Benjamin Greenman, Hersh Mehta
{blg59, hpm36}

CS6850: Final Paper
December 16, 2013

Abstract

Herd behavior in traditional commodities markets is a fairly well-studied subject. The earliest relevant literature dates back to the late 1980's [5], and since then a number of experiments and theoretical models have observed and characterized the behavior. These results established that herding does indeed exist and offer reasonable suggestions motivating and modeling the phenomena. However, high-frequency trading (HFT) has overhauled markets, and little to no published literature explores herding in this context.

We attempt a formal analysis of herding in a HFT setting. Using order book data we attempt to predict price shifts and herd behavior and then discuss our results, their implications, and our hypotheses.

1 Introduction

Financial markets are commonly modeled by a set of independent actors who combine public and private information in order to make investments. Public information is, as the name suggest, freely available to all decision-makers. On the other hand, private information is unique to each. A reasonable assumption when studying trade markets is that a minimum of information sharing occurs but for the most part private information remains so. At no point does all private information become public, thus the market is characterized by inequality and information asymmetry, with players vying to learn their opponents' secrets.

These models carry the assumption that individuals' choices are strongly guided by their private information. However, this assumption has been contradicted by empirical findings. Numerous studies have found that investors do not always act as individuals, guided by private information, but rather exhibit a tendency for collective action. Participants occasionally disregard their own private signal and instead buy with the crowd.

This behavior—ignoring private information and instead basing decisions off competitors' actions—is known as herding. Investors herd when they buy as a crowd, and a market in which this occurs is said to exhibit herd behavior. We are interested in identifying, classifying, and predicting herd behavior in financial markets. This has a variety of applications, most notable of which is to predict the formation of bubbles and prevent their subsequent collapses. Recent financial crises like the Dot-Com bubble and the Subprime Mortgage crisis provide strong motivation for controlling herd behavior. On a more micro scale, being able to predict when a market will herd has money-making potential for trading algorithms. By judiciously providing liquidity to influence herd events, and algorithm can maintain a slight edge on the market.

Herd behavior in traditional commodities markets is a fairly well-studied subject. The earliest relevant literature dates back to the late 1980's [5], and since then a number of experiments and theoretical models have observed and characterized the behavior. These results established that herding does indeed exist and offer reasonable suggestions motivating and modeling the phenomena. Bikchandania & Sharma identify two main reasons why rational human investors choose to herd: for reputation and compensation purposes [5]. Reputation-based herding is motivated by a trader's lack of trust in his or her private signal. If the manager directing the trader to buy or sell is considered unreliable, the individual on the trading floor may follow his competitors for lack of better alternatives. Compensation-based herding is motivated by a fear of doing poorly on the trade floor. Bikchandani and Sharma observe that when a trader's compensation is tied to his or her performance, they will buy with the crowd because this is seen as the "safe" move. Instead of using private information and making a risky, though potentially profitable investment, they hedge bets with the market. Additionally, the authors separate the notion of "spurious" herding, where investors facing similar constraints happen upon the same decision [5, 281]. At any rate, there have been sophisticated advances in the study of herding in traditional markets. We will recount a few of these in Section 2.

However, high-frequency trading (HFT) has revolutionized financial markets, and historical models may no longer apply. At the very least, the time is ripe for new models which are specifically tailored to predict herding in a high-frequency setting. While there is some literature on herding in HFT markets, it is relatively little and not quite as in-depth as past work on herding in traditional markets has been. Our goal in this class has been to analyze the presence, significance, and implications of herd behavior in an HTF setting.

This paper is outlined as follows: Section 2 gives more historical information and recounts the studies we found relevant to our work. In Section 3, we talk about the challenges and peculiarities we forsee in HFT markets. That is, we give a high-level description of our intuitions regarding herding in high-frequency markets. Section 4 contains the main part of this work, introducing our data source and documenting our experiments. Finally we conclude in Section 5 with a final overview of where we have come.

2 Background

Herding, as previously defined, is when traders ignore their private information and instead buy as a crowd. This is a fairly well-known and well-documented phenomena, particularly in traditional (pre-HFT) financial markets, and we recount some of the relevant literature here.

For the most part, works on herding can be put into one of two categories. Either they give an example of herding, identifying and giving reason for an observed behavior, or they present a theoretical model seeking to characterize herd behavior. We use this apparent split to guide our discussion, considering first the works that establish herding as a real phenomena and second the more academic hypotheses and conjectures that have stemmed from thought on herding. However, we make an exception for one notable model, the VPIN model of O'Hara, Easley, and Lopèz de Prado. This model is particularly influential for our work and how we interpret the data, so we give it special attention here.

2.1 Evidence of Herding

Many papers in recent memory document the existence of herd behavior in traditional financial markets. The earliest we know of is Shiller & Pound (1987). They conducted a survey of institutional investors, seeking to determine the influence of direct, personal communications on their decisions [19]. Results strongly indicated that investors follow one anothers' lead, which ultimately led to the realization that they were acting as a herd. Scharfstein & Stein (1990) explore herd behavior among firm managers, providing both rationale and an explanatory model [18, 3]. They also connect their model to stock markets, arguing that the financial sector as another location where their insights apply. Grinblatt, Titman, & Wermers studied mutual funds for evidence of herding and found that a large number of investors are momentum buyers/sellers; that is, they buy past winners and sell past losers [11]. This behavior is significant because it means the entire space of traders eventually converge on the same set of winners and losers. Lakonishok, Shleifer, and Vishny (1991) continue this line of reasoning in their exploration of herd behavior among tax-exempt funds [13]. Other examples include Trueman's study of herd behavior among analysts [21], Golec's observations on herding [10], and Sias's (2004) study of institutional herding [20]. The most recent example we encountered is that of Lin, Tsai, and Sun (2009), which examines herd behavior within the emerging Taiwanese market [14].

In summary, there is a large body of evidence that herding is a widespread phenomena among human traders, and a small but growing pool of work noting herd behavior in HFT data.

2.2 Theoretical Models

On the theoretical side, we have a variety of models that seek to capture the essence of herd behavior. An early, naïve model is that of Bikhchandi (1992), which gave the simplest effective model of herding [4]. They have investors enter the market one-by-one, in turn making decisions of whether to buy, sell, or hold based on their private signal, competitors' actions, and the transaction history. Building upon this model, Orlèan (1995) provided an asynchronous simulated market [15]. However, they gave all players an equal likelihood of copying one another, a misgiving which was corrected by Cont & Bouchad (2000) [8].

Models dealing with HFT data specifically include that of Park & Sabourian (2011) and Bootrz et al. (2012) [16, 6]. These focus on the transactions surrounding a single stock over time. Traders receive a private signal in the beginning of the simulation and the model observes whether they change from buying to selling (or vice-versa) at a later point in time. This model is of special interest to us as we conduct our study of HFT data, and we recall the lessons of Boortz et al. as we study our own dataset.

2.3 VPIN

The Volume Synchronized Probability of Informed Trading (VPIN) model, invented by O'Hara, Easley, and Lopèz de Prado, is a measure of the relative liquidity in a market [9]. Designed to inform market makers in an HFT setting, the model gives a measure of how likely it is that informed trading is occurring within the market. This helps the market maker identify when it is providing liquidity at a loss.

VPIN builds off a previous model, the PIN, by bucketing actions by volume. This is appropriate for high-frequency data, where a standard timestep no longer provides an accurate characterization of the market. Over the course of one second, a vast number of trades might take place. Or, depending on the time of day, extremely few trades might occur—HFT makes the frequency unpredictable. Choosing to group by volume ensures a stable unit of measure and moreover is not dependent on non-observable parameters and updates in stochastic time [9, 4].

This model is essential to our analysis in Section 4, and so we recount it briefly here. In particular, we are concerned, in our experiment, with predicting price movements for the purpose of tracking herd behavior. Using the VPIN metric gives a solid theoretical foundation upon which we can base our own inferences and conjectures.

Trades occur in the market over a period of time intervals; at each interval, traders arrive according to some Poisson process and an information event occurs with probability α.[1] The liquidity provider is an important player. At each time t he holds certain beliefs $P(t) = (P_n(t), P_b(t), P_g(t))$ about whether the news affecting the informed traders was good, bad, or neutral. He updates his beliefs at each time interval according to the expected value of the asset:

$$E[S_i|t] = P_n(t)S_i^* + P_b(t)\underline{S}_j + P_g(t)\overline{S}_i$$

S_i^* is the prior expected value of the asset.

With this value, the liquidity provider can determine the expected Bid $B(t)$ and Ask $A(t)$. These are defined in terms of μ, the proportion of informed traders, and ε, the proportion of uniformed traders, as follows:

$$B(t) = E[S_i|t] - \tfrac{\mu P_g(t)}{\varepsilon + \mu P_g(t)}(E[S_i]|t - \underline{S}_j) \quad A(t) = E[S_i|t] + \tfrac{\mu P_g(t)}{\varepsilon + \mu P_g(t)}(\underline{S}_j - E[S_i|t])$$

Read these as the expected value of the asset conditional on someone wanting to buy the assert from the liquidity provider [9, 10].

Lastly, we have the probability PIN that an order came from an informed trader. This value is key to decisions:

$$PIN = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon}$$

Moving forward, we understand that the market maker is continuously wary of the trades it observes. It constantly tries to determine whether an informed or noise trader made the order, and makes future decisions accordingly. If there seem to be many traders with significant private information, invisible to the maker, it will step out of the market.

[1] The information may be good or bad and the traders may be informed or uniformed, but these details are not essential to our use of the VPIN.

2.4 Comments

In summary, we have seen that herding does indeed exist withing both purely human and transitionary human-and-HF settings. Furthermore, we have discussed an array of models offering guiding insight for predicting herd behavior, each correcting an old belief or identifying a new, previously unconsidered parameter. We turn now towards herding in an HFT setting, describing the challenges we expect to see and our plans for addressing them.

3 Herding in a High-Frequency Setting

High-frequency trading is in-formally defined as the strategy of placing "a large number of orders" that can be "rapidly canceled" and are "held for a short while" [1, 2]. Here "a short while" can range from a few milliseconds up-to half an hour. These trades happen so quickly and so frequently that in the U.S. equity markets the average holding period for a stock is 22 seconds [1, 2]. This means trades can occur in much more rapid succession, giving a succession of small gains the potential to accumulate into large profits.

For the purposes of this endeavor, we assume that herding does occur in HFT settings. This assumption is reasonable given the observations of prior works; it has been shown that herding occurs in emerging high-frequency markets and the rational motivations for herding indicate that optimizing algorithms will identify and take advantage of situations where herding is likely to lead to the best outcomes [14, 6, 5]. However, it is worth mentioning that this assumption is not a given.

Financial data, particularly HFT data, is valuable and difficult to come by. Free sources, aside from what governments demand, are practically non-existant as proprietary services run a lucrative business tracking and distributing information on exchanges. Yet even these suffer insufficiencies. For one, some markets evade collection. European markets are sufficiently transparent but U.S. markets are privatized, with a significant portion of trades occurring in dark-pools, and thus more difficult to get information for [3]. Foreign exchange markets present a different challenge in that there is no central governing body documenting all exchanges [17]. Within the Forex markets, trading occurs amongst various liquidity providers, and there aren't individual exchanges, such as the NASDAQ or Dow Jones for stocks, where these transactions are cleared. So our research is limited by the sources available.

For data which is collected, one faces the additional problem of completeness. There are varying definitions of "high-frequency"; instead of milli- or micro- second data, some purportedly HFT sources offer data on the order of tenths of seconds or even whole seconds. Another issue is that quantities are not always reported alongside prices, and the buyer and seller are frequently anonymized. Without quantities, the interesting predictions we can do are sorely limited. Anonymization might not be such an issue, but traders in many cases are not even given unique identifiers. The data may not give a means of attributing the purchase of a few different securities to the single buyer who purchased them all.

Fortunately, one member of our group had an account with Interactive Brokers and we used this account to gather our data. Interactive Brokers (IB) is a service providing a web interface into securities markets [7]. Founded in 1977, the site "has grown internally to become one of the premier securities firms with over $4.8 billion in equity capital." For our purposes, Interactive Brokers offered price and quantity data approximately every five milliseconds. This was frequent enough for our purposes, and furthermore the data associated quantity information with every price. Hence IB provided us a record of recent transactions and a bit of information upon which to base conjectures.

That said, this data was certainly not complete. Though we were excited to have frequent updates and quantities, IB only provided and updated an order book: a summary of the current bids and asks. An order book consists of rows of the current bid and ask prices. These rows are ordered by price. The highest bid and the lowest ask are at the top of their respective columns. Associated with these prices are the quantities participants are seeking to work in. These quantities are not associated with individual traders but rather with the price as a whole. This make accounting for transactions difficult, as we cannot tell who participated in a given trade. An additional complication is the way in which data is reported. Changes to the order book appear one at a time, in sequence. That is, if trader A buys q units of trader B's stock at price

p, the order book is first changed to show that the aggregate buyers are seeking to buy q less stock at price p and second to show that the sellers have q units less to offer at price p. The transaction does not occur in one continguous step, thus making it impossible to discern market entry/exit from actual trades. This is a source of concern because in HFT, over 90% of market actions are entries and exits [2]. Unfortunately, this is a major source of inaccuracy in our analysis; we revisit this problem later. Furthermore, we have no means of associating one trade with a certain market participant. Changes in price and quantity appear without an identifier, thus, we cannot apply the models of, for example, Park & Sabourian or Boortz et al., which monitor individual traders over a period of time [6, 16].

Regardless, Interactive Brokers provided a sufficient data source for our purposes. Specifically, we looked at the spot EUR/USD exchange rate. After obtaining full access to the API and writing the initial plumbing code to interface with their quote dissemination servers, we began collecting data in early November. Although these early collections were admittedly sparse, they were useful for getting a general feel for our target market, which was foreign exchange currencies between the United States, Asia, and Europe.

Beginning around Thanksgiving break, we switched gears and began extracting data daily and for extended periods, building weekly sets of data. We base our experiments and presentations in Section 4 off one such week's worth of data, collected starting December 3.

3.1 Conjecture

High-frequency data differs from traditional data in that it moves much faster and it generated by precise machines, but the two still share the same underlying structure. At the end of the day, the objective is to make money given limited information about the market. Each agent has access to public information as well as their own private signal; a key objective of agent A is to learn as much as possible about the signals of the other agents so to make the most informed decision possible. Thus if we begin with what we know about traditional trading settings and add alterations to handle the peculiarities of HFT we can determine a predictor for price movements and, ultimately, herding.

Considering traditional markets, we recall one reason participants would herd was to supplement insufficient or untrusted private information [5, 291]. In other words, the participant relies more on outside data when less confident about the private signal. Conversely, participants who are very confident about their private signal are unlikely to herd. In fact, given sufficient confidence and opportunity for profit they may make larger-than-normal trades. Hence we take large-quantity trades to be an indicator of herding. Applying this intuition to HFT markets, we note that the average quantity of high-frequency trades should be low. Trading algorithms make money by providing liquidity to other participants [12]. They make money through many small, low quantity trades; therefore, we expect to see a low average trade quantity in an HFT market. This characteristic in mind, a large quantity trade would be quite out of the ordinary—something to take note of. Here we have the basis for our first conjecture regarding high-frequency data:

High-quantity trades significantly influence HFT price trends

That is, we expect trading algorithms to take note of and modify their position in accordance with the high-volume, presumably-high-confidence trades. For an algorithm bids heavily on a particular future is an uncommon event, worth paying attention to. We hypothesize that traders in the HFT market will follow this decision pattern. If they are at all inclined to herd, they will do so following the direction of large-quantity trades.

4 Experiment

We start by trying to re-create an observation that Almgren made within the Intrest Rate futures markets. While looking at the 1 year Eurodollar interest rate futures, he observed that market "micro-price" helped predict large price movements in the market. Specifically he looked at the best bid/ask prices, and the volume weighted average price and observed that while the market bid/asks didn't move frequently, change of trend in the average price forecast an impending shift in the market.

Below, we plot the EUR/USD Spot FOREX in an attempt to predict price movement.

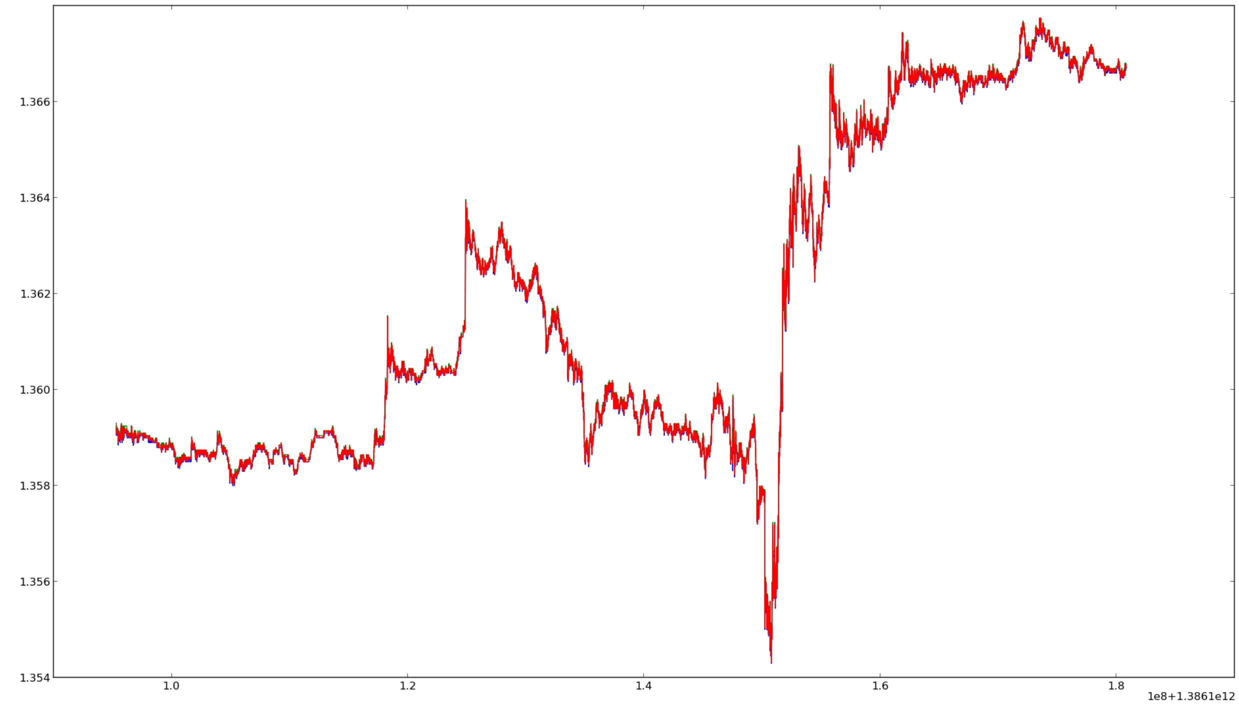

Figure 1: Bid, ask, and Volume-weighted price for EUR/USD Spot over 24-hour period. X-axis is time in milliseconds since EPOCH

Though difficult to read from Figure 1, the blue line is the best bid and the green line is the best ask. The red which overpowers the other two is the volume-weighted average.

Figure 2 is an enlargement of the large spike in price in Figure 1 preceding the sustained down-trend during the middle of the graph. Figure 2 shows us that we do not have the nice market micro-structure in the EUR/USD Spot which existed in the Eurodollar interest rates. Specifically, there is a lot of noise, and small fluctuations in the plot leading up to the large spike in figure two. This is because the Eurodollar has a large tick size, whereas the EUR/USD does not. The minimum tradeable unit is significantly larger in the Eurodollar, so prices do not move an entire tick size as frequently. Our graphs for every other day on which we collected data are similarly difficult to reason about. They exhibit patterns similar this one; that is, they are often characterized by a sharp rise or fall at one location. Thus we need a more descriptive picture.

In an attempt to generate a more informative picture and to test our conjecture that trade quantity motivates herding, we plot points of high quantity on top of Figure 1. The results are in Figure 3.

Points of high quantity are those which the quantity was over 2 standard deviations away from the mean. These $(time, price)$ points were then overlaid on the line graph and shifted up or down depending on whether they were for the ask book or the bid book. In all, there are 9,934 such outlying points, summarized in Figure 4.

Before analyzing the shifts and trends in this graph, we step back and review the premises for our model. Trading is happening in a high-frequency setting where informed traders, uniformed traders, and liquidity providers interact. We focus on the behavior of the HFT liquidity provider—these are the algorithms that

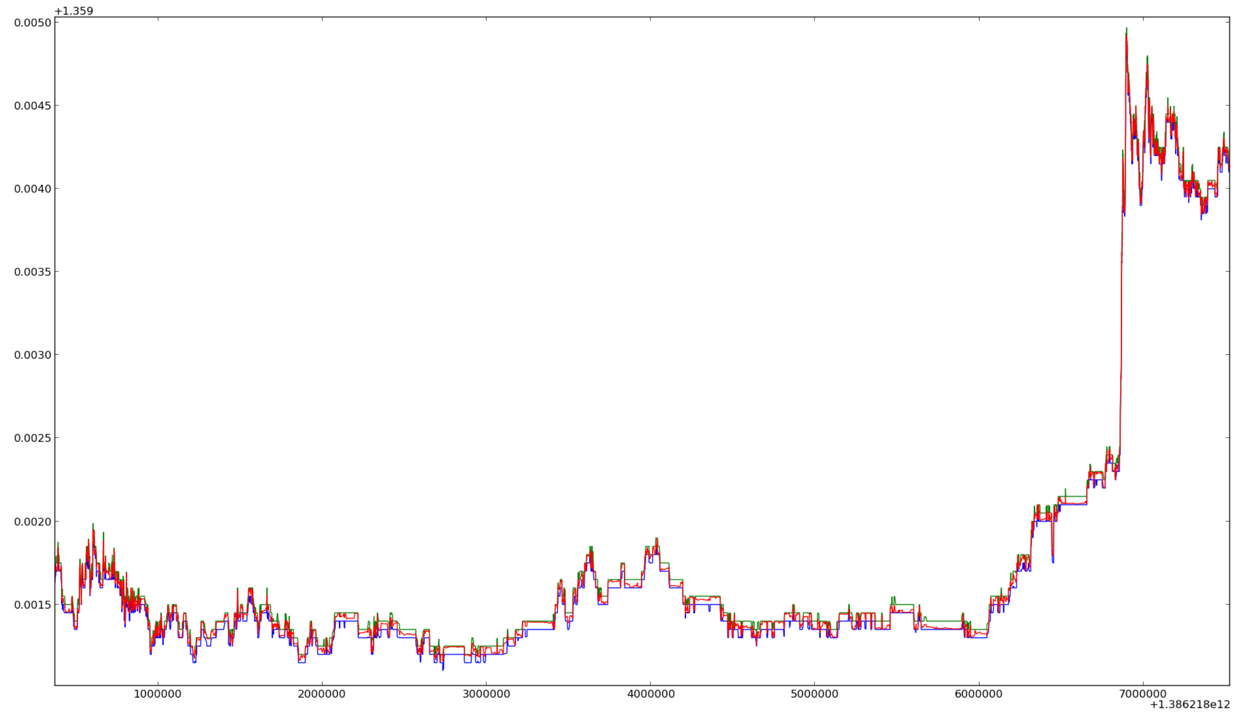

Figure 2: Portion of Figure 1, enlarged

are profiting through high-quantity, low-prices trades that happen at very high frequency. These liquidity providers do best in a fair, stable market, where drastic price shifts are rare. As such, they must continuously monitor their competitors (the informed traders) to guard themselves against private information within the market. If a trader gains a large information advantage, the liquidity provider is liable to suffer. Therefore providers use some metric like the VPIN to monitor the market and identify when to withdraw.

On to the graph, we discuss it in terms of three segments. Segment 1 begins at time 0 and continues until time 3.5. Segment 2 begins at time 3.5 and continues until time 4.5. Segment 3 begins at time 4.5 and continues until time 6. Segment 4 is the large spike we see from time 6 until time 6.5, and lastly Segment 5 begins at time 6.5 and continues until the end.

Throughout this discussion, we refer to outlying price points and liquidity providers interchangeably. This is because liquidity providers are the players who are making these markets—the support behind the large bid and ask quantities that define the outliers.

At the beginning, in Segment 1, we see that the prices are relatively stable. Prices fluctuate and even jump at one point, but the overall trend is a flat line. Accordingly, liquidity providers are active in the market as evidenced by the fair number of outlying points on both the bid and ask side of the bid/ask line.

This trend continues until we reach the downturn at Segment 2. Here we have a large density of outliers on both sides of the graph, indicating that liquidity providers are making offers in an attempt to stabilize prices. On the ask side we see the greatest density, which makes sense because the price is dropping and sellers prefer a higher price. The liquidity providers are losing money throughout Segment 2 as they fight an apparently unstoppable price trend. From our perspective as observers, we can look at this downtrend and hypothesize that a group of informed traders has some information advantage that is predicting the future prices for them. Liquidity providers have no such ability to see the future, and so they monitor the price history and plan their next move. After the small spike in the middle of this segment, we see liquidity providers leave the market.

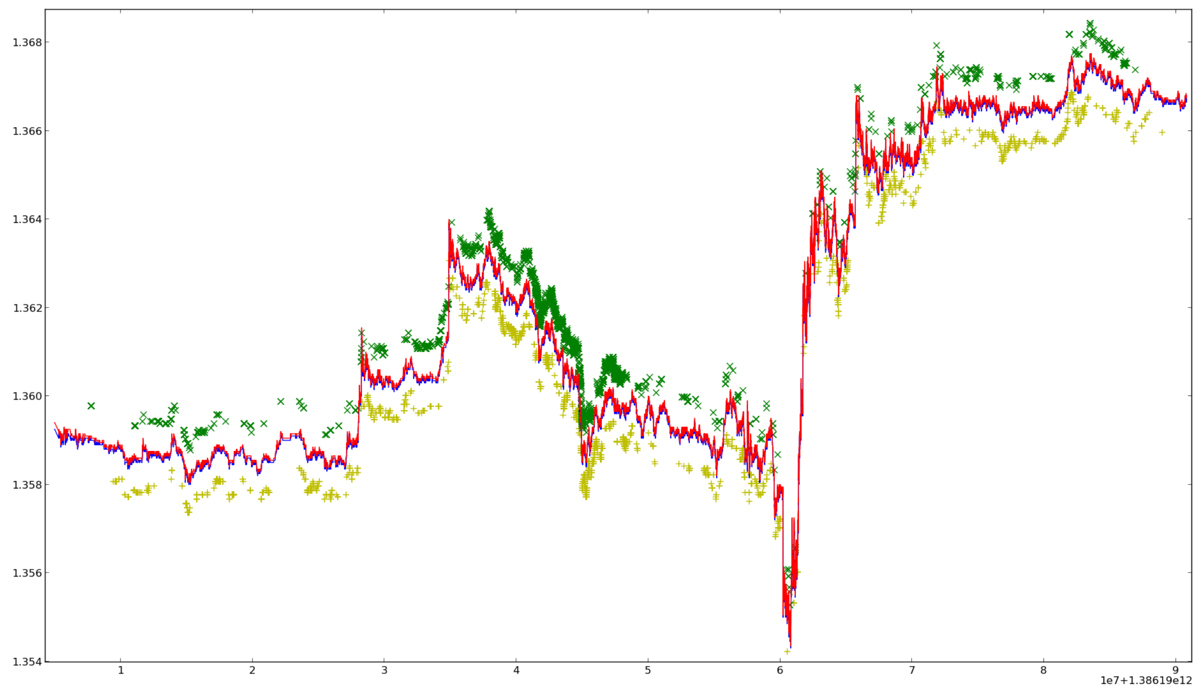

Figure 3: Price along side points indicating two quantities 2 std. dev. from the mean on bid book (Yellow) and ask book (Green).

	Bid	**Ask**
Outliers	4,525	5,409
Average Quantity.	8,471,857	9,656,184
Std. Deviation	6,196,046	6,432,082

Figure 4: Statistics on outliers from Figure 3

Their VPIN signal (or similar indicator) flagged the market as too dangerous to remain in—the likelihood that informed traders are making decisions has risen significantly. The signal produces this flag because it sees that the volume clock is out of sync with the price clock, which supports our hypothesis that volume plays a role in herding behavior. Unfortunately we cannot use this rule to predict herding as we had hoped, nevertheless it helps us identify herd behavior as it happens.

Returning to the graph, we see liquidity providers continue to leave the market throughout Segment 3. By the time we exit Segment 3, there is very little liquidity within the market; there is hardly any support for prices on the bid and ask sides. Hence movements in both directions are quick and exaggerated. This accounts for the spike we see in Segment 4. Without liquidity providers to support it, the market swings out of control momentarily. Finally, in Segment 5 the market stabilizes and liquidity providers return, yielding a pattern similar to what we saw in Segment 1.

Our initial objective was to predict herding behavior through careful monitoring of volume. We hypothesized that volume was an important contributor to herding. Unfortunately, our findings do not support this claim. That said, although quantity is not a predictive indicator, it does work as an on-line indicator for price shifts. Using the VPIN model, we can predict that liquidity provision will decrease after a herd-like

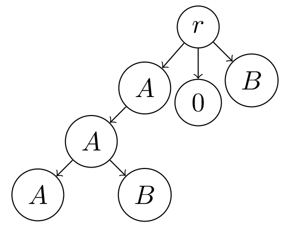

Figure 5: Example prefix tree, constructed from sequence $\{A, A, A, 0, B, A, A, B\}$

event. This is exactly what we saw in Segment 3 of Figure 3

As a result of the decrease in liquidity provision, we conclude that future price movements will be drastically over-exaggerated until the HFTs return to the market and start providing liquidity again. While it is difficult to predict the direction of this movement, we can nevertheless find a profitable trading strategy based on trading a derivative of the security (e.g. an option on the EUR/USD Spot FX) we are observing which will increase in price due to increase in volatility.

4.1 Prefix Tree

Unsatisfied by our initial attempt at characterizing herd behavior, we decided to take a different approach at identifying parameters that would predict price changes. Instead of working from our intuition or experience, which had thus far been insufficient at describing algorithms (or at least at describing the data available to us) we decided to apply machine learning techniques to identify parameters of interest. Our goal was to say something meaningful about trends in the bid and ask prices. To this end, we constructed a simple prefix tree modelling price patterns.

The prefix tree was constructed as follows. First, we extracted from the order book the state of the top bid and top ask at each point in time. Ignoring quantity and caring only for the price point, we stepped through the data and marked a A, 0, or B if the price had raised, stayed the same, or fallen since the previous timestep. This left us with a long string of prefix information. We then iterated over this string, constructing a tree of each newly encountered prefix. This process is best explained with an example.

Given the string $\{A, A, A, 0, B, A, A, B\}$, we begin with a tree consisting of only the root and consider the first element. Reading an A, we construct a new node labeled "A" and having count 1 and place a directed edge from the root to this node. This means we have found a new prefix. Having created a new node, we return to the root. Reading the next element, an A, we follow the edge to the node labeled "A" that we just created and increment its count to 2. This means we have followed an old prefix. Next we consider the third sequence element, A, and construct a new node labeled "A" and having count 1 as a child of the current node. Jumping back to the root because we just created a new node, the next two sequence elements result in the children "0" and "B" being added to the root. The last three elements represent one new sequence, which we reach by traveling from root to "A", incrementing the count to 3, from "A" to the next node labeled "A", incrementing its count to 2, and finally creating a new node "B" as a child of this node. Figure 5 shows the entire tree created following the above sequence.

In the presence of herding, we would expect to see a strong price trend and thus an unbalanced tree. The paths present in the final tree give the space of all encountered prefixes, which will be skewed left or right if herding occurs. Additionally the counts, which give the probability of moving the a node to any one of its children, should be skewed if herding occurred.

Our first attempt in constructing a prefix tree over an entire week's worth of data gave us a structure with a huge proportion of zeros. Almost 90% of the sequence elements were zeros, leading to a tree with a disproportionately high number of paths made almost exclusively of zeros. Figure 6 summarizes this data by giving the probability of choosing a zero from any node in the graph. To generate the plot we simply examined the at most 3 children of each node and compared the probabilities of advancing to each.

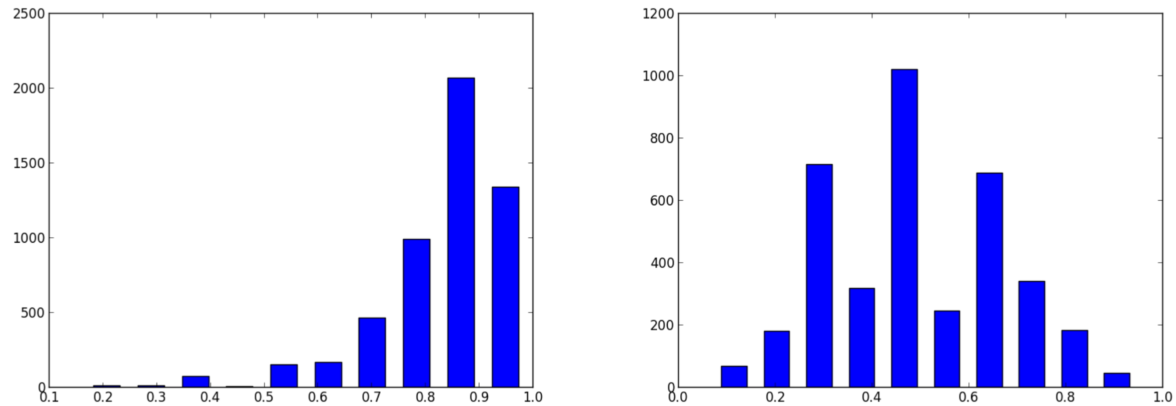

Figure 6: Prefix tree findings: the left is the probability of choosing 0 from a given node in the unpruned tree. The right is the probability of choosing A in the pruned tree.

At any rate, the tree full of zeros had no use to us as a prediction mechanism. We could not say anything meaningful about price trends if the vast majority of events led to no change at all. But looking back, finding this many zeros in the tree was not unexpected. We knew ahead of time that the order book changed in small ways between timesteps, and that many of these did not affect the best bid/ask. For instance any simple transaction between one of the other, non-top securities would lead to two order book changes (one documenting the buyer, the other documenting the seller) neither of which affected the top row. So we decided to prune the prefix string, omitting all the zero entries.

With this new, pruned input string containing only the directions A and B that the order book shifted between rounds, we expected to see a much more insightful predictor of prices. Having filtered the inconclusive data, asymmetries within the tree would now become apparent, even over-characterized. Yet our findings were decidely null.

The second plot in Figure 6 shows the probability of choosing A from any node in the tree. It follows a perfectly normal distribution, indicating one of two things. Either there was no herding throughout the week or our data failed to capture any interesting characteristics of the market. We fear it was the latter.

5 Conclusion

The study of financial markets and their intricacies is an intense and challenging area of study. Data is fiercely guarded, sold at a premium, and moves quickly. If not on guard continuously, significant yet subtle patterns can easily escape detection.

We attempted a disciplined study of financial markets, trying to characterize and predict herd behavior in an HFT setting. To this end we gathered weeks' worth of order book data and attempted fitting it to a few hypotheses and analysis techniques. Being able to predict herd behavior in HFT setting is certainly an interesting question, and we maintain that there must be some rational explanation for it, but we have not uncovered it. Collecting data, as noted above, was a challenge in itself. However, even when we had enough information to begin running experiments we met almost perpetual confusion and disappointment. Pronounced herd behavior does not take place on a daily basis. Intra-day trends take a more nuanced form and are harder to describe, much less to separate from noise. We attempted to correct this by examining data over the course of a week, but still no clear patterns emerged. This effort was perhaps foiled by our dataset having price information three separate markets: American, Asian, and European. Regardless, and although our intuitions were strong and based upon published and accepted findings for traditional markets, our results were not supportive of our hypotheses.

While we were unable to get a future predictor for herding behavior, we identified bid/ask quantity to be an important driver in the behavior of future market price and liquidity. By using the VPIN model as a foundation, we were able to draw insightful conclusions from our dataset and identify a strategy for predicting the increase in volatility of future price movements that occur as a result of liquidity providers exiting the market after having observed herd behavior which was started by an influx of informed traders.

Hence we come away from this project with valuable experience working with securities data and attempting to build models and find patterns in confusing data.

References

[1] Jonathan Ahlstedt and Johan Villysson, *High frequency trading*, (2012).

[2] Robert Almgren, *Interest rate futures*, Quantitative Brokers.

[3] Laura Nyantung Beny, *Do insider trading laws matter? some preliminary comparative evidence*, American Law and Economics Review **7** (2005), no. 1, 144–183.

[4] Sushil Bikhchandani, David Hirshleifer, and Ivo Welch, *A theory of fads, fashion, custom, and cultural change as informational cascades*, Journal of political Economy (1992), 992–1026.

[5] Sushil Bikhchandani and Sunil Sharma, *Herd behavior in financial markets*, IMF Staff papers (2000), 279–310.

[6] Christopher Boortz, Simon Jurkatis, Stephanie Kremer, and Dieter Nautz, *Herding in financial markets*, Sonderforschungsbereich 649: Ökonomisches Risiko-(SFB 649 Papers), Humboldt-Universität zu Berlin, Wirtschaftswissenschaftliche Fakultät.

[7] Interactive Brokers, *https://www.interactivebrokers.com/en/index.php?f=564*, Dec 2013.

[8] Rama Cont and Jean-Philipe Bouchaud, *Herd behavior and aggregate fluctuations in financial markets*, Macroeconomic dynamics **4** (2000), no. 2, 170–196.

[9] David Easley, Marcos M López de Prado, and Maureen O'Hara, *Flow toxicity and liquidity in a high-frequency world*, Review of Financial Studies **25** (2012), no. 5, 1457–1493.

[10] Joseph Golec, *Herding on noise: The case of johnson redbook's weekly retail sales*, Journal of financial and quantitative analysis **32** (1997), no. 03, 367–381.

[11] Mark Grinblatt, Sheridan Titman, and Russ Wermers, *Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior*, The American economic review (1995), 1088–1105.

[12] Xin Guo, *High-frequency trading*, (2012).

[13] Josef Lakonishok, Andrei Shleifer, and Robert W Vishny, *The impact of institutional trading on stock prices*, Journal of financial economics **32** (1992), no. 1, 23–43.

[14] William Lin, Shih-Chuan Tsai, and David Sun, *What causes herding: Information cascade or search cost?*, Available at SSRN 1540443 (2010).

[15] André Orléan, *Bayesian interactions and collective dynamics of opinion: Herd behavior and mimetic contagion*, Journal of Economic Behavior & Organization **28** (1995), no. 2, 257–274.

[16] Andreas Park and Hamid Sabourian, *Herding and contrarian behavior in financial markets*, Econometrica **79** (2011), no. 4, 973–1026.

[17] Lucio Sarno and Mark P Taylor, *Official intervention in the foreign exchange market: is it effective and, if so, how does it work?*, Journal of Economic Literature **39** (2001), no. 3, 839–868.

[18] David S Scharfstein and Jeremy C Stein, *Herd behavior and investment*, The American Economic Review (1990), 465–479.

[19] Robert J Shiller and John Pound, *Survey evidence on diffusion of interest and information among investors*, Journal of Economic Behavior & Organization **12** (1989), no. 1, 47–66.

[20] Richard W Sias, *Institutional herding*, Review of Financial Studies **17** (2004), no. 1, 165–206.

[21] Brett Trueman, *Analyst forecasts and herding behavior*, Review of financial studies **7** (1994), no. 1, 97–124.