Master Thesis

Overreaction and Noise Trading

Diego Pérez Gatti
Master in Economics

Advisor: Vicente Royuela Mora

UNIVERSITAT DE BARCELONA
SCHOOL OF ECONOMICS

Abstract

This master thesis examines whether the opening price of a trading session is a result of overreaction generated by the interaction of noise traders. In order to study the overreaction and noise trading, we analyze the price retracement pattern of the Ibovespa futures contract. We also perform an econometric analysis, using probit and logit regressions, to see if and how the extent of price movement, volatility and trading volume affect the price retracement and consequently the overreaction. We find evidence that, the opening price is an inefficient price level result of noise trading. We also find significant effects of our considered explanatory variables: move length affects negatively, while volatility and trading volume have a positive impact on overreaction.

Keywords: Overreaction, noise trading, efficient market hypothesis, arbitrage, Ibovespa, futures contract, price levels, probit, logit

Deadline: 1st September 2017

1. Introduction

Asset price behavior has important economic implications. The understanding of price behavior, can improve asset allocation and is a key element when addressing to one of the most important concepts for the Economic Science and Finance, Market Efficiency. There is a vast literature that analyses whether asset returns derived from the interaction of rational agents pursuing self-interest that fully incorporate all public information (see Cohen, 1966; Fama, 1965, 1970; Lo, 2004; Malkiel, 1985), which is widely known assumption of the classical theory of the Efficient Market Hypothesis (EMH). EMH advocates that, in efficient markets, competition will cause that public information regarding value will be fully and instantaneously reflected in prices (Fama, 1965). Therefore, an efficient price level must reflect all information available to the market. Moreover, prices are an equilibrium derived from the interaction of rational agents pursuing self-interest. Authors that defend market efficiency (see Fama, 1970; Samuelson, 1965) assume that price changes are, in the end, completely random and therefore, unpredictable, an assumption that will be challenged in this work.

On the other hand, the last thirty years witnessed the growing of a set of theories that advocate that agents are subject to "cognitive bias" (Tversky & Kahneman, 1975). As stated by Griffin and Tversky in an article from the Cognitive Psychology Journal, "One of the major findings that has emerged from this research is that people are often more confident in their judgement than is warranted by the facts" (Griffin & Tversky, 1992, p. 411). The same proposition is defined for financial markets: "In violation of Baye's rule, individuals do not consider prior probabilities when making their assessments, but rather arrive at subjective probabilities of occurrence based on how similar, or 'representative', the event is to their preconceived notions" (Madura & Richie, 2004, p. 92). Applying this concepts to asset valuation, prices can consistently deviate from its intrinsic value that cannot be explained by fundamentals. Therefore, showing that market agents are not fully rational.

Following the idea that agents are not fully rational, we arrive to the concept of noise, which can be described as the implications of human behavior that does not fit in the conventional notions of optimization (Black, 1986). He goes on saying "Noise makes financial markets possible, but also makes them imperfect. If there is no noise trading, there will be very little trading in individual assets. People will hold individual assets, directly or indirectly, but they rarely trade them" (Black, 1986, p. 530). Following his argument, noise provides liquidity and mispricing that makes trading desirable. Departing from this view, the concept of "noise trading" was developed, which states that, uninformed agents use irrelevant information as basis for their trading decisions.

Building on these ideas, using the framework of Madura & Richie (2004) as theoretical reference, we hypothesize that in general agents tend to overreact to public information released after the close of previous trading session generating predictable price change patterns. We develop an empirical study in which we analyze whether the first price of trading session is a result of overreaction generated by the interaction of noise traders. We also study the overreaction asymmetry argument, regarding positive and negative movements (see Caginalp, Desantis, & Sayrak, 2014; Da Costa, 1994; Hibbert et al., 2008; Kim, 2006).

In order to study the overreaction and noise trading, we analyze the price retracement pattern conditioned to different trigger sizes, which are used to qualify the price movement as overreaction. We also perform an econometric analysis, using probit and logit regressions, to see if and how the extent of price movement, volatility and trading volume affect the price retracement and consequently the overreaction.

We find evidence that, the opening price is an inefficient price level result of noise trading. We also find significant effects of our considered explanatory variables: move length affects negatively, while volatility and trading volume have a positive impact on overreaction.

The remaining sections are organized as follows: first the literature review, second empirical framework, third empirical analyses and results, and finally the conclusions.

2. Literature Review

2.1 Market Efficiency

A basic tenet of finance, is that an asset is a right in future cash flows, therefore the value of an asset depends on these cash flows. The most widely known theories in asset pricing try to determine the fundamental (intrinsic) value of an asset, which has to be distinguished from its price that we can observe in the market. According to Adam Smith (1776), the observed price or market price is subject to the demand and supply of the asset and therefore, can significantly deviate from its intrinsic value in the short run, but converging to the fundamental value in the long run. This relation between intrinsic value and market price is the start point of the market efficiency concept. A market is said to be efficient if expectations and consequently prices, correctly incorporates all information available. The EMH, which is a core assumption in classical finance, states that prices should "fully" reflect all public information available to the market, and information regarding a security is immediately incorporated to its price (Fama, 1970). "In an efficient market, *on the average*, competition will cause the full effects of new information on the intrinsic value to be reflected *instantaneously* in actual prices" (Fama, 1965, p. 7). In other words, any price deviation from its true value would be exploited by informed traders (*Arbitrageurs*) trying to maximize profits leading to the equilibrium price again, which will be a good estimate of its intrinsic value (Fama, 1965). More precisely, Eugene Fama (1970) divides this concept into three different categories conditioned to the nature of information subset: weak-form, semi-strong form, and strong-form of market efficiency. While the weak-form considers just historical prices, the semi-strong form tests if prices "fully" adjust to other new information that obviously became public (macroeconomic events, earnings reports, etc.). Finally, the strong-form of market efficiency accounts for the possibility of a specific group having monopolistic access to relevant information for price formation. He concludes that there is evidence that supports the EMH and therefore, "for the purposes of most investors the efficient markets model seems a good approximation to reality (Fama, 1970 p. 416)". In general, building in the EMH, we can assume that market prices are a result of the interaction of rational agents with complete information pursuing self-interest. The Nobel Laureate Paul Samuelson, through his article with self-explanatory name, "Proof that Properly Anticipated Prices Fluctuate Randomly" (Samuelson, 1965), was one of the first proponents to develop a consistent theoretical framework to show that stock prices evolve according to a random walk stochastic process. He starts the article proposing that in competitive markets, due to its nature, if one could be sure that a price would rise, it would have already risen and therefore, competitive prices would follow a random walk with no predictable bias. However, he emphasizes that in an empirical setting, such a theoretical proposition would be very hard to prove. In his words: "from a non-empirical base of axioms you never get empirical results" (Samuelson, 1965 p. 42) Andrew W. Lo perfectly summarizes above ideas about the EMH, "In an informationally efficient market, price changes must be unforecastable if they are properly anticipated, i.e., if they fully incorporate the information and expectations of all market participants". He continues with "the more efficient the market, the more random the sequence of price changes generated by such a market, and the most efficient market of all is one in which price changes are completely random and unpredictable" (Lo, 2004 p. 2). However, in the last three decades occurred several market events, which can be defined as large price deviations (from its intrinsic value) for a relatively long-time period, that casted reasonable doubts on the above assumptions. Most of these events consisted in stock market bubbles and crashes such as the US stock market crash of 1987 (Dow Jones index plunged over 20% in a single day), the Japanese stock market bubble in early 90's (Nikkei index almost doubled in two years, and one year later it was at

50% of its peak), and the bubble of US high tech stocks (driven by the internet), after reaching its peak in 2000, saw some of the largest and profitable companies lose 80% of its value three years later.

2.2 Mean Reversion and Overreaction in asset returns

"What goes up must come down" firstly advocated by Isaac Newton in 1670, turned out to be a stylized fact about stock markets returns (Hillebrand, 2003). This means that reverting patterns are inherent to the human behavior, which is a mixed function of emotion and rationality. The Nobel Prize Robert Shiller, in his book, Irrational Exuberance, named after the speech of Federal Reserve ex-chairman Alan Greenspan, proposes that stock market prices overreact to changes in valuation. Alan Greenspan, on December 1996, used the term "Irrational Exuberance" to describe the behavior of stock market investors (Shiller, 2014) suggesting that investors are not fully rational.

Several studies present mean reversion as a consequence of overreaction, i.e., reversion of price behavior after large price movements, which also present the concept of underreaction, which can be understood as positive (negative) cumulative returns following large positive (negative) price movements (see Madura & Richie, 2004). As noted by Maheshwari & Dhankar (2014) the overreaction hypothesis states that in case of good news or periods of euphoria (optimism) as well as in case of bad news, market prices may react excessively, deviating from its value. However, after some time the stock prices revert to its fundamental value, suggesting an overreaction. The first work to consider the overreaction effect was De Bondt & Thaler (1985) in which they study the empirical case of US stock market. They found evidence that important news events are followed by abnormal price movements, e.g. stock prices tend to overreact. Using NYSE monthly data (1926-1982) they created two different portfolios, winners and losers, which consisted in stocks that experienced extreme appreciation (winners) and depreciation (losers) in its prices in the past five years. Moreover, they showed that contrarian strategies of selling "winners" and buying "losers" stocks, based on overreaction effect can generate excess of return, which is a significant blow in the EMH, a standard assumption of classical theory.

Various studies suggest that excess of returns can be earned by investors exploiting mean reversion behavior or overreaction (Spierdijk, Bikker, & Hoek, 2010). Jegadeesh (1990) and Lehmann (1990) for example, generated abnormal returns also applying contrarian strategies (selling winners and buying losers), in this case based on short term reversals. Some works have found evidence that returns tend to show long-term mean reversion bias (negative autocorrelation), while in the short run they tend to show small positive serial correlation (underreaction) as presented by Poterba and Summers (1988). Other studies as Balvers, Wu, & Gilliland (2000) advocate that contrarian strategies generates abnormal returns, despite mean reversion evidence over long horizons is not conclusive for US stock prices, which may be a result of an absence of reliable long time series. Nevertheless, Jegadeesh & Titman (1993) presented a study that used relative strength strategy, which consists buying winners and selling losers, thus the opposite strategy of the contrarian strategy. They suggest that even though there is evidence of return reversals (overreaction) in the short-term (1 week or 1 month) and in the very long-term (3 to 5 years), is possible to retrieve excess or returns from a momentum (underreaction) strategy using a 3 to 12 months horizon.

What is clear is that the presence of overreaction/underreaction phenomenon is important for an efficient asset allocation and ultimately to state whether there is market efficiency in stock market in the sense of EMH. Moreover, this switch of pattern behavior corroborates the idea that market agents are not fully rational. As stated by the Nobel Laureate and one of the precursors of Behavioral Economics Daniel Kahneman "we observed systematic biases in our own decisions, intuitive preferences that consistently violated the rules of rational choice." (Kahneman, 2011, p. 12). This psychological bias, and resulting mispricing, can be found in the literature as price noise.

2.3 Noise and its Importance to Finance

According to Fischer Black, noise is a statement of the implications of human behavior that does not fit conventional notions of optimization. It is what makes our market observations imperfect, regardless the field in Economics we are studying (Black, 1986). Black, developed models in Finance, Econometrics, and Macroeconomics (see Fischer Black 1972, 1986, 1995, 1987), that were connected through a unique link, noise. He usually referred to these models as Equilibrium Models rather than Rational Equilibrium Models, mainly due to the fundamental role played by noise in such models. The importance of noise for Finance, is very well summarized in the following quote: "Noise makes financial markets possible, but also makes them imperfect. If there is no noise trading, there will be very little trading in individual assets. People will hold individual assets, directly or indirectly, but they rarely trade them" (Black, 1986, p. 530). Noise provides mispricing and market depth (liquidity) that makes trading profitably, and consequently desirable. The concept of noise trading was firstly exploited by (Kyle, 1985) and (Black, 1986), which consists, in uninformed agents trading on noise (irrelevant information) in contrast with sophisticated (informed) agents trading on relevant information. However, it was not clear what is noise and what is information, since all market participants believe they are trading based on information.

De Long, Shleifer, Summers, & Waldmann (1990) designed a market model that represented a contest between noise (uninformed) traders driven by market sentiment, and consequently, susceptible to systematic psychological bias and sophisticated (informed) traders driven by rational expectations, not susceptible to psychological bias. In the model *Arbitrageurs* (sophisticated traders) may restrict their arbitrage positions (limited arbitrage in contrast with perfect arbitrage) due to "noise trader risk"[1] pressured by their operational time horizon (finite). Building in the same rationale, Shleifer & Summers (1990) advocate that there are two types of risk that limit arbitrage. First, they consider the fundamental risk, which is derived from the uncertainty in the fundamentals of a specific stock or market. Let´s assume for instance, that a specific asset is being traded above its expected value and an *Arbitrageur* is selling it short. However, at some point in the future a better than expected result is released, this new information shifts the expected value of the underlying asset, resulting in a loss for this agent. The second source of risk, is the above mentioned "noise trader risk", since the sophisticated trader has a specific time horizon to liquidate his position, the possibility that the mispricing - which is a function of noise/uninformed traders psychological bias and market power- may be even greater at this date. However, this implicit market power equilibrium is a weak assumption, since informed/Arbitrageurs (sophisticated) traders, if there are any, can be understood as institutional traders, that account for 80% of market transactions, and therefore would be impossible for noise traders to compete with them.

3. Empirical Framework

This section describes the empirical approach by focusing on the applied case of study, the BM&FBovespa and Ibovespa Futures Contract, the description of the key variables and the empirical strategy.

3.1 BM&FBovespa – The Brazilian Futures and Stock Exchange

The BM&FBovespa is the São Paulo Stock, Mercantile & Futures Exchange, a company that is a result of the merge of São Paulo Stock Exchange (Bovespa) and the Brazilian Mercantile and Futures Exchange (BM&F), i.e. the Brazilian market for securities, commodities and futures. It is one of the largest exchanges in the world in terms of market value, the second largest in Americas and the largest in South America. The trading value in Bovespa´s equity market in 2016 totaled BRL 1.84 trillion (**USD 561 billion – 01/2017**)

[1] "the risk that noise traders' beliefs will not revert to their mean for a long time and might in the meantime become even more extreme"(De Long et al., 1990).

with an average daily trading value of BRL 7.41 billion (**USD 2.26 billion – 01/2017**). The BM&F segment (securities, commodities and futures) traded the total amount of BRL 59.60 trillion (**USD 18.17 trillion** – 01/2017) (BM&FBovespa, 2017). The main benchmark indicator is the Bovespa index, it is being broadcasted since 1968, which is a weighted theoretical portfolio with the stocks that represent 85% of the total of transactions and volume traded in the last twelve months, and were traded at least in 95% of trading days.

The Ibovespa Futures Contract is a derivative referred in the Bovespa index, characterized by high liquidity (90.000 contracts traded daily – BRL 5.8 billion of daily financial volume) [2] and moderate volatility. Since it is a derivative, it keeps the tradability features inherent to this kind of contract. Due to its liquidity and tradability, an investor can open a position in the buy side or sell side, been used either for hedge or for speculative positions. Its specific characteristics allows sophisticated traders to perform its arbitrage strategies and therefore, correct for any mispricing they believe exists.

Find below the Ibovespa Futures contract specifications considered here:

1. Underlying asset: The Ibovespa index
2. Price quotation: Index points times the Brazilian Real (R$) value of each point, as established by BM&F.
3. Minimum price fluctuation: Five index points.
4. Contract size: The Ibovespa futures times the Brazilian Real value of each point, as established by BM&F.
5. Contract months: Even-numbered months.
6. Last trading day: The Wednesday closest to the 15th calendar day of the contract month. Should this day be a holiday or a non-trading day at BM&F, the last trading day shall be on the following business day.
7. Day settlement of accounts (variation margin): The positions outstanding at the end of each session shall be marked to that day's settlement price, as determined by BM&F rules and regulations. The corresponding amount shall be cash settled on the following business day.

The variation margin shall be calculated up to and including the last trading day by the following formulas:

(a) For the positions initiated on the day

$$AD_t = (PA_t - PO)\ x\ M\ x\ N$$

(b) For the positions outstanding on the previous day

$$AD_t = (PA_t - PA_{t-1})\ x\ M\ x\ N$$

Where:

AD_t = the variation margin value, in Brazilian Reals, corresponding to date "t";

PA_t = the daily settlement price, in index points, for the corresponding contract month on date "t";

PO = the opening traded price, in index points;

M = the Brazilian Real value of each index point, as established by BM&F;

N = the number of contracts;

PA_{t-1} = the previous day's settlement price, in index points, for the corresponding contract month.

[2] Considering just the full contract. In case we consider the mini contract, which consists in a contract that represents 20% of the value of the full contract, the daily financial volume would reach around BRL 12 billion.

The variation margin value (AD_t) calculated as shown above, if positive, shall be credited to the buyer and debited to the seller. Should the calculation above present a negative value, it shall be debited to the buyer and credited to the seller.

3.2 Trade Session Structure at the Open

The active trading hours are usually from 9:00 to 17:55h [3] , however, from 8:45 to 9:00, there is the pre-market, where an auction takes place, in which bid and ask[4] orders can be placed and showed in the order book with no guarantee that they will be executed. It is highly important to visualize how the offer book is presented, in order to understand the possible effects that it may exert in a trader's mind (noise trader). Considering the book offer organization, regardless the side of the book, the orders at the top are the most probable to be executed. The bid side is organized in ascending order, i.e. the highest bid is placed at the top and therefore, it has higher probability to be executed. The ask side is exactly in the opposite way, in descending order, it means the lowest ask will be in a situation of preference. During the auction, there are no other price references than the offer book and closing price of the previous day. Since, at this moment, all relevant information is public, this offer contest dynamics, in which, just the best bids and asks are rewarded with the execution, will result in an accumulation of orders towards the direction agents processed the information, generating a price overreaction. Most important economic events, such as interest rate decisions, are released before or after trading hours (when markets are closed), in order to do not cause any kind of speculative distress during the trading session. Even if we consider international economic events from major financial markets, such as US and Europe, they will be released during non-trading hours due to the Brazilian time zone (UTC-3). It is important to mention that there are minor economic events that are released during trading hours that could affect the market eventually. However, due to its restricted importance, in general its effects are negligible. Building on the EMH argument, that prices should reflect all public information immediately, would be reasonable to assume that public information should be already contained in the opening price. Therefore, the price at the opening, probably capture and highlight any psychological bias affecting traders at that moment.

3.3 Price at the Close - the most important price level

The stock price at the close is the most important price information. This anecdotal evidence is widely known among finance researchers and practitioners, that is supported by the amount of studies that use this price level as reference. Almost the totality of works that consider one price information per unit of time use the last price available for this time horizon, regardless the time frame considered (intra-daily, daily, weekly, monthly, yearly and so on). Furthermore, financial returns are usually calculated based on the last price available in a fashion, that is, closing price in time *t* minus the closing price in *t-1* is equal the return in time *t* (see Bae, Chan, & Ng, 2004; Chen, 2012; Chuang, Liu, & Susmel, 2012; Darrat, Rahman, & Zhong, 2003; Hibbert, Daigler, & Dupoyet, 2008; Shan, Taylor, & Walter, 2014; Umutlu, Akdeniz, & Altay-Salih, 2010). It is well known that institutional traders have as function, provide liquidity to the market. Therefore, is reasonable to assume that institutional investors are responsible for most of market trading volume, specially, in a time that high frequency trading is a common practice. Nevertheless, institutional players, in aggregate, are responsible for 80% of trading volume (Puckett & Yan, 2011). "Institutional players place enormous importance on closing stock prices as benchmark of Value. Closing prices are used to calculate portfolio returns, tally the net asset values of mutual funds, and as a basis for certain types contracts and for after-hours trading" (Cushing & Madhavan, 2000). Consequently, following above affirmation, is logical to consider the closing price as a valid benchmark of value to develop our analysis.

[3] On contract expiration dates the trading session closes at 17:00.
[4] The difference between bid and ask prices is the widely known as bid-ask spread.

3.4. Empirical Strategy

This master thesis studies the behavior pattern of opening prices with respect to the closing price of the previous day. More precisely, we analyze whether the price at the open is an inefficient price level, result of the interaction of noise traders, i.e. traders that are reacting (overreacting) to irrelevant news (noise), and under which circumstances this situation happens considering some exogenous variables. Furthermore, we consider the hypothesis that the opening price may be a result of overreaction, showing a predictable behavioral pattern that can be used in a trading strategy to achieve excess of returns, therefore contradicting the EMH. We approach this matter, analyzing if the opening price retraces or not to the closing price of the previous day using a set of triggers to characterize the overreaction, and how the explanatory variables (difference close to open, trade volume and volatility) help to explain such dynamics. The results of this study may unveil interesting evidence: if price at the open is an inefficient price level, traders may tend to react (overreact) to unimportant news (noise), and subsequently there is a predictable behavioral pattern that can lead to excess of returns and asymmetric behavior with respect to positive and negative price movements.

In this work, the market concept will follow De Long et al. (1990) and Shleifer & Summers (1990) as theoretical reference. Furthermore, the idea of two different agents, Noise traders driven by market sentiment, and Sophisticated traders (Arbitrageurs) driven by rational expectations is a good approximation of market reality. However, the notion of a contest between these two groups, is at least naïve, since sophisticated traders (Arbitrageurs) are responsible to provide liquidity to the market, in a time that automated trading and High Frequency Trading (HFT) are a common practice. Moreover, this class of agent accounts for 80% of market volume (as explained previously), thus, an impossible competition for noise traders.

Based in the aforementioned, is reasonable to assume three possible market dynamics taking place. The first scenario, as in De Long at al. (1990), can be seen as competition, i.e., noise traders versus sophisticated traders, which in this case, perform an arbitrage strategy, and therefore will be referred as Arbitrageurs. Given that Arbitrageurs are driven by rational expectations, the new information[5] (irrelevant from a rational perspective) released previous to the trading session did not change Arbitrageurs' short-term perception of value. However, as suggested by Griffin & Tversky (1992), unsophisticated (noise) traders tend to place too much weight on new information, affecting noise trader's short-term measurement, consistent with agents driven by market sentiment. In such dynamics, given the obtained opening price (generated by noise traders)[6] is different from the closing price in the previous day, Arbitrageurs start place arbitrage orders, leading the price to reach a short-run equilibrium, that is, the closing price of the previous day, explained by the assumption that there is no rational shift on value due to the new information available. In other words, a price retracement occurs, showing a mean reversion pattern motivated by the Arbitrageurs' strategy, a consequence of noise trader's overreaction to the new information, therefore, a situation where the price at the opening is an inefficient price level.

The second feasible scenario, is that the new information available is relevant, changing the short-term value of the asset from a rational perspective. In this case, given the opening price, generated by the interaction of noise traders, is different from the price at the close (previous rational measurement of value) of the previous day, the informed sophisticated traders (Arbitrageurs[7] in the previous scenario) place their orders along (in the same side) with noise traders generating a momentum or trend. In this scenario, the price will not retrace to the price at the close of the previous day, actually, will possibly drift away, because the value

[5] Described as noise according to Black (1986)

[6] Arbitrageurs need a price reference to further find a price deviation to trade on, therefore, is reasonable to assume that opening price is generated by noise traders. Moreover, there is the anecdotal evidence (widely known among traders and practitioners) that institutional traders do not trade at the opening, supported by 15 years of work experience with the stock, monitoring institutional players.

[7] What define an Arbitrageur is the strategy, that is, the agent must apply an arbitrage strategy, which is not the case in that scenario.

from a rational perspective changed. In such case, the opening price can be understood as noise traders' underreaction to the new available information, and therefore, an efficient price level, at least in some extent. It is worth note that, this result can also be explained by an alternative scenario. Imagine that after the opening price formation, the sophisticated trader rationalizes about the true value of the asset and find that the mispricing is such that considering his operational time horizon (finite) and the extent of price retracement needed, an arbitrage strategy for this price deviation is too risky.[8] In such situation, he would rather adopt a momentum strategy trying to profit from the resulting trend, consistent with the analyses of Shleifer & Summers (1990), categorizing a herd behavior as advocated by Froot, Scharfstein, & Stein (1992). This behavior is consistent with more pronounced price deviations that can be verified in the empirical analyses.

3.5 Data, Variables and Descriptive Analyses

Data

The original data sample is from BM&FBovespa database, and consists on business days from January 2, 2003 to December 30, 2013, totaling 2,720 observations. This period (01/2003 – 12/2013) was selected in order to avoid the political turbulence started 2014 that could negatively affect the time series analysis. Each event consists on date, opening price (O), maximum price (Max), minimum price (Min), closing price (C), negotiated financial volume and contracts negotiated in the period.

Variables

Since the aim of this work is to analyze mispricing or price deviation, we consider only the cases that there is a price difference form the price at the open in t with respect to the price at the close in $t-1$. From now on, for the analyses and notation simplicity, the difference from the opening price in t and the closing price in $t-1$, may be referred as opening gap. We start explaining the criteria used to define opening gaps and its closure (price retracement). Since there are many works that found overreaction asymmetry regarding positive and negative movements (see Caginalp, Desantis, & Sayrak, 2014; Da Costa, 1994; Hibbert et al., 2008; Kim, 2006), we decided to differentiate between positive and negative gaps. A positive opening gap is created (**gapposit** =1) if the opening price in t is greater than the closing price in t-1. In this case, the price retracement occurs (**gappositf** =1) if the minimum quotation in t is less than or equal to the closing price in t-1. A negative gap occurs (**gapneg** = 1) when the opening price in t is less than the closing price in t-1, it closes (**gapnegf** = 1) if the maximum quotation in t is greater than or equal to the closing quotation in t-1. Using mathematical notation:

Positive opening gap: $Gapposit = 1\ if\ O_t - C_{t-1} > 0$
Negative opening gap: $Gapneg = 1\ if\ O_t - C_{t-1} < 0$

Closure (price retracement) of positive gap: $Gappositf = 1\ if\ Min_t \leq C_{t-1}$
Closure (price retracement) of negative gap: $Gapnegf = 1\ if\ Max_t \geq C_{t-1}$

Gappositf and **gapnegf** are going to be the endogenous variables for the first set of regressions, in which we developed a probit and a logit approach. These are binary variables, assuming the value 0 if there was no price retracement (gap does not close) and 1 in case there was price retracement (gap close).

[8] Consistent with the concept of noise trader risk as suggested by De Long, Shleifer, Summers, & Waldmann (1990)

It is reasonable to assume the price retracement is a function of the retracement extension or length, i.e. depends on gap size. Nevertheless, gap size is expected to affect negatively the probability of price retracement (closing the gap), i.e. bigger the gap, smaller the probability. The gap size variable consists in the difference in points between the closing price in t-1 and the opening price in t. In addition, we created the ratio between the opening and the closing price in t-1, which captures the size in percentage terms. That percentage size will be used to check for different triggers size when accounting for overreaction, that is, it will make possible understanding how the overreaction behavior is conditioned to the retracement length.

Volatility is one the most important feature when studying asset returns, mainly by its role when measuring asset risk (Bae et al., 2004; R. Y. Chou, Chou, & Liu, 2009; R. Y. Chou & Liu, 2010; Li & Hong, 2011). A variable that captures the volatility was imperative for this study, since at times when the market is more volatile, the price experiences greater variations, what implies a greater probability of overreaction. In this case we created the volatility variable based on price range, the price range has been increasingly used in the academic literature to measure volatility (H.C. Chou & Wang, 2006). The range variable is measured by the difference between the maximum and minimum quotation of the day divided by the previous closing price.

It is well known in financial literature that trading volume is of significant importance when explaining price behavior and is the most used covariate when analyzing asset pricing and market risk (see Bialkowski, Darolles, & Le Fol, 2008; G. Caginalp & Desantis, 2008; Gunduz Caginalp & Desantis, 2011; H.-C. Chou & Wang, 2006; Darrat et al., 2003). It is expected that volume has a positive impact in the probability of price retracement, since it supports and provides consistency to price movements. Since the volume variable showed nonstationary behavior, we decided to create a variable (volumeHP) that would soften the stochastic component of the financial volume, so we used a Hodrick-Prescott filter for cyclical components, which according to Hodrick & Prescott (1997) can be represented as follows:

$$y_t = g_t + c_t \qquad \text{for } t = 1, \dots, T. \tag{1}$$

Where the time series y_t is the sum of a growth component g_t and cyclical component c_t. The measure of the smoothness of the $\{g_t\}$ path is the sum of the squares of its second difference. The c_t are deviations from g_t, which over longtime periods, their average is near zero (Hodrick & Prescott, 1997).

$$\underset{\{g_t\}_{t=-1}^{T}}{Min} \left\{ \sum_{t=1}^{T} c_t^2 + \lambda \sum_{t=1}^{T} [(g_t - g_{t-1}) - (g_{t-1} - g_{t-2})]^2 \right\} \tag{2}$$

When the smoothness penalty $\lambda \longrightarrow 0$, g_t would just be the series y_t itself, whereas when $\lambda \longrightarrow \infty$ the procedure amounts to a regression on a linear time trend, that is, produces a series whose second difference is exactly 0 (Hamilton, 2016).

We extend the descriptive analyses (time horizon) for five days, i.e., check for price retracement in the following five days (trading week) after opening the gap. For this purpose, we create two variables capturing if the opening price in t retraced to closing price in $t-1$ (gap close) for the period t up to $t+5$. In this case, two count variables were created, which can assume value 0 in case the price did not retrace and from 1 up to 6 otherwise. Specifically, it will assume value of 1 (gap closed in t), 2 (closed in t+1), 3 (closed in t+2), 4 (closed in t+3), 5 (in t+4) and 6 if the gap closed in t+5 (working days). In order to perform this study, we had to create ten more variables (five different lags forward for minimum and maximum prices). That is, Min_{t+1}, Min_{t+2}, Min_{t+3}, Min_{t+4}, Min_{t+5}, Max_{t+1}, Max_{t+2}, Max_{t+3}, Max_{t+4} and Max_{t+5}. Using these variables as references, we can follow whether the opening price retraced up to five days. Then, a positive gap closes if the minimum quotation in t up to $t+5$, is less than or equal to the closing price in

$t-1$ and a negative gap closes if the maximum quotation in t up to $t+5$ is greater than or equal to the closing quotation in $t-1$. In mathematical notation and departing from the previous situation:

Positive gaps

Closure (price retracement) in t: $gappositf = 1\ if\ Min_t \leq C_{t-1}$

Closure (price retracement) in $t+1$: $gappositf = 2\ if\ Min_{t+1} \leq C_{t-1}$, given it did not close in t.

Closure (price retracement) in $t+2$: $gappositf = 3\ if\ Min_{t+2} \leq C_{t-1}$, given it did not close in t and $t+1$.

Closure (price retracement) in $t+3$: $gappositf = 4\ if\ Min_{t+3} \leq C_{t-1}$, given it did not close in t, $t+1$, and $t+2$.

Closure (price retracement) in $t+4$: $gappositf = 5\ if\ Min_{t+4} \leq C_{t-1}$, given it did not close in t, $t+1$, $t+2$ and $t+3$.

Closure (price retracement) in $t+5$: $gappositf = 6\ if\ Min_{t+5} \leq C_{t-1}$, given it did not close in t, $t+1$, $t+2$, $t+3$ and $t+4$.

Negative gaps

Closure (price retracement) in t: $gapnegf = 1\ if\ Max_t \geq C_{t-1}$

Closure (price retracement) in $t+1$: $gapnegf = 2\ if\ Max_{t+1} \geq C_{t-1}$, given it did not close in t.

Closure (price retracement) in $t+2$: $gapnegf = 3\ if\ Max_{t+2} \geq C_{t-1}$, given it did not close in any previous day.

Closure (price retracement) in $t+3$: $gapnegf = 4\ if\ Max_{t+3} \geq C_{t-1}$, given it did not close in any previous day.

Closure (price retracement) in $t+4$: $gapnegf = 5\ if\ Max_{t+4} \geq C_{t-1}$, given it did not close in any previous day.

Closure (price retracement) in $t+5$: $gapnegf = 6\ if\ Max_{t+5} \geq C_{t-1}$, given it did not close in any previous day.

Descriptive analyses

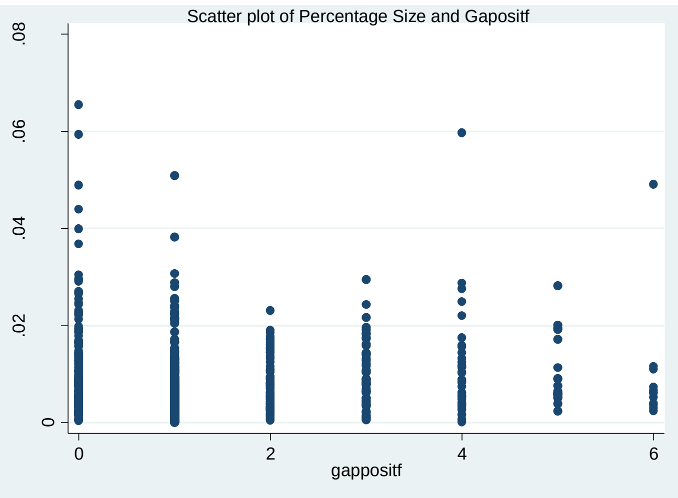
Table 1 presents the summary of variables of interest. We can see that the mean size of the first price movement of the day is 260.2 points or 0.54%. This is a useful information since it is going to be used as basis for the overreaction triggers. The biggest opening price movement is 4,375 points or 9.78%, which is a huge movement, considering that the underlying asset is a market index, and therefore should represent the whole market pricing. Moreover, price movements of 10% trigger a mechanism that paralyzes the trading session for thirty minutes, to avoid the spread of panic situations. The percentage range shows a mean of 0.0233, that is an average intraday volatility of 2.33%. The variables **gappositf** and **gapnegf** present mean values of 0.62 and 0.58 respectively, suggesting that lower values present higher frequency, since the possible values drift from 0 to 6.

Variable	Obs	Mean	Std. Dev.	Min	Max
gappositf	2751	0.624864	0.92655	0	6
gapnegf	2751	0.583788	0.900189	0	6
sizeabs	2751	260.2328	338.0369	5	4375
sizeperc	2751	0.0054	0.007082	7.23E-05	0.097875
Rangeperc	2751	0.023385	0.012532	0.003577	0.122691
volumeHP	2751	3.58E+09	1.63E+09	6.29E+08	7.05E+09

Table 1: descriptive analyze of main variables

Figures 1a and 1b contrast the percentage size against with our two main variables (gapositf and gapnegf), we can see that, the most extreme opening price movements (close to 10%) occurs when considering negative movements, an interesting evidence of asymmetric behavior. Moreover, it suggests that investors tend to overreact more to bad news than to good news.

(a)



(b)

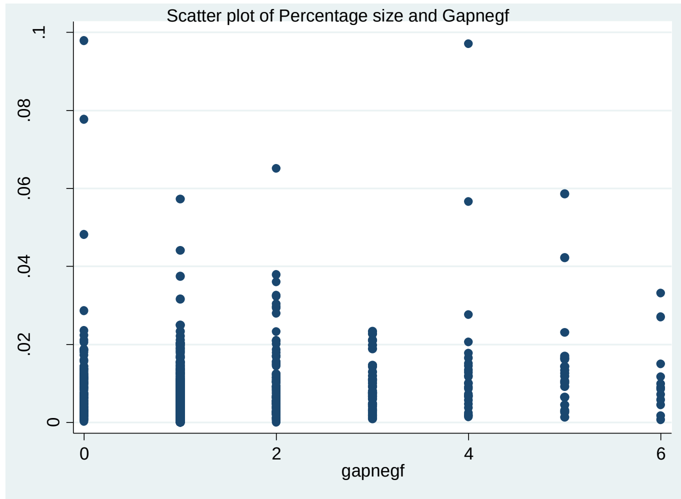
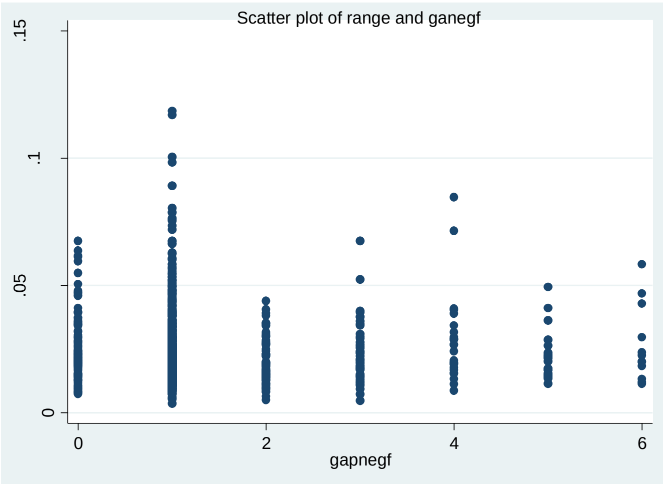


Figure 1: scatter plot of size with respect to positive and negative gaps

Figures 2a and 2b consider volatility with respect the two variables of interest. Higher volatility favors price retracement. We can also note that, for negative gaps there is no observation of value 0 with higher volatility, corroborating the idea of asymmetric behavior. It means that, whenever there is high volatility a mean reversion pattern appears, suggesting overreaction.

(a)



(b)

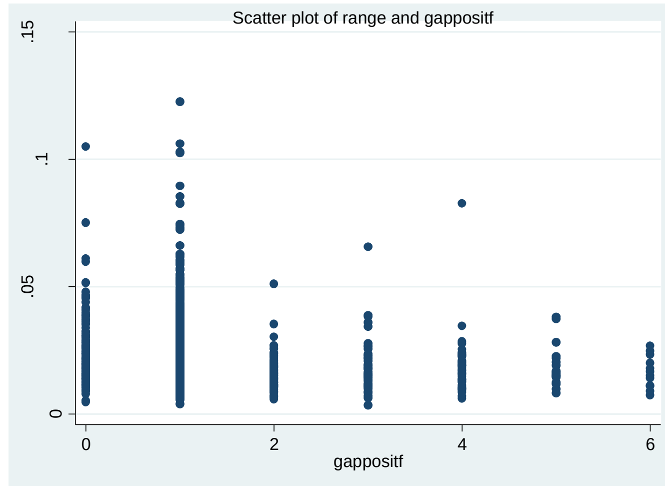


Figure 2: scatter plot of volatility with respect to negative and positive gaps

4. Empirical analyses and results

4.1 Count data analyses

By the following tables, we analyze how often the price retracement occurs conditioned to different trigger size. In table 2, our variables of interest are studied separately without using any trigger. We can note that, the table regarding positive gaps shows that in 87.93% of observations there are price retracement considering up to 5 days. Moreover, we can see that in 73.03% of events the reverting pattern occur intraday, i.e. in the same day that the gap is created. Looking to negative gaps, we find the same behavior, though showing a stronger mean reversion pattern. In just 8.79% of the situations there is no price retracement, presenting a strong support to the concept of noise traders (traders tend to trade on noise).

gappositf	Freq.	Percent	Cum.
0	174	12.07	12.07
1	1,054	73.09	85.16
2	92	6.38	91.54
3	50	3.47	95.01
4	41	2.84	97.85
5	19	1.32	99.17
6	12	0.83	100
Total	1,442	100	

gapnegf	Freq.	Percent	Cum.
0	115	8.79	8.79
1	991	75.71	84.49
2	97	7.41	91.9
3	47	3.59	95.49
4	27	2.06	97.56
5	20	1.53	99.08
6	12	0.92	100
Total	1,309	100	

Table 2: variables of interest with no trigger

Since most of overreaction studies the trigger used to define overreaction seems to be arbitrary, we chose different trigger sizes from 0.5% up to 2% to show in the following tables, 3 and 4. Even though we checked opening movements considering triggers up to 5%, there is a trade-off between increasing trigger and number of observations, that is, higher the trigger less observations we have. Despite that in general, showing the same price behavior when looking to higher triggers, the asymmetric behavior pattern became clearer as we increase the threshold. While when accounting for positive gaps the momentum or trend pattern increases with the trigger, reaching 50% with a 5% trigger, for negative gaps it remains under 30%, clearly showing a systematic bias in traders' behavior.

From the tables 3 and 4, when considering positive gaps, we find that the mean reversion pattern gradually decreases from 87,93%(no trigger), 77.34% (0.5% trigger), 68.63% (1% trigger), 59.6% (1.5% trigger) to 55.77% (2% trigger). When considering negative opening gaps, we found an interesting pattern, if we take into account just the same day the gap occurred, we find the same behavior as in positive gaps. However, if we consider the following five days, the price retracement level remains almost constant. It means, that the intraday price retracement decreases, while in the following days it increases, leaving the percentage of observations that do not experience mean reversion around 20%. This evidence supports the overreaction and noise traders' hypothesis for negative gaps, while also corroborating for the asymmetric behavior pattern.

Trigger	0.50%		
gappositf	Freq.	Perc	Cum.
0	116	22.66	22.66
1	254	49.61	72.27
2	52	10.16	82.42
3	37	7.23	89.65
4	28	5.47	95.12
5	17	3.32	98.44
6	8	1.56	100
Total	512	100	

Trigger	1%		
gappositf	Freq.	Perc	Cum.
0	64	31.37	31.37
1	69	33.82	65.2
2	23	11.27	76.47
3	22	10.78	87.25
4	17	8.33	95.59
5	6	2.94	98.53
6	3	1.47	100
Total	204	100	

Trigger	0.50%		
gapnegf	Freq.	Perc	Cum.
0	74	16.55	16.55
1	232	51.9	68.46
2	63	14.09	82.55
3	32	7.16	89.71
4	21	4.7	94.41
5	16	3.58	97.99
6	9	2.01	100
Total	447	100	

Trigger	1%		
gapnegf	Freq.	Perc	Cum.
0	39	20.42	20.42
1	70	36.65	57.07
2	37	19.37	76.44
3	15	7.85	84.29
4	13	6.81	91.1
5	13	6.81	97.91
6	4	2.09	100
Total	191	100	

Table 3: variables of interest with trigger of 0.5% (left column) and 1% (right column)

Trigger	1.50%			Trigger	2%		
gappositf	Freq.	Percent	Cum.	gappositf	Freq.	Percent	Cum.
0	40	40.4	40.4	0	23	44.23	44.23
1	22	22.22	62.63	1	17	32.69	76.92
2	12	12.12	74.75	2	1	1.92	78.85
3	11	11.11	85.86	3	3	5.77	84.62
4	8	8.08	93.94	4	5	9.62	94.23
5	5	5.05	98.99	5	2	3.85	98.08
6	1	1.01	100	6	1	1.92	100
Total	99	100		Total	52	100	
Trigger	1.50%			Trigger	2%		
gapnegf	Freq.	Percent	Cum.	gapnegf	Freq.	Percent	Cum.
0	15	18.07	18.07	0	8	17.02	17.02
1	27	32.53	50.6	1	13	27.66	44.68
2	20	24.1	74.7	2	13	27.66	72.34
3	6	7.23	81.93	3	4	8.51	80.85
4	7	8.43	90.36	4	4	8.51	89.36
5	6	7.23	97.59	5	3	6.38	95.74
6	2	2.41	100	6	2	4.26	100
Total	83	100		Total	47	100	

Table 4: variables of interest with trigger of 1.5% (left) and 2% (right)

4.2 Econometric approach

Since the aim of this work is to check for overreaction and noise trading, the empirical model consists in analysing the probability of intraday price retracement considering the gap size, volume and volatility through non-linear estimates (logit and probit). We consider as dependent variables the information whether the price retraced or not, and we regress it against the gap size, the trading volume and volatility. According Wooldridge (2012), the main difference between probit and logit specification lies on the shape (sigmoide) of the cumulative distribution function (cdf). While logit assumes a cdf for the standard logistic random variable, probit assumes a cdf for the standard normal random variable.

Probit specification:

$$P(gapositf = 1|\, size, volatility, volume) = \Phi(\alpha_i size + \beta_i volatility + \gamma_i volume + \varepsilon_i) \quad (3)$$

$$P(gapnegf = 1|\, size, volatility, volume) = \Phi(\alpha_i size + \beta_i volatility + \gamma_i volume + \varepsilon_i) \quad (4)$$

Where Φ is the standard normal cumulative distribution function, which can also be expressed as the integral: $\int_{-\infty}^{\alpha_i size+\beta_i volatility+\gamma_i volume+\varepsilon_i} \phi(v)dv$, (Wooldridge, 2012).

Logit specification:

$$P(gapositf = 1|\, size, volatility, volume) = \frac{e^{\alpha_i size+\beta_i volatility+\gamma_i volume+\varepsilon_i}}{1+e^{\alpha_i size+\beta_i volatility+\gamma_i volume+\varepsilon_i}} \quad (5)$$

$$P(gapnegf = 1|\, size, volatility, volume) = \frac{e^{\alpha_i size+\beta_i volatility+\gamma_i volume+\varepsilon_i}}{1+e^{\alpha_i size+\beta_i volatility+\gamma_i volume+\varepsilon_i}} \quad (6)$$

The logit and probit approaches are used for analyzing the intraday dynamics. In order to capture behavior asymmetry in case it exists, we develop two models for each specification, a model considering just positive gaps and a model accounting for negative gaps.

4.3 Econometric results

For each approach (probit and logit), we present two estimations, for positive and negative gaps. Constructing the empirical framework, we use a threshold of 0.5% as the trigger that allowed the gap to qualify for the overreaction analysis.

Due to its nature, the interpretation of binary response models become more complex and less intuitive than in linear models: the analysis of the coefficients allows only to determine if a certain variable has a positive or negative impact on the probability of the event occur, that is, only the signal is evaluated. Regarding the marginal effects of the explanatory variables, it is known that they are not constant, as they vary according to the position in the cumulative distribution function.

In table 5, we report the results for the estimation of equations 3, 4, 5, 6, in which column (1) refers to the probit specification for positive gaps, column (2) to the logit specification for positive gaps, column (3) to the probit specification for negative gaps and finally column (4) to the logit specification for negative gaps. We can see that there are no major differences in fitting between the logit and probit specifications, even when comparing the log likelihood and information criteria, such as, AIC and BIC.

	(1) Positive gap *probit*	(2) Positive gap *logit*	(3) Negative gap *probit*	(4) Negative gap *logit*
main				
sizeabs	-0.00176^{***}	-0.00293^{***}	-0.00173^{***}	-0.00305^{***}
	(0.000225)	(0.000399)	(0.000303)	(0.000545)
rangeperc	36.69^{***}	62.21^{***}	30.72^{***}	53.79^{***}
	(5.865)	(10.52)	(5.876)	(10.62)
volumeHP	$9.91e-11^{**}$	$1.68e-10^{**}$	$1.56e-10^{***}$	$2.71e-10^{***}$
	(4.04e-11)	(6.71e-11)	(4.80e-11)	(7.98e-11)
_cons	-0.353^{*}	-0.620^{*}	-0.431^{**}	-0.734^{**}
	(0.202)	(0.340)	(0.214)	(0.354)
N	512	512	447	447
ll (null)	-354.8757	-354.8757	-309.5134	-309.5134
ll (model)	-301.3081	-301.2653	-268.3984	-267.2994
AIC	608.6162	608.5305	542.7968	540.5988
BIC	621.3311	621.2455	555.1045	552.9064

Standard errors in parentheses;

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: estimation results of probit and logit specifications for positive and negative gaps

The variable size presents a negative coefficient, meanwhile the variables range and volume show a positive impact on price retracement. The results are consistent with the expectations, since all explanatory variables presented the expected influence on the endogenous one. Moreover, the coefficients show that the variables have a statistically significant effect when explaining overreaction.

TRUE			
Classified	Close	Do not close	Total
Positive	173	84	257
Negative	81	174	255
Total	254	258	512
Sensibility			68.11%
Specificity			67.44%
Positive forecast value			67.32%
Negative forecast value			68.24%
False + for true ~D			32.56%
False - for true D			31.89%
False + for classified +			32.68%
False - for classified -			31.76%
Correctly classified			67.77%

Table 6: Classification and prediction of positive gaps.

It can be seen from table 6 that the probit model for positive gaps showed a percentage of 67.77% of gaps correctly classified, achieving a good predictive power as explicit in the indicators of sensitivity and specificity with values of 68.11% and 67.44% respectively.

TRUE			
Classified	Close	Do not close	Total
Positive	171	79	250
Negative	61	136	197
Total	232	215	447
Sensibility			73.11%
Specificity			63.26%
Positive forecast value			68.40%
Negative forecast value			69.04%
False + for true ~D			36.74%
False - for true D			26.29%
False + for classified +			31.60%
False - for classified -			30.96%
Correctly classified			68.68%

Table 7: Classification and prediction of negative gaps.

The probit negative model presented a percentage of 68.68% of correctly classified gaps, also yielding a good predictive power predicting correctly 73.71% of the gaps that closed and 63.23% of the gaps that did not close.

We opted by the average marginal effect because it accounts for the entire sample, although the marginal effect at the average has reached similar results. The variable size has a negative effect on the probability of the event occur, based on figures 3a and 3b, it is clear that the bigger the absolute size of the first price movement, smaller the probability of price retracement in both models. These results follow the expected effect, since the larger the window in the graph, greater the extent of the movement required to fill it.

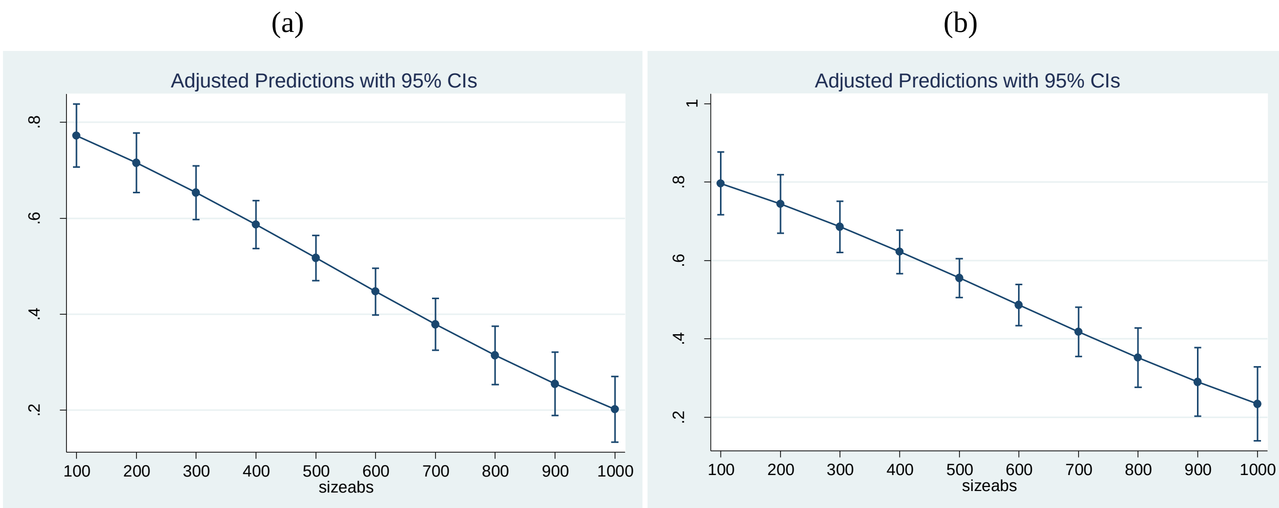


Figure 3: marginal effects of size for positive and negative gaps

Figure 3, shows the probability of closing positive (panel a) and negative (panel b) gaps conditioned to their absolute size with 95% confidence interval. keeping the other independent variables at the mean, range (2.34%) and VolumeHP (R$ 3.86 billion), with the absolute size variable assuming the value of one hundred points, a probability of closure of 77.21% for positive gaps and 79.68% for negative gaps are reached, decreasing to 20.15% and 23.38% respectively, with the absolute size rising to a thousand points. The results corroborate the asymmetric behavior found on the count data analyses.

Using the average marginal effect as basis, we arrive at the interpretation that an increase in one hundred points in the absolute size decreases by an average of 5.71% and 5.68% (for positive and negative gaps respectively) in the probability of price retracement.

The rangeperc variable, has a positive effect on the dependent variable, we can notice that the higher its values, greater the probability of price retracement. Since range is a measure of volatility, it is expected that in moments of greater volatility the probability of price retracement will be greater. Therefore, the result presented meets the expectation.

Figures 4a and 4b presents how the probability of price retracement conditioned to the range behaves with a 95% confidence interval. Keeping the other covariates at the mean, the range variable at the level of 1%, we find a probability of closure of 28.74% for positive gaps and 30.39% for negative gaps. Considering the volatility at 8% the probability rises to 97.77% and 94.92% respectively. According to the table of average marginal effects, a 1% increase in volatility increases by 11.92% and 10.09% for positive and negative gaps respectively the likelihood of price retracement.

(a) (b)

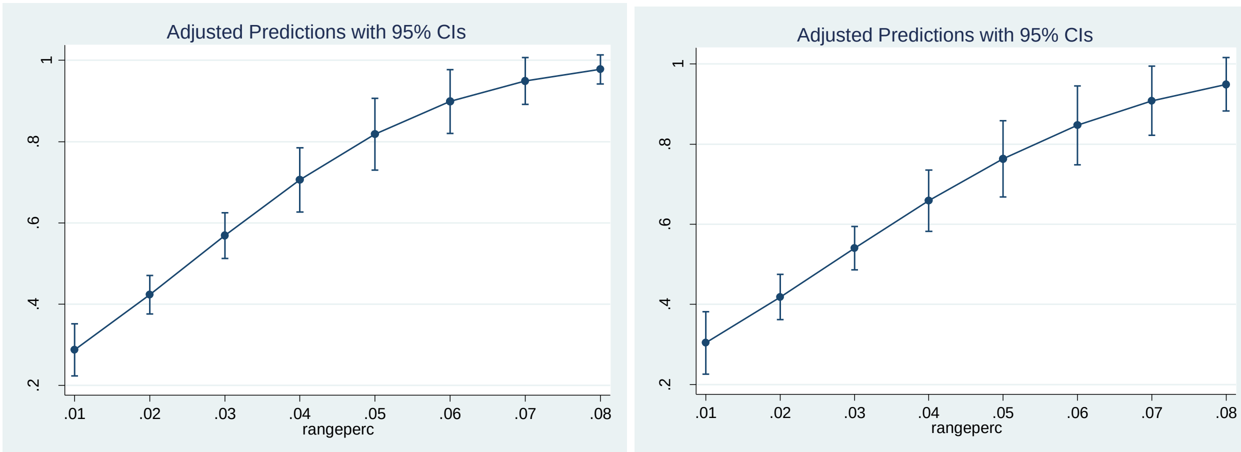


Figure 4: marginal effects of volatility for positive and negative gaps

Figures 5a and 5b show the marginal effect of volume on the probability of price retracement with 95% IC. When the variable volume assumes the value of 1.5e+09, the probability is 39.35% for positive gaps and 35.77% for negative gaps. When the volume reaches a value close to its maximum of 7.5e+09 the probability jumps to 62.73% and 71.65% respectively. This marginal effect analysis presents a clear asymmetric component; the volume affects more heavily negative than positive gaps.

(a) (b)

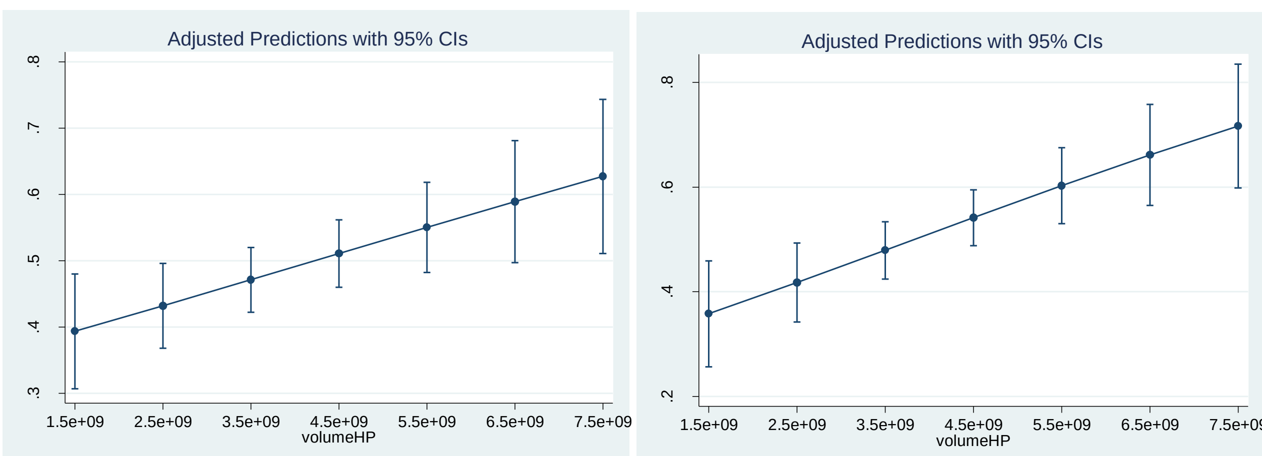


Figure 5: marginal effects of trading volume for positive and negative gaps

As a consequence of such results, it is possible, for example, to perform a simulation of a real market situation. In this way, the importance of this work can be better demonstrated.

At the opening of the trading session shows that the future Bovespa index contract opened with a negative gap of two hundred points and the variables range and volume are at the mean levels with values of 2.69% and R $ 3.86 billion respectively. According to the model for negative gaps, there is a probability of 74.46% of price retracement. Therefore, in case the investor opts for buying a contract of the future Bovespa index at this moment, there is a probability of 74.46% of success in the transaction. This predictable pattern, which can yield excess of returns if used in a trading strategy, represents a significant stroke to the EMH, since it assumes that price changes must be completely random and unpredictable.

4. Conclusion

The ultimate objective of this master thesis, was to check the hypothesis that the opening price is a result of overreaction and noise trading, and consequently an inefficient price level. Furthermore, we also intended to demonstrate that there are predictable patterns when studying the dynamics of price behavior at the opening, contradicting the Efficient Market Hypothesis (EMH).

We analyse the price retracement considering different triggers. We find that the opening price is frequently an inefficient price level derived from the interaction of uninformed traders. This mispricing is corrected by sophisticated traders (responsible for 80% of trading volume) acting as *Arbitrageurs* leading the price to the equilibrium again. We can note that without using any trigger we found that 87.93% of positive and 91.21% of negative gaps retrace to the closing price of the previous day. Therefore, a very strong and predictable retracement pattern, consistent with the first scenario mentioned in the empirical strategy part. It is interesting to realize that, there is a very clear asymmetric behavior when accounting for positive and negative gaps, which is accentuated as the trigger increases. While looking to positive gaps the momentum/trend or herd behavior increases with the trigger, reaching 50% with a 5% trigger, for negative gaps this pattern remains under 30%, showing a systematic bias in trader´s behavior. The evidence that herd behavior increases with the trigger is consistent with the alternative scenario. It means that, Arbitrageurs take into account "noise trader risk" when deciding the trading strategy, since under higher price deviations, sophisticated traders tend to place orders along with unsophisticated traders generating a momentum or trend.

From the econometric analyses, we inferred how gap size, volatility and trading volume affect the probability of price retracement and consequently overreaction, since price retracement is what defines the overreaction concept. Moreover, we found the expected relationship between endogenous and explanatory variables. Based on four econometric models, using two different approaches, we arrived to predictions permitting their use in real market situations. In such case, this work can improve asset allocation for individual investors, mutual funds and hedge funds for example.

Another important contribution of this study, relies on the findings regarding price efficiency and psychological bias that affect investors, shedding light in the discussion regarding the Efficient Market Hypothesis and alternative theories, central to financial theory.

The econometric framework used in this work, allowed the analyses of price retracement and overreaction in t, that is, the intraday dynamics. This feature is a limitation of this study; however, it also can be seen as a motivation for future works, in which we could extend the econometric analyses to five days. The initial intention of this master thesis was to consider as dependent variable the price retracement for five days and develop two more approaches, a negative binomial and survival analyses. This objective will be accomplished in the future.

References

Bae, K.-H., Chan, K., & Ng, A. (2004). Investibility and return volatility. *Journal of Financial Economics*, *71*(2), 239–263. https://doi.org/10.1016/S0304-405X(03)00166-1

Balvers, R., Wu, Y., & Gilliland, E. (2000). Mean Reversion across National Stock Markets and Parametric Contrarian Investment Strategies. *The Journal of Finance*, *55*(2), 745–772. https://doi.org/10.1111/0022-1082.00225

Bialkowski, J., Darolles, S., & Le Fol, G. (2008). Improving VWAP strategies: A dynamic volume approach. *Journal of Banking and Finance*, *32*(9), 1709–1722. https://doi.org/10.1016/j.jbankfin.2007.09.023

Black, F. (1972). Capital Market Equilibrium with Restricted Borrowing. *The Journal of Business*, *45*(3), 444. https://doi.org/10.1086/295472

Black, F. (1986). Noise. *Journal of Finance*, *41*(3), 529–543. https://doi.org/10.1111/j.1540-6261.1986.tb04513.x

Black, F. (1987). General Equilibriium and Business Cycles. *National Bureau of Economic Research*, (950). Retrieved from http://www.nber.org/papers/w0950.pdf

Black, F. (1995). Interest Rates as Options. *The Journal of Finance*, (7). Retrieved from https://www.jstor.org/stable/pdf/2329320.pdf?refreqid=excelsior%3Aa7b066ae5dcae948a070a16f00576514

BM&FBovespa. (2017). Market performance. Retrieved August 1, 2017, from http://www.bmfbovespa.com.br/en_us/news/market-performance-6.htm

Caginalp, G., & Desantis, M. (2008). Stock price dynamics: nonlinear trend, volume, volatility, resistance and money supply. *Quantitative Finance*, *11*(6), 849–861. https://doi.org/10.1080/14697680903220356

Caginalp, G., & Desantis, M. (2011). Nonlinearity in the dynamics of financial markets. *Nonlinear Analysis: Real World Applications*, *12*, 1140–1151. Retrieved from http://ssrn.com/abstract=1910296

Caginalp, G., Desantis, M., & Sayrak, A. (2014). The nonlinear price dynamics of U.S. equity ETFs. *Journal of Econometrics*, *183*, 193–201. https://doi.org/10.1016/j.jeconom.2014.05.009

Chen, S.-S. (2012). Revisiting the empirical linkages between stock returns and trading volume. *Journal of Banking & Finance*, *36*(6), 1781–1788. https://doi.org/10.1016/j.jbankfin.2012.02.003

Chou, H.-C., & Wang, D. (2006). Using Conditional Autoregressive Range Model to Forecast Volatility of the Stock Indices. In *Proceedings of the 9th Joint Conference on Information Sciences (JCIS)*. https://doi.org/10.2991/jcis.2006.175

Chou, R. Y., Chou, H.-C., & Liu, N. (2009). Range Volatility Models and Their Applications in Finance. https://doi.org/10.1007/978-0-387-77117-5_83

Chou, R. Y., & Liu, N. (2010). The economic value of volatility timing using a range-based volatility model. *Journal of Economic Dynamics and Control*, *34*(11), 2288–2301. https://doi.org/10.1016/j.jedc.2010.05.010

Chuang, W.-I., Liu, H.-H., & Susmel, R. (2012). The bivariate GARCH approach to investigating the relation between stock returns, trading volume, and return volatility. *Global Finance Journal*, *23*(1), 1–15. https://doi.org/10.1016/j.gfj.2012.01.001

Cohen, K. J. (1966). Review: The Random Character of Stock Market Prices. *The Journal of Business*, *39*(4), 532–535. Retrieved from http://www.citeulike.org/group/2384/article/1204379

Cushing, D., & Madhavan, A. (2000). Stock returns and trading at the close. *Journal of Financial Markets*, *3*, 45–67. Retrieved from http://ac.els-cdn.com.sire.ub.edu/S1386418199000129/1-s2.0-S1386418199000129-main.pdf?_tid=33bb1582-6bbb-11e7-8e6e-00000aacb360&acdnat=1500383923_2f6950e0dc42540f4fa5aa149642eebd

Da Costa, N. C. A. (1994). Overreaction in the Brazilian stock market. *Journal of Banking and Finance*, *18*, 633–642. Retrieved from http://ac.els-cdn.com.sire.ub.edu/0378426694000115/1-s2.0-0378426694000115-main.pdf?_tid=7be7b46a-714b-11e7-ad18-00000aacb35f&acdnat=1500995647_e5b45b57805a1f201e7f7acadb3bc265

Darrat, A. F., Rahman, S., & Zhong, M. (2003). Intraday trading volume and return volatility of the DJIA stocks: A note. *Journal of Banking & Finance*, *27*(10), 2035–2043. https://doi.org/10.1016/S0378-4266(02)00321-7

De Bondt, W. F. M., & Thaler, R. (1985). American Finance Association Does the Stock Market Overreact? *Source: The Journal of Finance*, *40*(3), 793–805. Retrieved from http://www.jstor.org/stable/2327804

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise Trader Risk in Financial Markets. *Source: The Journal of Political Economy*, *98*(4), 703–738. Retrieved from http://www.jstor.org/stable/2937765

Fama, E. F. (1965). Random Walks in Stock Market Prices. *Financial Analysts Journal*, *21*(5), 55–59. https://doi.org/10.2469/faj.v21.n5.55

Fama, E. F. (1970). Efficient Capital Markets : A Review of Theory and Empirical Work. *Journal of Finance*, *25*(2). https://doi.org/10.2307/2325486

Froot, K. A., Scharfstein, D. S., & Stein, J. C. (1992). Herd on the Street: Informational Inefficiencies in a Market with Short-Term Speculation. *Source: The Journal of Finance THE JOURNAL OF FINANCE a*, *47*(4), 1461–1484. Retrieved from http://www.jstor.org/stable/2328947

Griffin, D., & Tversky, A. (1992). The Weighing of Evidence and the Determinants of Confidence. *Cognitive Psychology*, *24*. Retrieved from http://ac.els-cdn.com.sire.ub.edu/001002859290013R/1-s2.0-001002859290013R-main.pdf?_tid=e3738b06-6c8a-11e7-b8e1-00000aacb35d&acdnat=1500473123_ce55494ff3d039c45a1cdaf6c8e0fa76

Hamilton, J. D. (2016). *Why You Should Never Use the HodrickPrescott Filter*. *UCSD Department of Economics Working Paper*. https://doi.org/10.3386/w23429

Hibbert, A. M., Daigler, R. T., & Dupoyet, B. (2008). A behavioral explanation for the negative asymmetric return–volatility relation. *Journal of Banking & Finance*, *32*(10), 2254–2266. https://doi.org/10.1016/j.jbankfin.2007.12.046

Hillebrand, E. (2003). *Mean Reversion Models of Financial Markets*. Retrieved from http://elib.suub.uni-bremen.de/diss/docs/E-Diss549_diss02.pdf

Hodrick, R. J., & Prescott, E. C. (1997). Postwar U . S . Business Cycles : An Empirical Investigation. *Journal of Money, Credit and Banking*, *29*(1), 1–16. https://doi.org/10.2307/2953682

Jegadeesh, N. (1990). Evidence of Predictable Behavior of Security Returns. *Source: The Journal of Finance*, *45*(3), 881–898. Retrieved from http://www.jstor.org/stable/2328797

Jegadeesh, N., & Titman, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *Source: The Journal of Finance*, *48*(1), 65–91. Retrieved from http://www.jstor.org/stable/2328882

Kahneman, D. (2011). *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux.

Kim, S. (2006). Mean Reversion of Short-Horizon Stock Returns : Asymmetry Property, 137–163.

Kyle, A. S. (1985). Continuous Auctions and Insider Trading. *Econometrica, 53*(6), 1315–1336.

Lehmann, B. N. (1990). Fads, Martingales and Market Effciency. *The Quaterly Journal of Economy*, *CV*(1).

Li, H., & Hong, Y. (2011). Financial volatility forecasting with range-based autoregressive volatility model. *Finance Research Letters*. https://doi.org/10.1016/j.frl.2010.12.002

Lo, A. (2004). The adaptive market hypothesis: market efficiency from an evolutionary perspective. *The Journal of Portfolio Management*, *30*(5), 15–29. Retrieved from http://opim.wharton.upenn.edu/~sok/papers/l/JPM2004.pdf

Madura, J., & Richie, N. (2004). Overreaction of Exchange-Traded Funds During the bubble of 1998-2002. *The Journal of Behavioral Finance*, *5*(2), 91–104.

Maheshwari, S., & Dhankar, R. S. (2014). A Critique of Overreaction Effect in the Global Stock Markets over the past three decades. *IOSR Journal of Business and Management Ver. V*, *16*(4), 2319–7668. Retrieved from www.iosrjournals.org

Malkiel, B. G. (1985). *A random walk down Wall Street*. Norton. Retrieved from https://www.mendeley.com/import/?url=http://oden.cbuc.cat/mendeley/enviaamendeley.php?bibid=.b1105887%26inst=UB%26llen=cat

Puckett, A., & Yan, X. S. (2011). The Interim Trading Skills of Institutional Investors. *The Journal of Finance*, *66*(2), 601–633. https://doi.org/10.1111/j.1540-6261.2010.01643.x

Samuelson, P. A. (1965). Proof That Properly Anticipated Prices Fluctuate Randomly. *Industrial Management Review*. https://doi.org/10.1109/SP.2006.16

Shan, Y., Taylor, S., & Walter, T. (2014). The role of “other information” in analysts’ forecasts in understanding stock return volatility. *Review of Accounting Studies*, *19*(4), 1346–1392. https://doi.org/10.1007/s11142-013-9272-5

Shiller, R. J. (2014). The Stock Market in Historical Perspective, 1–10. Retrieved from https://press.princeton.edu/chapters/s10421.pdf

Shleifer, A., & Summers, L. (1990). The Noise Trader Approach to Finance. *Journal of Economic Perspectives*, *4*(2), 19–33. https://doi.org/10.1257/jep.4.2.19

Smith, A. (1776). The Wealth of Nations: An Inquiry into the Nature and Causes of the Wealth of Nations, 842. https://doi.org/10.7208/chicago/9780226763750.001.0001

Spierdijk, L., Bikker, J. A., & Hoek, P. Van Den. (2010). Mean Reversion in International Stock Markets : An Empirical Analysis of the 20th Century.

Tversky, A., & Kahneman, D. (1975). Judgment under Uncertainty: Heuristics and Biases. In *Utility, Probability, and Human Decision Making* (pp. 141–162). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-010-1834-0_8

Umutlu, M., Akdeniz, L., & Altay-Salih, A. (2010). The degree of financial liberalization and aggregated stock-return volatility in emerging markets. *Journal of Banking & Finance*, *34*(3), 509–521. https://doi.org/10.1016/j.jbankfin.2009.08.010

Wooldridge, J. M. (2012). *Introductory Econometrics: A modern approach* (5th ed.).

A Appendix

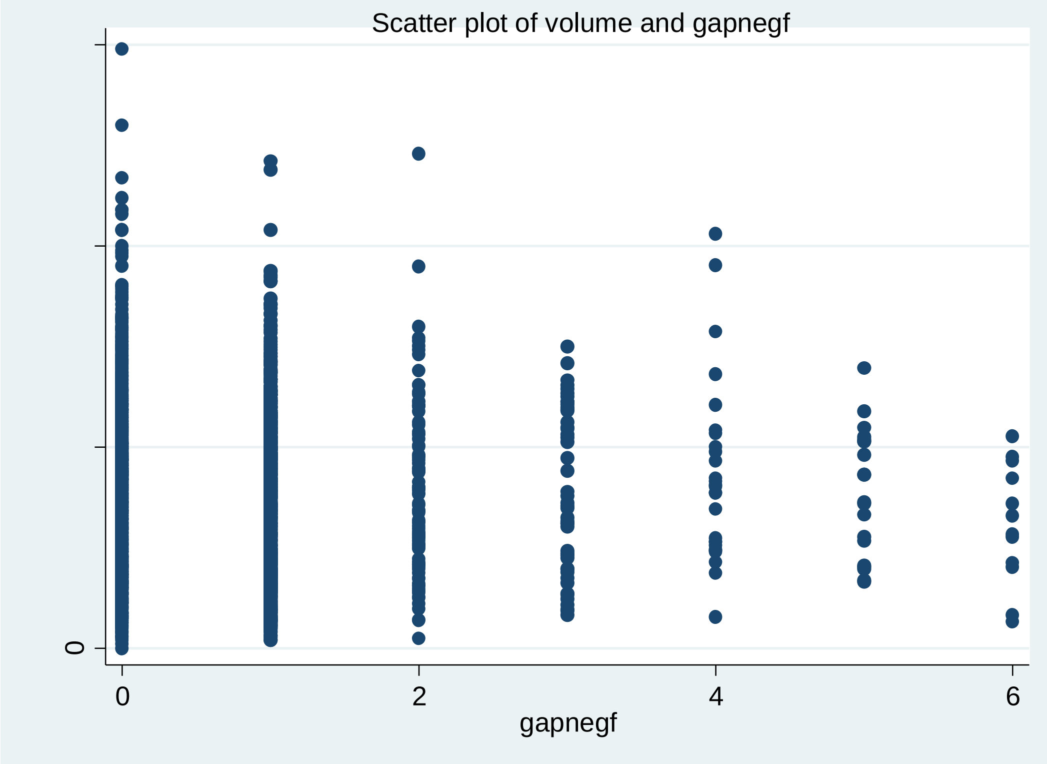


Figure A 1: Scatter plot of volume and gapnegf

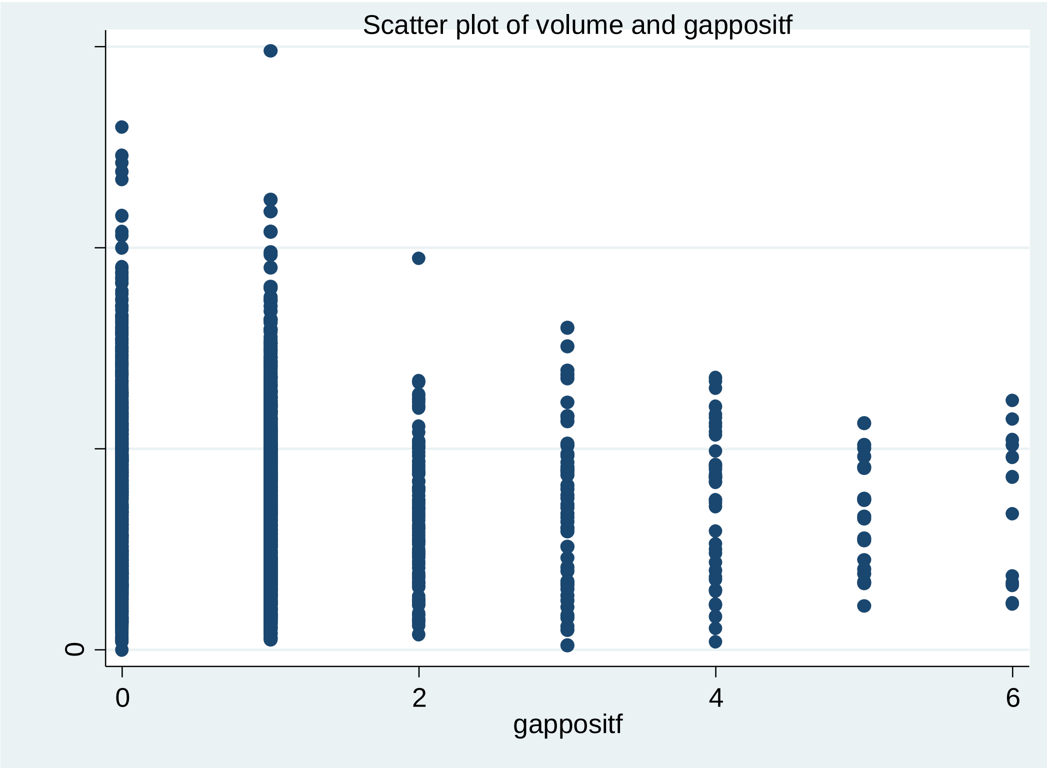


Figure A 2: scatterplot of volume and gappositf

```
Probit regression                               Number of obs   =        512
                                                LR chi2(2)      =     107.14
                                                Prob > chi2     =     0.0000
Log likelihood = -301.30808                     Pseudo R2       =     0.1509
```

gappositf	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sizeabs	-.0017579	.0002273	-7.73	0.000	-.0022033	-.0013124
rangeperc	36.69234	5.224267	7.02	0.000	26.45297	46.93172
volumeHP	9.91e-11	4.08e-11	2.43	0.015	1.92e-11	1.79e-10
_cons	-.3534032	.1941384	-1.82	0.069	-.7339075	.0271011

Table A 1: probit estimation for positive gaps

```
Probit regression                               Number of obs   =        447
                                                LR chi2(2)      =      82.23
                                                Prob > chi2     =     0.0000
Log likelihood = -268.39841                     Pseudo R2       =     0.1328
```

gapnegf	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sizeabs	-.0017296	.0002483	-6.97	0.000	-.0022163	-.0012429
rangeperc	30.72199	5.152196	5.96	0.000	20.62387	40.82011
volumeHP	1.56e-10	4.61e-11	3.38	0.001	6.57e-11	2.47e-10
_cons	-.4312963	.2142232	-2.01	0.044	-.8511661	-.0114266

Table A 2: probit estimation for negative gaps

```
Logistic regression                               Number of obs   =        512
                                                  LR chi2(2)      =     107.22
                                                  Prob > chi2     =     0.0000
Log likelihood = -301.26527                       Pseudo R2       =     0.1511
```

gappositf	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sizeabs	-.0029263	.0003995	-7.33	0.000	-.0037092	-.0021434
rangeperc	62.2134	9.36401	6.64	0.000	43.86028	80.56653
volumeHP	1.68e-10	6.76e-11	2.49	0.013	3.58e-11	3.01e-10
_cons	-.6201924	.3244488	-1.91	0.056	-1.2561	.0157156

Table A 1: logit estimation for positive gaps

```
Logistic regression                               Number of obs   =        447
                                                  LR chi2(2)      =      84.43
                                                  Prob > chi2     =     0.0000
Log likelihood = -267.29939                       Pseudo R2       =     0.1364
```

gapnegf	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sizeabs	-.0030532	.0004628	-6.60	0.000	-.0039603	-.002146
rangeperc	53.78697	9.32037	5.77	0.000	35.51938	72.05456
volumeHP	2.71e-10	7.79e-11	3.48	0.001	1.18e-10	4.24e-10
_cons	-.7340211	.3544916	-2.07	0.038	-1.428812	-.0392302

Table A 4: logit estimation for negative gaps

Measures of Fit for probit of gappositf

Log-Lik Intercept Only:	-354.876	Log-Lik Full Model:	-301.308
D(508):	602.616	LR(2):	107.135
		Prob > LR:	0.000
McFadden's R2:	0.151	McFadden's Adj R2:	0.140
Maximum Likelihood R2:	0.189	Cragg & Uhler's R2:	0.252
McKelvey and Zavoina's R2:	0.348	Efron's R2:	0.187
Variance of y*:	1.534	Variance of error:	1.000
Count R2:	0.678	Adj Count R2:	0.350
AIC:	1.193	AIC*n:	610.616
BIC:	-2566.453	BIC':	-94.659

Table A 2: measures of fit for probit of positive gaps

Measures of Fit for probit of gapnegf

Log-Lik Intercept Only:	-309.513	Log-Lik Full Model:	-268.398
D(443):	536.797	LR(2):	82.230
		Prob > LR:	0.000
McFadden's R2:	0.133	McFadden's Adj R2:	0.120
Maximum Likelihood R2:	0.168	Cragg & Uhler's R2:	0.224
McKelvey and Zavoina's R2:	0.364	Efron's R2:	0.179
Variance of y*:	1.572	Variance of error:	1.000
Count R2:	0.687	Adj Count R2:	0.349
AIC:	1.219	AIC*n:	544.797
BIC:	-2166.637	BIC':	-70.025

Table A 3: measures of fit for probit of negative gaps

Measures of Fit for logit of gappositf

Log-Lik Intercept Only:	-354.876	Log-Lik Full Model:	-301.265
D(508):	602.531	LR(2):	107.221
		Prob > LR:	0.000
McFadden's R2:	0.151	McFadden's Adj R2:	0.140
Maximum Likelihood R2:	0.189	Cragg & Uhler's R2:	0.252
McKelvey and Zavoina's R2:	0.313	Efron's R2:	0.188
Variance of y*:	4.791	Variance of error:	3.290
Count R2:	0.676	Adj Count R2:	0.346
AIC:	1.192	AIC*n:	610.531
BIC:	-2566.538	BIC':	-94.744

Table A 7: measures of fit for logit of positive gaps

Measures of Fit for logit of gapnegf

Log-Lik Intercept Only:	-309.513	Log-Lik Full Model:	-267.299
D(443):	534.599	LR(2):	84.428
		Prob > LR:	0.000
McFadden's R2:	0.136	McFadden's Adj R2:	0.123
Maximum Likelihood R2:	0.172	Cragg & Uhler's R2:	0.230
McKelvey and Zavoina's R2:	0.351	Efron's R2:	0.183
Variance of y*:	5.068	Variance of error:	3.290
Count R2:	0.689	Adj Count R2:	0.353
AIC:	1.214	AIC*n:	542.599
BIC:	-2168.835	BIC':	-72.223

Table A 8: measures of fit for logit of negative gaps