the 15th and 85th percentiles of 100 bootstrap resamples.

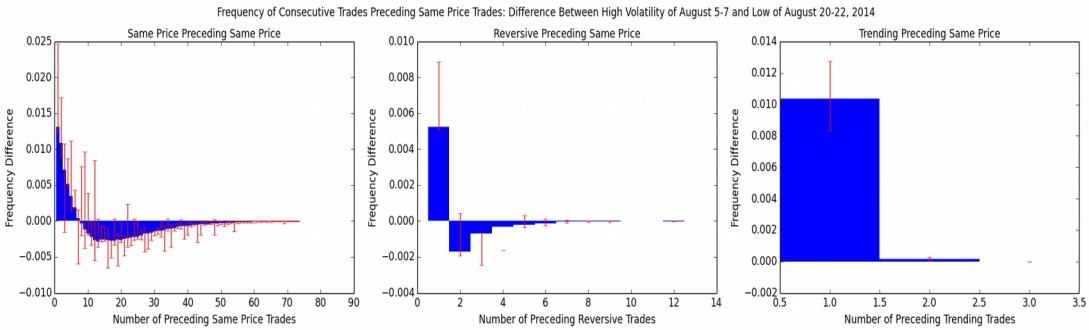

Figure 15: Frequency of trade types preceding same price trades, difference between high VIX and low VIX for E-Mini S&P 500 in August 2014.

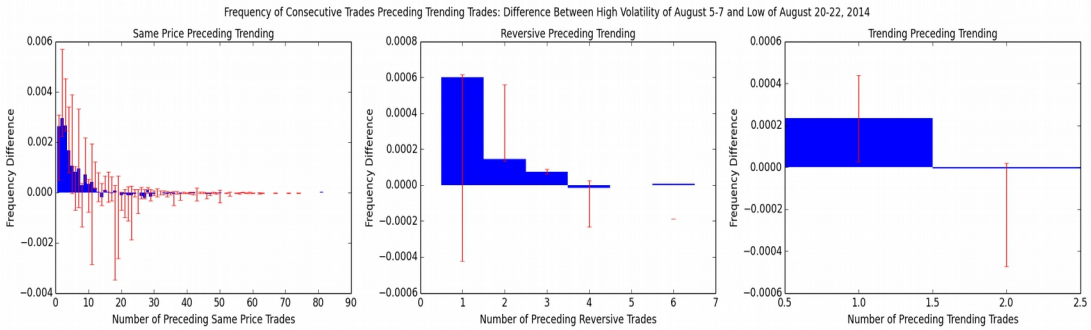

Figure 16: Frequency of trade types preceding trending trades, difference between high VIX and low VIX for E-Mini S&P 500 in August 2014.

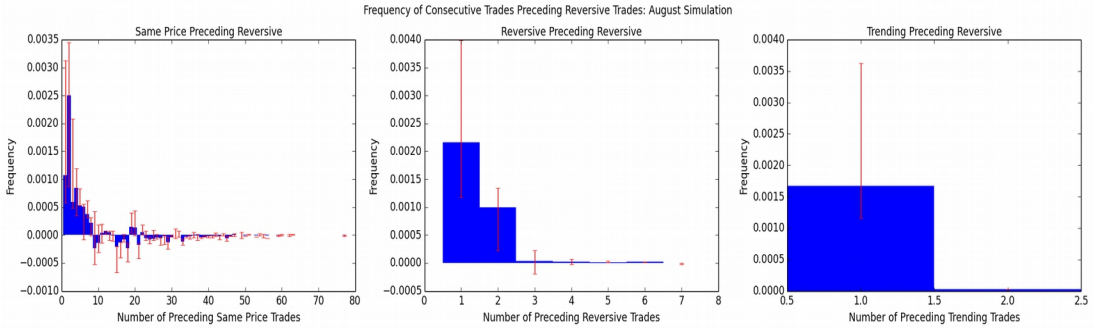

Figure 17: Brownian motion simulation: frequency of trade types preceding reversive trades, difference between high volatility and low volatility.

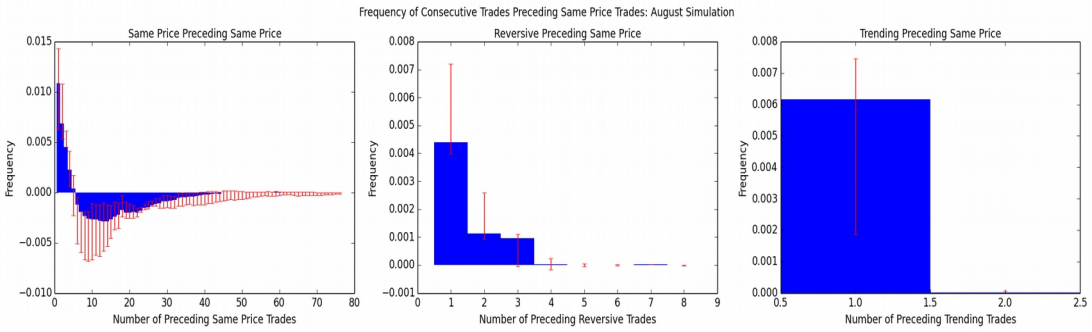

Figure 18: Brownian motion simulation: frequency of trade types preceding same price trades, difference between high volatility and low volatility.

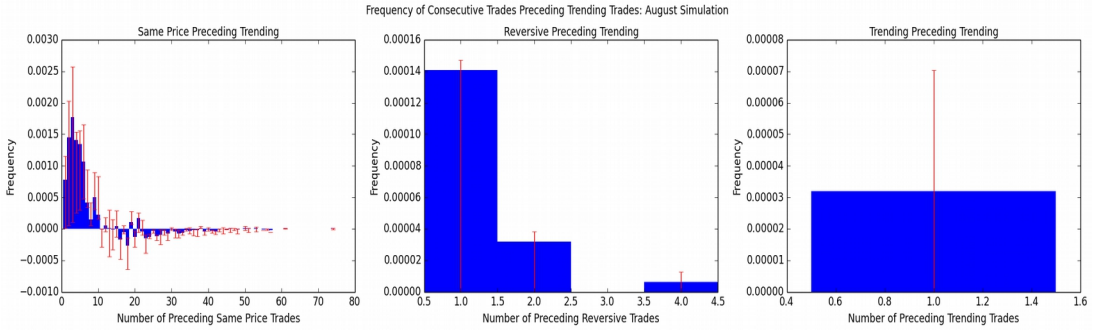

Figure 19: Brownian motion simulation: frequency of trade types preceding trending trades, difference between high volatility and low volatility.

References

- Abad, D.; Yagüe, J. 2012. From PIN to VPIN: An introduction to order flow toxicity. The Spanish Review of Financial Economics 10.2, 74-83.

- Andersen, T.; Bondarenko, O. 2014. Assessing Measures of Order Flow Toxicity and Early Warning Signals for Market Turbulence. Review of Finance, Forthcoming.

- CFTC Press Release. Commodity Futures Trading Commission. Web. 2015 <http://www.cftc.gov/PressRoom/PressReleases/pr7156-15>

- CME DataMine. CME Group. Web. 2015. <http://www.cmegroup.com/market-data/datamine-historical-data/marketdepth.html>

- Easley, D.; Kiefer, N.; O'Hara, M.; Paperman, J. 1996. Liquidity, Information, and Infrequently Traded Stocks. Journal of Finance 51, 1405-1436.

- Easley, D.; Lopez de Prado, M.; O'Hara, M. 2011. The Microstructure of the 'Flash Crash': Flow Toxicity, Liquidity Crashes and the Probability of Informed Trading. The Journal of Portfolio Management 37.2, 118-128.

- Easley, D.; Lopez de Prado, M.; O'Hara, M. 2012. Flow Toxicity and Liquidity in a High Frequency World. Review of Financial Studies 25.5, 1457-1493.

- Easley, D.; Lopez de Prado, M.; O'Hara, M. Bulk Classification of Trading Activity. 2012. Johnson School Research Paper Series 8.

- Hasbrouck, J. 2003. Intraday Price Formation in US Equity Index Market. Journal of Finance, 58(6), 2375-2399.

- Hope, B.; Albanese, C.; Viswanatha, A. 2015. Navinder Sarao's 'Flash Crash' Case Highlights Problem of 'Spoofing' in Complex Markets. The Wall Street Journal. Web. <http://www.wsj.com/articles/navinder-saraos-flash-crash-case-highlights-problem-of-spoofing-in-complex-markets-1430943635>

- Hull, J. 2012. Options, Futures, and Other Derivatives. Upper Saddle River, N.J.: Pearson/Prentice Hall.

- NYSE Custom TAQ. NYSE Market Data. Web. 2014. <http://www.nyxdata.com/Data-Products/Custom-TAQ/>

- Sussman, A.; Tabb, L.; Iati, R. US Equity High Frequency Trading: Strategies, Sizing and Market Structure. 2009. Tabb Group.

- Yahoo Finance! Volatility S&P 500 (VIX). Web. 2014. <http://finance.yahoo.com/q?s=%5EVIX>