

MSc. Thesis

JSE Market Micro-Structure

Brett Schorn Du Preez
Supervisor: Dr Diane Wilcox
Co-supervisor: Dr Timothy Gebbie

Programme in Advanced Mathematics of Finance, University of the
Witwatersrand, Johannesburg.



16 January 2015

Declaration

I declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Science in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University.

Brett Schorn Du Preez
16 January 2015

Abstract

Stylized facts play a significant role in the testing whether models agree with known statistical anomalies and phenomena that occur in financial markets or not. Thus, we can use these stylized facts as a modeling tool or just to understand the general behavior of financial markets better. In the paper by Bouchaud *et al* in 2004 [1] we see the promotion of a new stylized fact that correlations in trade signs fail to die out, even after large lags. In fact, Bouchaud *et al* expressed the correlations as a slow power-law decay over trade ticks. In the results of our empirical study of JSE and BM&FBOVESP we find that the selected stocks show the this same power-law decay of correlations of trade signs. We also find that the stocks behave in a way which may allow for price manipulation at high enough trading rates as discussed by Gatheral [2].

For my parents
Michael and Marlene

Acknowledgement

I would like to thank Prof. Wilcox, for providing Bloomberg and Thomson-Reuters Tick History as part of her research programme funding and allowing me to work on the mini supercomputer she built using her NRF research grant, without which my project would not have been possible. The computational side of my project was done in MATLAB provided by the AMF program.

I would also like to express my sincere appreciation to my supervisory team, Prof. Wilcox and Dr. Gebbie, for their tireless assistance, encouragement and their support through hard times.

Thanks to Michail Scholiadis for helping with proof reading of my finalized thesis.

Finally I would like to thank the NRF for the funding that allowed me to conduct my research.

Contents

1	Introduction	9
1.1	Efficient Market Hypothesis vs “Zero Intelligence Agent” Model	9
1.2	Market Forces	10
1.2.1	Lit Order-book Market	10
1.2.2	Dark Pools	12
1.3	Limit Order-book Dynamics	13
1.4	Optimal Trading Strategy	13
2	Preliminary Market Micro-Structure	16
2.1	Introduction	16
2.2	Intra-Day Volumes	17
2.3	Intra-Day Returns	21
2.4	Intra-Day Spreads and Liquidity	32
2.5	Basic Price Impact Analysis	39
2.6	Summary and Conclusions	44
3	Trade Sign Behavior	46
3.1	Auto-Correlation of Trade Signs	46
3.2	Trade Sign Clustering	50
4	Cost of Trading Models	55
4.1	Special Cases	56
5	Price Impact Statistics	58
5.1	Introduction	58
5.2	Diffusion Function	60
5.3	Empirical Response Function and Market Impact	63
5.4	Distribution of Impact Variables	68
5.5	Stock Price Modelling for Price Impact	71
5.6	Fitting a Bare-Impact Function	72
5.6.1	Fitting β	73
5.6.2	Fitting ℓ_0 and R_0	77
5.6.3	BGPW Model Fitting Results	78
6	Conclusion	80

A	Stocks Used In Study	90
A.1	JSE Stocks	90
A.2	BM&FBOVESPA Stocks	92

List of Tables

1	Intra-day Trade Statistics	22
2	Intra-day five minute return statistics	25
3	Detrended daily average effective spread statistics	39
4	Trade sign statistics	50
5	AGL u_ℓ statistics	69
6	Fitting statistics	79

1 Introduction

In empirical studies on stock price data, such as [1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21], one finds several statistical anomalies in the behavior of the stock price. Many of these have been quantified or have been explained and formed into stylized facts.

In 2004, Bouchaud *et al* [1] formulated a new stylized fact that trade sign correlations die off over a relatively long period as a slow power-law decay. This new stylized fact lead Bouchaud *et al* [1] to contradict the widely accepted idea that market impact is permanent. In Section 5.5 we discuss how a permanent would cause the average response to be significantly amplified with a large lag which we do not see in the empirical results. They provided a strong argument suggesting that market impact has to be temporary and suggested that it should decay as a power-law. Subsequent to the Bouchaud *et al* [1] study, the effect has anecdotally been significantly reduced by market activities¹. However, the importance of quantifying such effects remains, particularly in the context of emerging markets such as South Africa and Brazil, where market developments have lagged those in developed markets.

To understand the complexity of the correlation, one has to first understand the dynamics of the limit order-book and the process involved in the optimal execution of a trade, explained by Obizhaeva and Wang in 2006 [22] and Almgren and Chriss in 2000 [23], the competition between liquidity takers and liquidity providers, and how these affect the stock price over a short time interval.

In what follows, we take a look at some of the fundamental ideas behind some conclusions of past empirical studies as well as some practical applications of these ideas.

1.1 Efficient Market Hypothesis vs “Zero Intelligence Agent” Model

The Efficient Market Hypothesis states that all of the information available in a market will be reflected in the fair-value price of a stock and the price

¹Private communication with T. Gebbie.

emerges for a consensus amongst rational market participants who would be likely to arbitrage away any deviation from the fair price [24, 25]. Thus price changes would be the result of unpredictable changes in fundamental information. Thus, the current stock price would be the best predictor of its future price and therefore, the stock price should be a random walk in time.

In financial markets, we see stock prices exhibiting random walk like behaviors. However, the volatilities observed in financial markets appear to be too high to be compatible with rational pricing. We also see that the frantic movement in the stock prices fails to fit the hypothesis and the assumptions of fully informed agents and a fair price seem too strong to be realistic.

The *Zero Intelligence* agent model suggests that prices follow a pure random walk because, if an agent has zero intelligence and decides randomly whether to buy or sell, their actions will be interpreted by other agents as containing some information [26, 27, 28, 29, 30, 31, 32]. The mere fact that the agent bought or sold would cause a shift in the stock price to a new reference price from which the process could repeat itself, giving the stock price a random walk like behavior.

Although both ideas seem to be rather impractical, they can be used as viewpoints from which various anomalies can be understood. The true behavior of a stock price should lie somewhere in the middle of these two ideas. If we take a closer look at what drives the stock price, we find that it is not a trivial random walk, but more a fine-tuned competition between opposing market forces. Thus, it is vital that we understand the behavior of market forces before we can understand the behavior of a stock price.

1.2 Market Forces

1.2.1 Lit Order-book Market

Market forces are the forces of supply and demand representing the influence of market participants on the price and liquidity of equity in a market. On exchanges, these forces are proxied limit orders and market orders. Limit orders are orders to buy or sell a specified volume of a stock at a specified

price, whereas market orders are orders to buy or sell a specified volume of a stock without a specified price. It is through these two types of orders that we can separate traders into two broad categories: liquidity takers, and liquidity providers.

Liquidity takers use market orders to initiate trades. In general, these trades are time sensitive and are entered for immediate execution. Liquidity takers often try to anticipate future movements in stock prices due to some information they believe to be asymmetric and in their trade's favor. They will try to take full advantage of the information for profit. Often this information is misinterpreted and does not affect the stock price at all, in which case the bid-ask spread is lost. Alternatively it could have the opposite effect to what was expected and more may be lost from the transaction.

Investors who are impatient or need to hedge or liquidate their position as a matter of urgency take advantage of the fact that market orders are executed immediately, even though it costs them the bid-ask spread.

Liquidity providers use limit orders to avoid taking unintended naked positions. Their limit orders provide liquidity in the market by increasing the volumes in the order-book and thus giving the market orders something to execute against. Liquidity providers are able to make small margins of profit from the bid-ask spread if the stock price doesn't change while the order remains within the order-book queue since the liquidity takers will buy for higher than they sell for.

Now we can give some thought to the competition between liquidity takers and liquidity providers. Let's assume that a large number of market orders to buy enter the system. The liquidity providers could see this as the liquidity takers being informed that the stock is under priced and they might want to increase their ask price because the stock price is likely to go up in the near future if the liquidity takers are correctly informed. The liquidity takers would then have to pay a considerably higher price for the stock and they may not benefit from the transaction in the end. Thus the more prudent approach would be to divide one's larger orders into a number of smaller trades and spread them as far as possible across the day so that as much information as possible can be with-held from the liquidity providers. If the liquidity takers' views do not become transparent through the trade and no information can be taken out of the trade the stock price should settle at its

previous value. A mean reversee behavior can be seen in stock prices caused by liquidity providers closing out their position and trying to keep the price from moving as it would be costly for them if the stock price had to shift too far from its original position.

1.2.2 Dark Pools

A dark pool of liquidity is an order matching facility which allows institutions to execute large block [33] trades with a degree of anonymity and reduced trading costs as a result of lower slippage. Dark pools are not available to the public, but are used mainly by financial institutions when executing large block trades or shifting large volumes of equity between funds. These dark pool transactions could be executed through an electronic communication system which bypasses the exchange when matching orders, or an over the counter transaction between market participants.

The primary advantage of using a dark pools, especially for large institutions, is to trade large volumes of stocks while withholding information that is key to the institution's operations from other large institutions who would be likely to take advantage of this information. The fact that the size of the trade is not revealed until the price is agreed upon by the participants would also appeal larger institutions as it would reduce the market impact of the trade.

The three main types of dark pools are comprised of: independent companies, broker-owned dark pools and dark pools created by public exchanges. Independent company dark pools such as: Instinet Crossing, POSIT and Liquidnet, among others, provide the platform for institutions to execute their block trades outside of an exchange environment. Broker-owned dark pools allow the clients to interact with each other without disclosing any information prior to the agreement to execute a trade. Some examples of broker-owned dark pools include: CITI markets and banking, Credit Suisse and CrossStream. Public exchange dark pools, like International Securities Exchange and NYSE Euronext, allow market larger participants to execute their large volume trades in a similar environment to the exchange, but with the added benefit of anonymity.

The main purpose of dark pools is to allow large volumes to be traded at the current fair value, where as lit markets are used to determine the current

fair value. These are two different yet essential components of an effective market.

1.3 Limit Order-book Dynamics

In their paper in 2006 [22] Obizhaeva and Wang considered the limit order-book market. This market is arguably the closest to the definition of a what one would refer to as a centralized market. They suggested that if one takes a one-sided view the rate at which limit order-book would converge, to a so called steady state after a trade has been executed is exponential (provided that there are no fundamental shocks). The density of the steady state has the form:

$$q_t(P) = q1_{\{P \geq A_t\}}, \quad (1)$$

where t is relatively large, $A_t = V_t + s/2$ and $V_t = F_0 + \lambda x_0$, F_t is the fundamental value of the stock, A_t is the ask price, V_t is the mid-price and s is the spread at time t . Directly after a trade they suggest that the ask price will have the form:

$$A_t = V_t + s/2 + x_0(1/q - \lambda)e^{-\rho t}, \quad (2)$$

where x_0 is the size of the initial trade and $\rho \geq 0$ is the speed at which the order-book converges and $V_t = V_{0+}$ in absence of new trades and changes in the fundamental value, which can be used when measuring the resilience of a limit order-book. This is an important idea to appreciate, as most impact models assume that the order-book will reach a new steady state directly after a trade has taken place. One can see that a more realistic view would be to work under the assumption that the order-book will converge in some manner to a new steady state, but not immediately after a trade has been executed because the executed trade would affect the supply of the stock and limit orders would flow in to meet the gap in supply.

1.4 Optimal Trading Strategy

In previous optimal execution studies [21, 34, 35, 36, 37, 38, 39, 40, 41, 42], we see numerous methodologies, strategies and systems for optimal order

control. One of the focus in these studies is optimizing the execution of orders in order to reduce the cost of trading. In this study we look at the optimal trading strategy discussed by Obizhaeva and Wang [22].

Using the idea of the flow of limit orders Obizhaeva and Wang [22] discuss a so called “optimal” trading strategy, that is a trading strategy that would result minimal costs when executing a given trade. They suggested that an optimal trading strategy should be a combination of discrete trades coupled with a continuous stream of trades.

Using this framework they discovered that the optimal way of executing a trade would involve a discrete trade at the beginning of the trading period large enough to entice new limit orders into the order-book, but not high enough to incur unnecessary costs, followed by a continuous stream of trades just large enough to execute against the incoming limit orders and a large discrete trade towards the close of the trading period to execute the remaining portion of the trade.

Upon closer investigation they found that the size of the discrete trades, and consequently the cost of the trading strategy, were inversely proportional to the length of the trading period.

This provides a reasonable explanation for the long-term correlation between the trade signs. One of the shortfalls of this idea is the fact that they do not consider the interference caused by other trades i.e. they look at a market with only one market order trader.

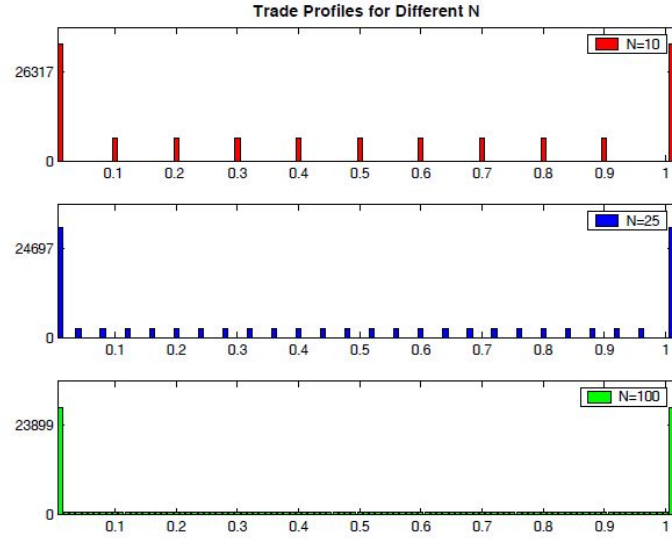


Figure 1: Optimal execution strategy, found by Obizhaeva and Wang [22] , with fixed discrete trade intervals. This figure plots the optimal trades for trade periods of 10, 25 and 100 for respectively the top, middle and bottom panels.

2 Preliminary Market Micro-Structure

Market architecture, as described by Madavahn [43] is the set of rules governing the trading process. This architecture has facets that are market specific, but [44] suggests that there are some statistical rules that govern market structure across markets. Understanding market micro-structure and architecture is necessary for risk management when trading on multiple trading venues and in managing optimal trade execution services. It is also necessary for understanding the relationship between regulations, high-frequency trading and market feed-backs. In this section we examine the behavior of intra-day trade statistics and compare our results with previous studies. We do so for two markets: the JSE and the BM&FBOVESPA. We discuss the possible causes of certain behaviors and the implications of our findings. We also examine some basic forms of price impact where to find that a one size fits all approach is not always suitable; with factors such as liquidity and market capitalization playing roles in determining the shape of a price impact function.

2.1 Introduction

Following the work done by O'Hara [45], Hasbrouck [46] and Madhavan [43] among others, in this section we will investigate some of the fundamental micro-structures on the JSE. We will look at the shapes of intra-day volume curves, intra-day return curves and other basic indicators of market behavior. One of our main interests in this section is to better understand how price movement relates to traded volumes. Although one would expect to find that a large volume traded would have a greater effect on the price than a smaller volume, this is not the case. It is often noted that a low volume can be associated with a large price movement and a high volume with a small price movement.

In this section we have decided to select five indicative stocks out of the 46 available for the JSE data so that our plots can be interpreted more efficiently by the reader. These five stocks include:

- Anglo American Plc, a diversified mining and natural resource groups.
- Aspen Pharmacare, a supplier of branded and generic pharmaceuticals in approximately 100 countries.

- Bhp Billiton Plc, a diversified natural resources company.
- Mtn Group, a telecommunications provider.
- Standard Bank Group Limited, a banking and financial services provider.

These five have been selected as some have specific similarities while differing from the others. AGL and BIL are very liquid stocks with a large market cap, MTN and SBK are what some would call “super” liquid stocks, and APN is one of the more illiquid stocks in the Top40. This gives us a good mix of related and unrelated industries, market caps, and liquidities for comparison.

2.2 Intra-Day Volumes

In past empirical studies [1, 9, 47, 48, 49, 50, 51, 46, 52, 44, 53] it has been found that intra-day volumes generally have a U-shape with some seasonal events which affects the volume of most stocks. To avoid averaging out the effects of the different opening times in Europe and USA during daylight savings time, we decided to split the two time periods and look at the data in each separately.

In Figure (2) we only look at trades that take place while both Europe and the US are not in daylight savings time. For this we had to exclude the data where any daylight saving is taking place, in particular the data belonging to the period starting 14 March 2010 and ending 07 November 2010. We can clearly see the U-shape that has come to be expected when looking at intra-day volume curves. One can also clearly see the effects of the European markets opening on the volume traded. One sees no influence from the US on South African markets outside of daylight savings time as their markets open as our trading day comes to an end.

In Figure (3) we only look at trades that take place while both Europe and the US are in daylight savings time. Thus, we look only at the data for which the two daylight saving periods intersect in 2010, which is period starting 28 March 2010 and ending 31 October 2010. Again we can clearly see the U-shape that has come to be expected when looking at intra-day volume curves. The influence on volume of European markets opening during daylight savings time can be seen as our stocks come out of the morning

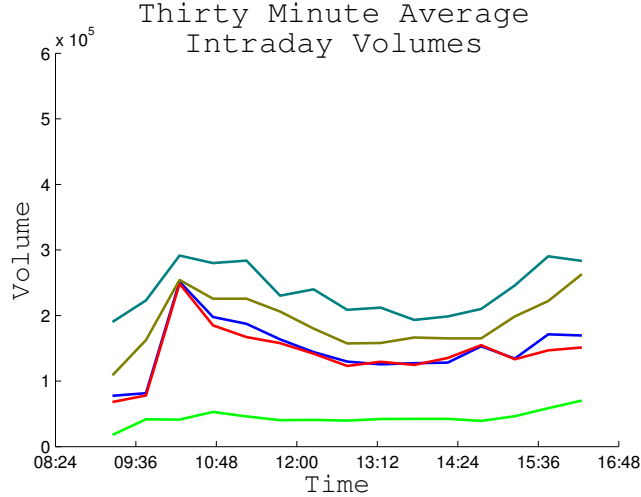


Figure 2: Thirty minute average volume curves for AGL, APN, BIL, MTN and SBK. This graph was done for the period outside of daylight savings time in 2010, excluding data belonging to the period starting on the 14 March 2010 and ending on the 07 November 2010.

auction period. In this period the effects of the US markets opening before our afternoon action period is important, but one also has to take in to account the natural rise in trading activity as we approach the afternoon auction. These together have a large effect on the volume traded.

In Figure (4) we only look at trades that take place while both Europe and the US are not in daylight savings time. For this we had to exclude the data where any daylight saving is taking place. To achieve this, we excluded the data belonging to the period starting 14 March 2010 and ending 07 November 2010. After excluding daylight savings, we took the volumes traded in thirty minute periods normalized by total volume traded on the day and averaged these thirty minute quantities across the year. In this average relative volume curve we get a better idea of how similar the behavior of intra-day volume is for different stocks. In particular the behavior of APN becomes more similar to the other stocks. The effects of foreign markets discussed in the average daily volume curves are more obviously seen in the average relative volume curve. It is interesting to note that the stocks with similar market cap and liquidity are almost indistinguishable outside of daylight savings time.

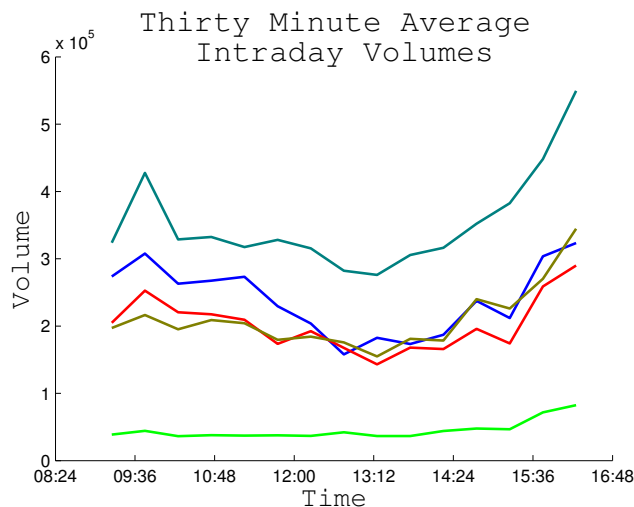


Figure 3: Thirty minute average volume curves for AGL, APN, BIL, MTN and SBK. This graph was done for the period of daylight savings time in 2010, so we have used data belonging to the period starting on the 28 March 2010 and ending on the 31 October 2010. The influence on volume of European markets opening during daylight savings time can be seen as our stocks come out of the morning auction period.

In Figure (5) we only look at trades that take place while both Europe and the US are in daylight savings time. Thus we look only at the data for which the two daylight saving periods intersect in 2010, which is period starting on the 28 March 2010 and ending on the 31 October 2010. For this graph we followed the same averaging procedure as we did for the non-daylight savings time data. In this average relative volume curve we get a better idea of how similar the behavior of intra-day volume is for different stocks. The effects of foreign markets discussed in the average intra-day volume curves are more obviously seen in the average relative volume curve.

When one looks at the seasonal effects in the JSE average intra-day volume curves for daylight savings time and non-daylight savings time, we see a significant increase in the amount of volume being traded when foreign markets, such as Europe and USA, open. In Figure (3) and Figure (2) one can see a clear U-shape in both volume curves which is consistent with previous empirical findings.

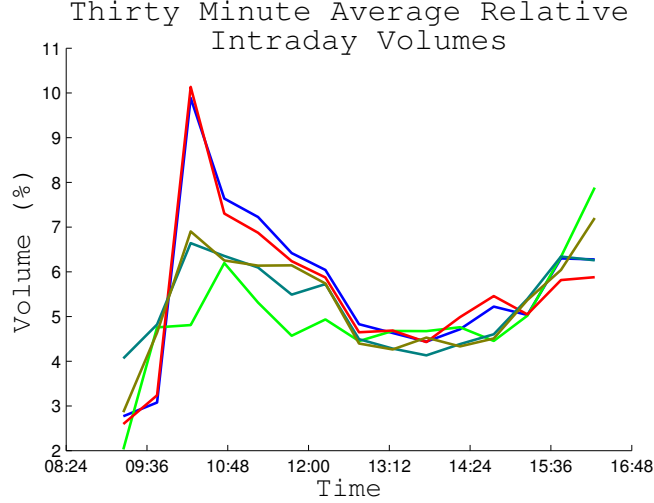


Figure 4: Thirty minute average relative volume curves for AGL, APN, BIL, MTN and SBK. This graph was done for the period outside of daylight savings time in 2010, so we have excluded data belonging to the period starting on the 14 March 2010 and ending on the 07 November 2010. In this average relative volume curve we get a better idea of how similar the behavior of intra-day volume is for different stocks.

In Table (1) we have tabulated some indicative values related to volume and liquidity of the five stocks which we have chosen. One can see that the most liquid stock MTN has an average of 2405 trades executed each day with an average of 5722900 shares traded each day and SBK which is also a very liquid stock has slightly less, but they have similar prices. On the other end of the scale we have APN which is one of the less liquid stocks in our data only executing an average of 770 trades per day with an average of 86470 shares traded each day. We also see AGL and BIL having large quantities of volume being moved, but they also have fairly high stock prices. AGL and BIL are both liquid stocks, but they are not as liquid as MTN and SBK. The liquidity of each stock seems to be influenced by not only the movement of volume but also the price of the stock. It is important to note that we are taking a market-order perspective and that the direction of trading has been ignored.

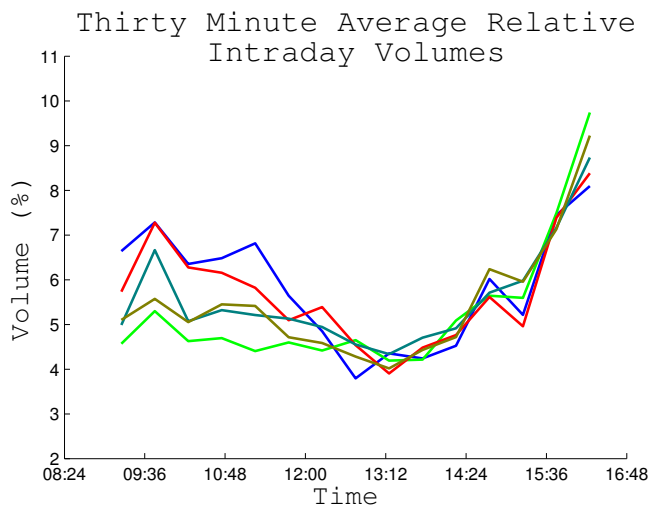


Figure 5: Thirty minute average relative volume curves for AGL, APN, BIL, MTN and SBK. This graph was done for the period of daylight savings time in 2010, so we have used data belonging to the period starting on the 28 March 2010 and ending on the 31 October 2010. In this average relative volume curve we get a better idea of how similar the behavior of intra-day volume is for different stocks.

2.3 Intra-Day Returns

In Figure (6) we have plotted the five minute returns for AGL in the period starting on 10 March 2010 and ending 09 March 2011. The five minutes returns for AGL can be seen as noise. The spikes in the returns occur at times of market stress. Such as the flash crash on 06 May 2010.

The distribution of stock returns has been a topic which has been debated in many papers like [54, 55, 56, 57, 58], among others. Aparicio and Estrada [54] make a strong argument that daily returns do not follow a Normal distribution because of the leptokurtic behavior of said returns. In Figure (7) and (8) we see that intra-day returns also have a leptokurtic behavior. This could indicate that returns may have a leptokurtic behavior on multiple grains and may warrant further investigation. Leptokurtic behavior has also been viewed as the result of looking at price movement independently

Ticker	<i>AGL</i>	<i>APN</i>	<i>BIL</i>	<i>MTN</i>	<i>SBK</i>
Average Trades Per Day	2088	770	1540	2405	1789
Average Trade Price	31064	8459	23788	11870	10712
Average Volume Per Day	3519495	869470	3185600	5722900	3830400

Table 1: Intra-day trade statistics for AGL, APN, BIL, MTN, SBK. In this table we have displayed some of the trade statistics for our five indicative stocks.

from volume; that once prices changes are corrected for volume effects, the distribution can become more Normal [59].

The intra-day pattern in price variance was explained by Madhavan [60] in terms of a structural model. They argued that the volatility is significantly higher earlier in the day because of the higher inflow of information, which is a reasonable explanation for the peaks in both average returns and average absolute returns in the morning and afternoon auctions.

The intra-day average absolute returns for non-daylight savings and daylight savings time are shown in Figure (11). The average absolute returns seem to be closely correlated for similar stocks and seasonal effects were obvious in all five stocks. What is interesting is the fact that all five stocks seem to have the same shape average absolute returns in the daylight savings time data, which would also be the case for the non-daylight savings data if not for the spike in AGL and BIL at 10am. It is also interesting that the average absolute returns seem to have a U-shape which we see arising in intra-day statistics.

Hasbrouck [51] argued that the higher price volatility at market open is primarily due to higher volume of information being available. These studies both find the impact of trading to be more significant at market open and during the morning. This is clear in Figure (11) and Figure (12), as one can see in both figures price movements are generally larger in the mornings.

In Figures (13) and (14) we have plotted the volatility of five minute returns and thirty minute returns for each day in our data set. In order for us to compare the two sets of daily volatilities we have scaled the five minute

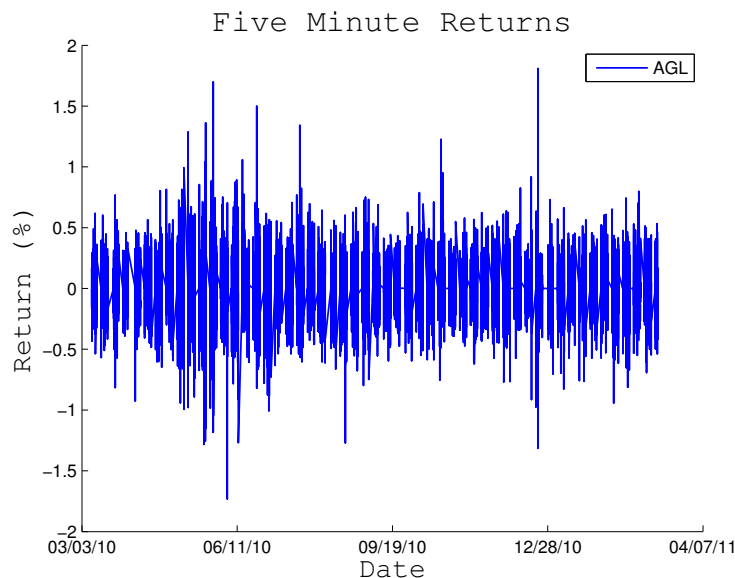


Figure 6: Five minute returns for AGL. Other than a few spikes, which we have assumed to be caused by minor crashes and fundamental shocks, the returns for AGL are what we have come to expect from return behavior. The five minute returns of other stocks displayed the same behavior as that of AGL.

return volatilities by a factor of root 6. We see, as one would expect, that the daily volatilities using five minute returns has a mean greater than that of the volatilities using thirty minute returns. This stems from the fact that as one takes larger intervals in any data set out of a time series, one sees a more prominent trend and the noise will be reduced. In terms of stock prices this comes about as a direct result of the mean reversion caused by the liquidity providers trying to keep the stock price from drifting too far from the current value.

It is interesting to note that for the four liquid stocks we see that the daily volatilities of the five minute returns all have a similar pattern which leads us to believe that seasonality plays a role in the volatility of intra-day returns. We also notice a few similarities when using the thirty minute returns, but what really stands out is the behaviour of the daily return volatilities in our illiquid stock. The daily return volatilities for APN are generally larger than those of the more liquid stocks, this is clearer when using thirty minute

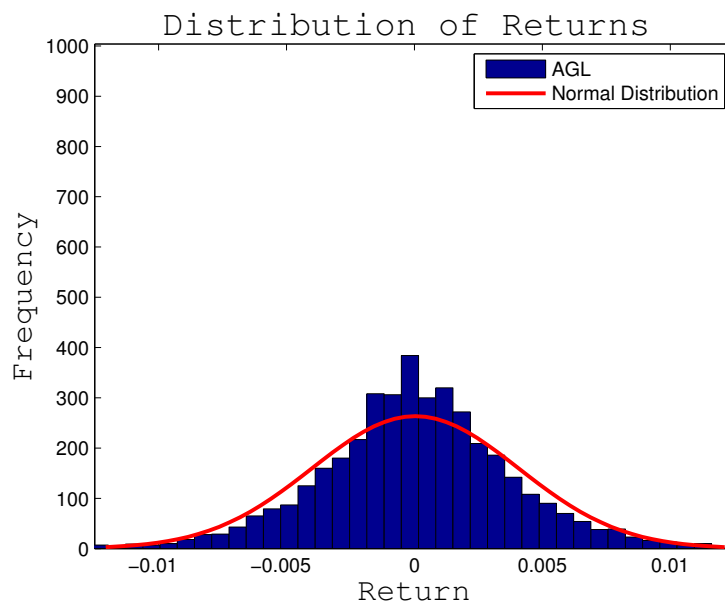


Figure 7: Distribution of thirty minute returns of AGL plotted against an equivalent Normal distribution.

returns, which leads us to believe that the spikes in the daily thirty minute return volatilities of the other stocks would point towards low liquidity on those days.

Through this we also learn that one should not choose a specific interval size when working with intra-day data, but rather use a few interval sizes as we see above. If we had taken only five minute return data we would have lost important information.

We noted in our study that there is no particular pattern with regards to the average five minute returns. As one sees in Figure (9) there is no distinguishable trend. As one can see in Table (2) that the means of returns are almost insignificant whereas the volatility is relatively large and similar for all but our illiquid stock.

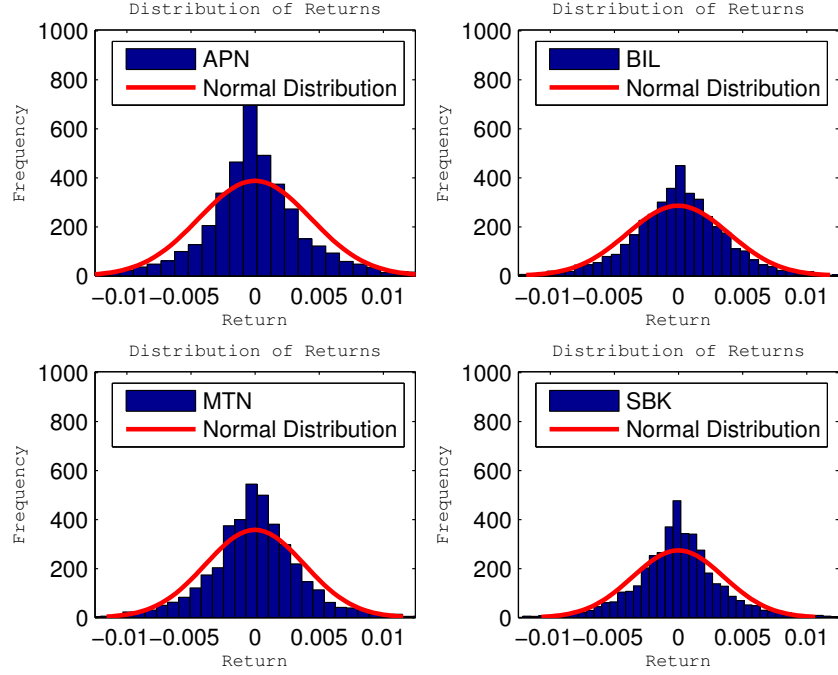


Figure 8: Distribution of thirty minute returns of APN, BIL, MTN and SBK plotted against an equivalent Normal distribution.

Ticker	<i>AGL</i>	<i>APN</i>	<i>BIL</i>	<i>MTN</i>	<i>SBK</i>
Mean	0.00054	-0.00012	-0.00052	-0.00017	-0.00032
Volatility	0.18	0.26	0.19	0.18	0.17
Skewness	0.0036	0.1676	-0.0764	0.0911	0.1357

Table 2: Intra-day five minute return statistics for AGL, APN, BIL, MTN, SBK. We have tabulated some return statistics for our five indicative stocks. We see that the illiquid stock has returns which are more volatile than the others which is to be expected. It is interesting to note that the other four stocks have returns with almost the same volatility.

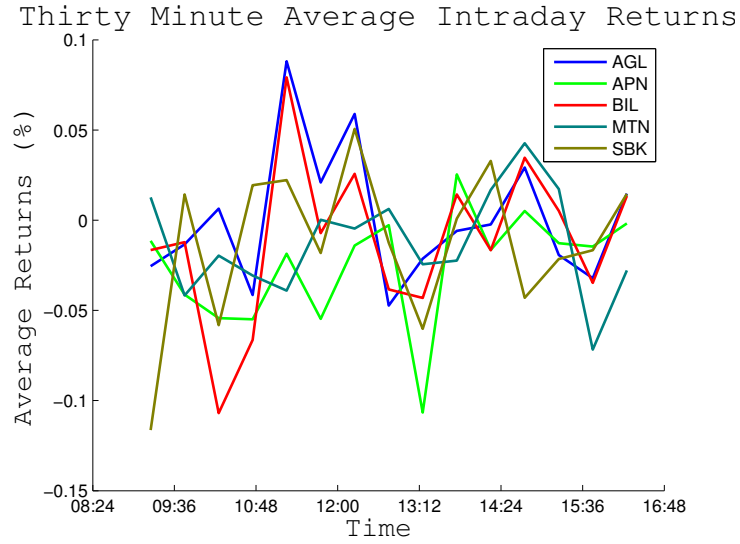


Figure 9: Thirty minute average returns for AGL, APN, BIL, MTN and SBK. This graph was done for the period outside of daylight savings time in 2010, so we have excluded data belonging to the period starting on the 14 March 2010 and ending on the 07 November 2010. In this figure we have plotted the returns of each stock for the sixteen thirty minute periods in our trading day, averaged across the year. One sees that, other than the few points where the returns of two or more stocks seem to behave similarly, there is no significant relationship between the intra-day returns of our five chosen stocks.

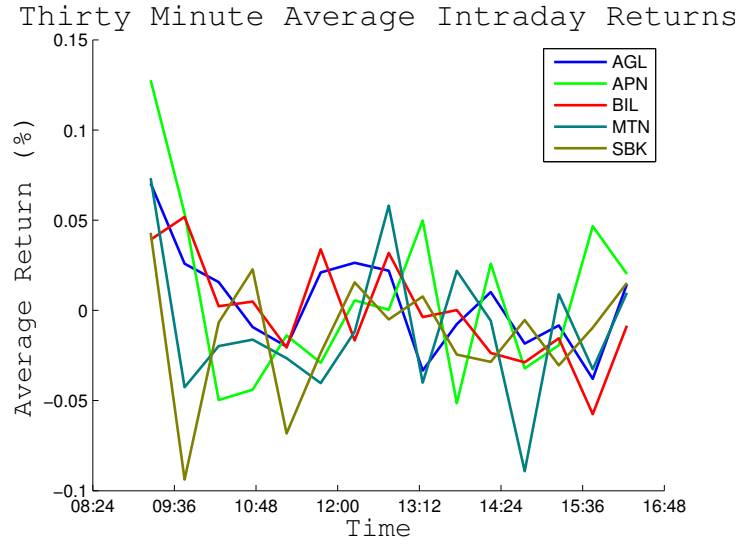


Figure 10: Thirty minute average returns for AGL, APN, BIL, MTN and SBK. This graph was done for the period of daylight savings time in 2010, so we have used data belonging to the period starting on the 28 March 2010 and ending on the 31 October 2010. In this figure we have plotted the returns of each stock for the sixteen thirty minute periods in our trading day, averaged across the year. One sees that, other than the few points where the returns of two or more stocks seem to behave similarly, there is no significant relationship between the intra-day returns of our five chosen stocks.

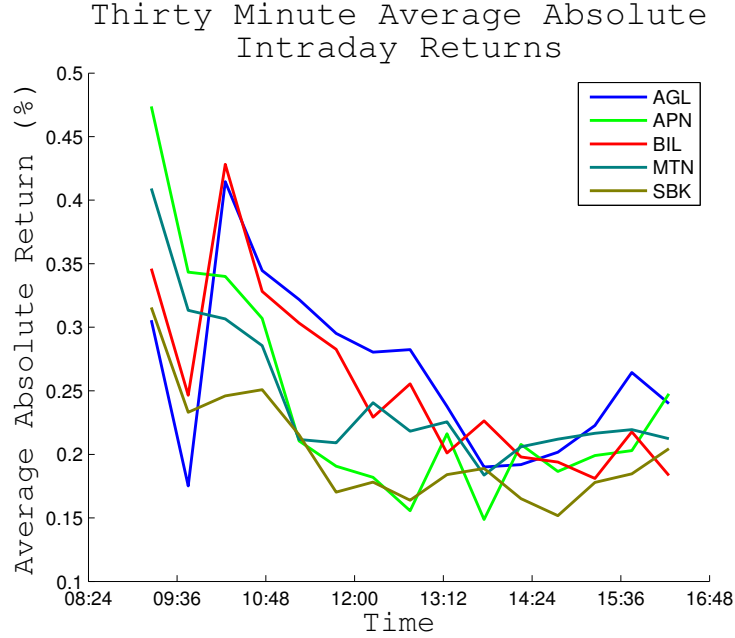


Figure 11: Thirty minute average absolute returns for AGL, APN, BIL, MTN and SBK. This graph was done for the period outside of daylight savings time in 2010, so we have excluded data belonging to the period starting on the 14 March 2010 and ending on the 07 November 2010. Using absolute average returns as a proxy for return volatility we see that volatility is high coming out of and going into auction periods, but we can also see the volatility does not remain constant during continuous trading which is similar to a result found by [61]. One can also see that, other than the spike in AGL and BIL, the return volatility for these five stocks behave similarly through the day.

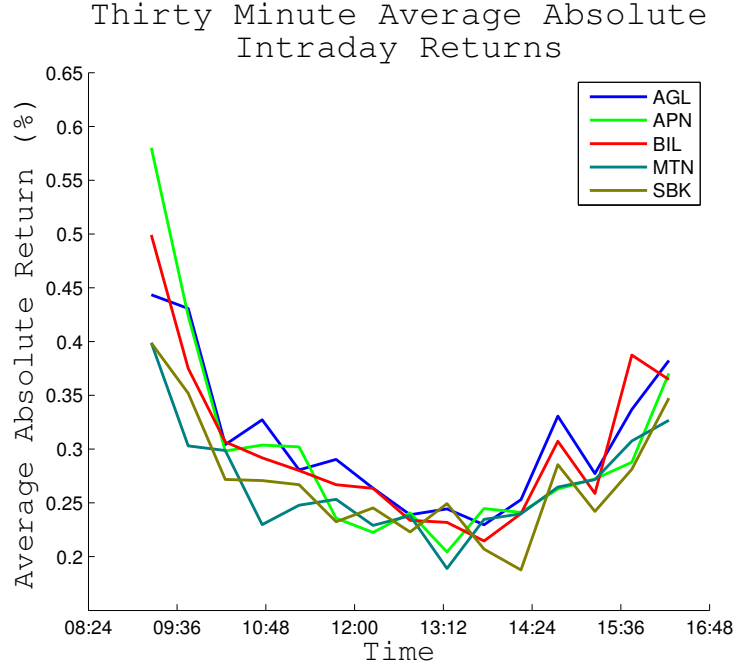


Figure 12: Thirty minute average absolute returns for AGL, APN, BIL, MTN and SBK. This graph was done for the period of daylight savings time in 2010, so we have used data belonging to the period starting on the 28 March 2010 and ending on the 31 October 2010. Using absolute average returns as a proxy for return volatility we see that volatility is high coming out of and going into auction periods, but we can also see the volatility does not remain constant during continuous trading which is similar to a result found by [61]. One can also see that the return volatility for these five stocks behave similarly through the day.

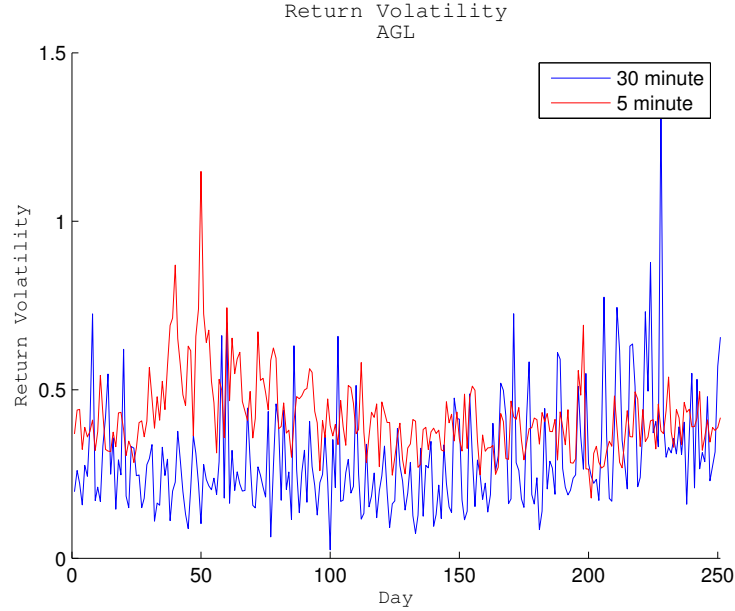


Figure 13: Daily Volatility of returns of AGL. In the figure we have plotted the volatility of five minute returns and thirty minute returns for each day in our data set. To avoid misinterpretation we have scaled our five minute return volatilities to a thirty minute scale so that we can compare the two sets of results. One can see that over the 251 days the mean of the average thirty minute intra-day return volatility is significantly smaller than that of our shorter time frame results. We noted that in the thirty minute return volatilities there is clear grouping of spikes after the 200th trading day in our period that is not present in the five minute return volatilities, and clear grouping of spikes around the 50th day in the five minute return volatilities that is absent from the thirty minute return volatilities. Along with these three differences, we also see that there is a significant difference in trend of the return volatilities through the year. Hence, there is no fixed relationship between volatilities over different intervals, confirming appropriateness of nondeterministic stochastic models (stochastic volatility or at least non constant local volatility).

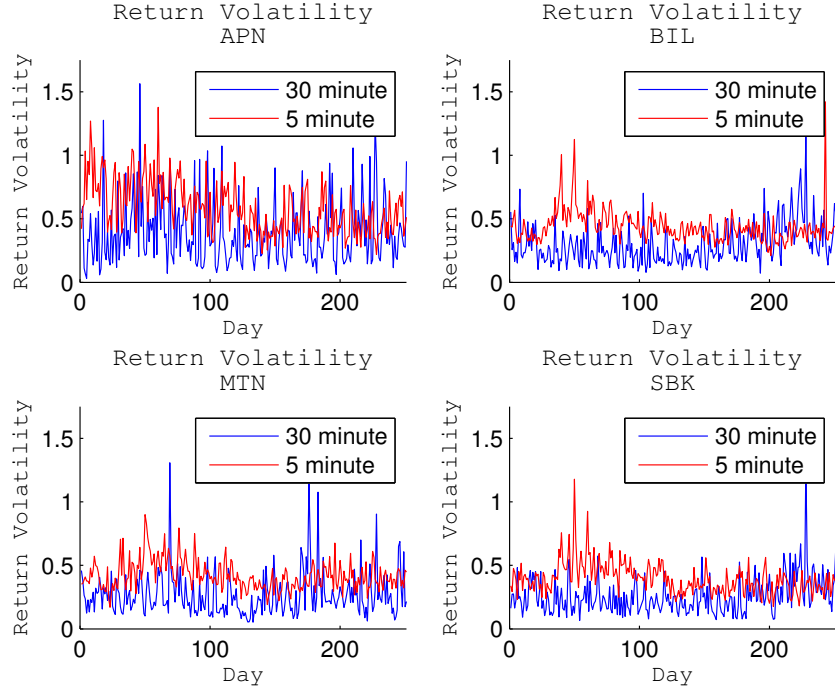


Figure 14: Daily Volatility of returns of APN, BIL, MTN and SBK. In the figure we have plotted the volatility of five minute returns and thirty minute returns for each day in our data set. To avoid misinterpretation we have scaled our five minute return volatilities to a thirty minute scale so that we can compare the two sets of results. One can see that over the 251 days the mean of the average thirty minute intra-day return volatility is significantly smaller than that of our shorter time frame results. We noted that in the thirty minute return volatilities there is clear grouping of spikes after the 200th trading day in our period that is not present in the five minute return volatilities, and clear grouping of spikes around the 50th day in the five minute return volatilities that is absent from the thirty minute return volatilities. Along with these three differences we also see that there is a significant differences in trend of the return volatilities through the year. Hence, there is no fixed relationship between volatilities over different intervals, confirming appropriateness of nondeterministic stochastic models.

2.4 Intra-Day Spreads and Liquidity

In financial markets liquidity can be defined as the degree to which a stock can be traded in the market with no affect on the stock's value. Many liquidity measures such as spreads and price impact, among others, are used to gauge the degree of liquidity displayed by an asset. Factors which contribute to liquidity include: transaction costs, rate of order execution as well as order-book depth and breadth and resilience.

Transaction costs consist of bid-ask spreads as well as other costs incurred when an order is executed, such as admin costs, slippage e.t.c. Rate of order execution refers to how fast an incoming trade is executed. This can be affected by communication latency if the trading sever is remotely located, as was the case for the JSE equity trading sever for the period we have focused on. Order-book resilience as we briefly explained in Section 1 refers to the rate at which new limit orders arrive to fill the order-book. Order-book depth refers to the range of prices available in the order-book. And the breadth of an order-book refers to the amount of volume available in said order-book. In Figure (15) we have an illustration of the examples of depth and breadth when taking a one-sided view of an order-book.

Sarr and Lybek [62] explained that liquid markets and, by extension, liquid stocks should exhibit the following characteristics:

- Low transaction costs which would require a small bid-ask spread and low implicit costs.
- Efficient order execution in other words incoming orders should be matched and cleared with minimal delay.
- Order-books should be deep and broad by having large volumes as well as a wide spread of order prices.
- Strong order-book resilience which would mean that limit-orders flow in at a rate high enough to keep the spreads and prices from drifting to far from its current position after an order has been executed.

One of the key factors in the liquidity of a stock is the behavior of its spreads. Preceding research has found that bid-ask spreads are largely determined by trading activity, incoming information, competition between

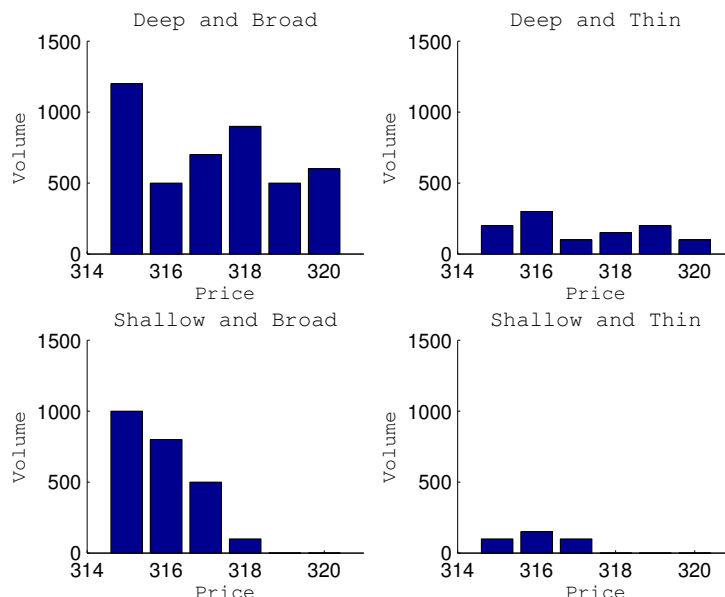


Figure 15: In this figure we have given some examples to illustrate how one would identify different depths and breadths in the sell side of an order-book. As one can see a deep order-book is characterized by having a wide range of different priced orders, while a shallow order-book would only have small range of prices attached to the orders. A broad order-book has large amounts of volume available while a thin order-book does not have much volume to offer.

market participants and risk. McNish and Wood [63] found that there is an inverse relationship between trading activity and the level of spreads. They also found intra-day bid-ask spreads on the NYSE to have what they refer to as a crude reverse J-shape. In more recent studies this has come to be known as a U-shape or an L-shape, but the results are still similar.

Our results regarding intra-day spreads are in line with previous empirical studies. Again we have filtered the data such that it is split into two groups, the first containing only those entries which occur outside of daylight savings time and the second containing only those entries which occur within daylight savings time. Daylight savings time in 2010 for the UK started on Sunday March 28 and ended on Sunday October 31, whereas daylight savings time for the USA started on Sunday March 14 and ended Sunday November 7, thus we excluded any data from 14 March 2010 to 28 March 2010 as well as

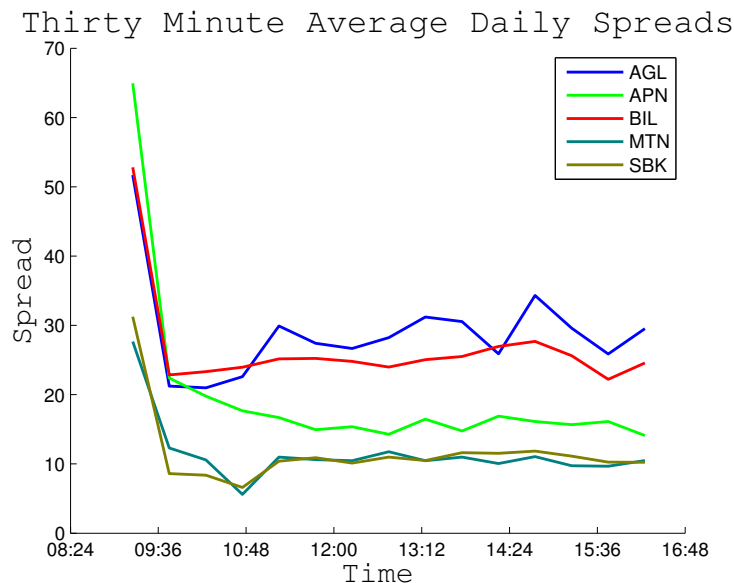


Figure 16: Thirty minute average daily spreads for the period of daylight savings time. One can see in the figure that spreads drop in the after morning auction and climb slowly when it approaches the end of the trading period, which is consistent with past studies.

31 October 2010 to 7 November 2010.

One can see in Figure (16) and Figure (17) that spreads drop after the morning auction and climb slowly towards the end of the day, which is consistent with past studies. The large drop after the morning auction is due to the fact that incoming orders can be somewhat erratic during the auction period leading to large and sometimes negative spreads. Although there is a significant inflation in the size of spreads within the daylight savings time period as seen in Figure (16), it is still clear that we have a similar spread behavior in both data sets.

So far we have discussed some of the factors which contribute towards the liquidity of a stock such as, the behavior of volume traded throughout the day, the amount of volume matched per trade and the behavior of the bid-ask spreads. We will now look at determining periods of high and low liquidity, but first we have to have a brief discussion on effective spreads.

Chordia *et al* [49] described the effective spread as being twice the absolute

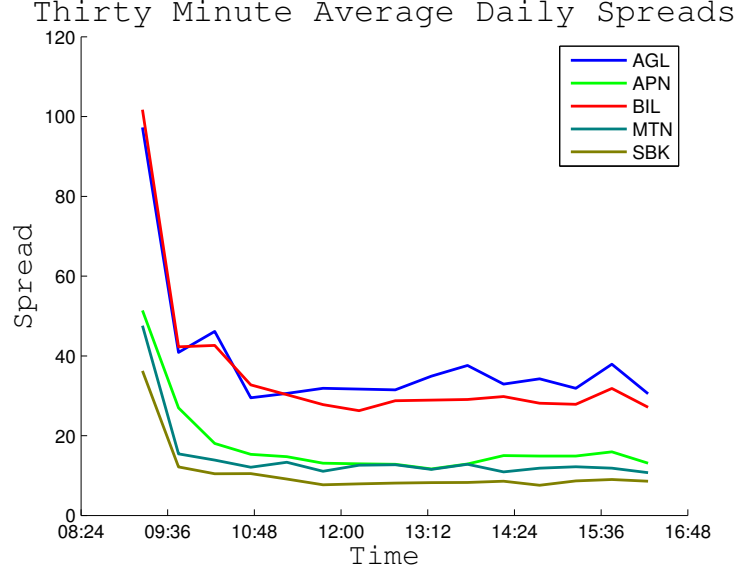


Figure 17: Thirty minute average daily spreads for the period outside of daylight savings time. One can see in the figure that spreads drop in the after morning auction and climb slowly as it reaches the close of the trading period, which is consistent with past studies.

distance between the trade price and the mid-price at the time of the trade averaged over the day. One can see how this statistic incorporates the effects of the order-book depth and is more insight into the liquidity of a stock. In our results we found the intra-day behavior of the un-averaged effective spreads to be random with no particular trend, not too dissimilar to the behavior of intra-day returns.

However this effective spread statistic can be used to determine days of high or low liquidity within a period. Since the average spread of a stock may vary throughout the year due to changes in the stock price and market behaviour, we have detrended our effective spread linearly across our 251 trading days. Using these detrended effective spread we are able to compare liquidity between trading days. Chordia *et al* [49] defined a “low liquidity” day as one with a detrended effective spread which is at least one standard deviation above the mean. A day with a detrended effective spread which is within one standard deviation of the mean is then classified as a “high liquidity” day, with no middle ground between the two classes.

To compute our daily average effective spread we took our five minute intra-day effective spreads, which are calculated by taking double the absolute difference between the trade price and the mid-price, and averaging across the day. Then to remove drift we detrend the daily average ESPR (effective spread). Once this is done we are able to calculate the mean and standard deviation and decide which days are “low liquidity” days and which are not.

A problem arises, as one can see in Figure (18), when we have outliers in the data caused by fundamental shocks. These outliers artificially inflate our average statistics and hence cause us to classify fewer days as “low liquidity” days. To compensate for this problem we have removed outliers from the data when calculating a trimmed mean and standard deviation to use for our decision bound. The detrended average daily effective spreads are plotted in Figure 18 against the decision bound as well as the trimmed decision bound.

To identify possible seasonality in stock liquidity we compared the detrended average daily effective spreads. The detrended ESPR is plotted in Figure (19) for the other four stocks in our 251 trading day period. In all five stocks one sees a group of low liquidity days found between trading day 201 and 205 which corresponds to the 24th and the 31st of December 2010.

There are no other groupings of low liquidity days which are common among the stocks which leads us to conclude that the last five trading days in the year have a significant effect on the average liquidity of a stock.

A drop in liquidity could cause an opportunity for price manipulation. Gatheral [2] proposed that if a stock has a power-law price impact function and a low liquidity, then there would be room for arbitrage. In this period of low liquidity the opportunity for arbitrage could be subdued by the general lack of trading causing a shortage in order-book depth. There would also be a risk that once the price has moved sufficiently for one to profit that there would be no matching orders to execute the closing out of one’s position. It is beyond the scope of this project to acquire the data and re-construct the order-book to investigate further.

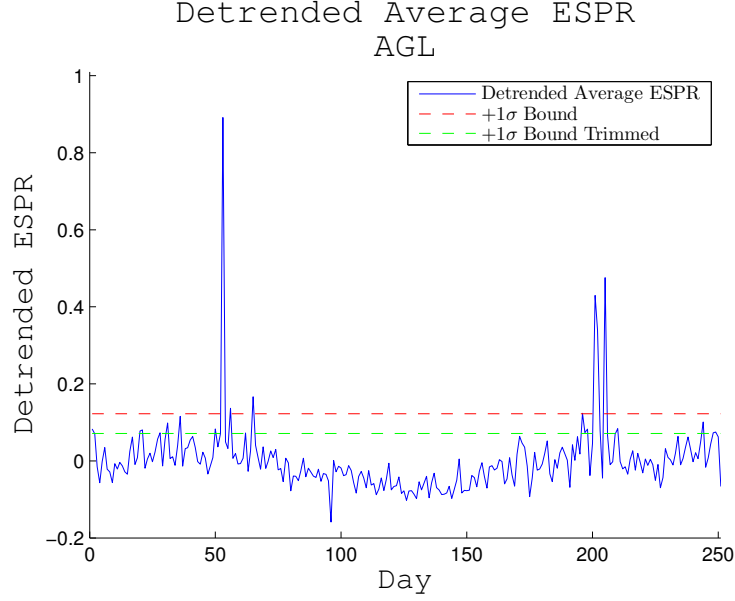


Figure 18: The detrended average daily effective spreads are plotted in order to determine periods of high and low liquidity in AGL. The red dashed line is the mean plus one standard deviation bound which is used as described by Chordia *et al* [49] to determine whether it has been a high or a low liquidity day. One can see that there is at least one outlier in this data set. To avoid over inflation of our mean and standard deviation we have removed outliers and used the trimmed statistics to get a more reasonable one standard deviation bound. Our trimmed bound is plotted as a green dashed line. In either case our low liquidity days lie above our bound, but our trimmed bound clearly has more data above it than our untrimmed bound.

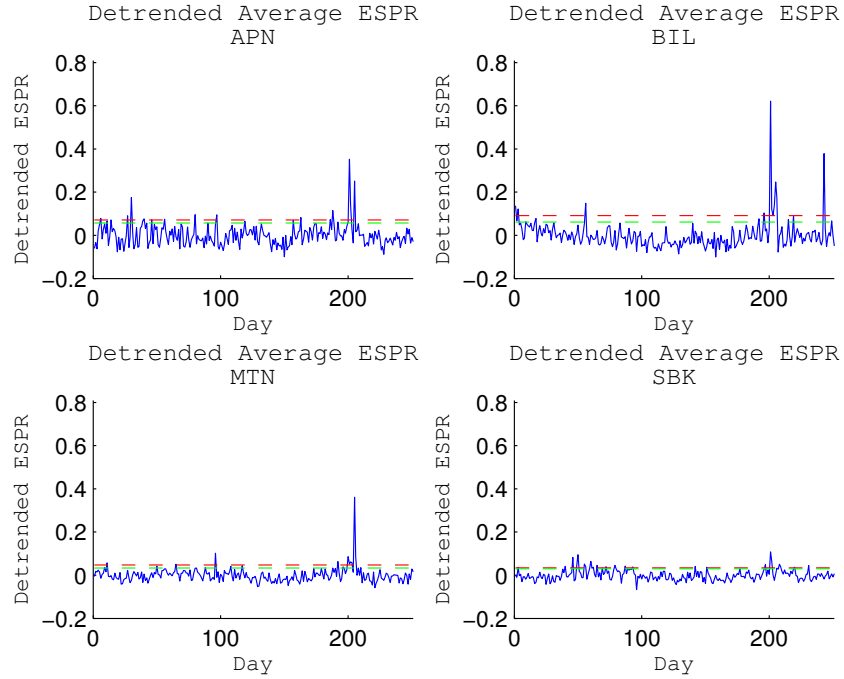


Figure 19: The detrended average daily effective spreads are plotted in order to compare periods of high and low liquidity in our five stocks. The red dashed line is our mean plus one standard deviation bound which is used as described by Chordia *et al* [49] to determine whether it has been a high or a low liquidity day. One can see that there is at least one outlier in each data set, so to avoid over inflation of our mean and standard deviation we have removed outliers and used the trimmed statistics to get a more reasonable one standard deviation bound. Our trimmed bound is plotted as a green dashed line. In either case our low liquidity days lie above our bound, but our trimmed bound clearly has more data above it than our untrimmed bound.

Ticker	<i>AGL</i>	<i>APN</i>	<i>BIL</i>	<i>MTN</i>	<i>SBK</i>
Mean	0.0484	0.0359	0.0386	0.0207	0.0185
Standard Deviation	0.0741	0.0352	0.0525	0.0265	0.0157
Trimmed Mean	0.0405	0.0332	0.0340	0.0190	0.0174
Trimmed Standard Deviation	0.0307	0.0239	0.0274	0.0146	0.0129
Low Liquidity Days	7	14	11	10	23
Low Liquidity Days (Trimmed)	22	21	18	19	26
High Liquidity Day %	91	91	93	92	89

Table 3: Detrended daily average effective spread statistics in rands for AGL, APN, BIL, MTN and SBK. As described by Chordia *et al* [49] we have used a measure of one standard deviation above the mean of the average daily effective spread to determine whether it has been a “high liquidity” or a “low liquidity” day. A problem arises when outliers in the data caused by fundamental shocks artificially inflate the mean and standard deviation. To avoid this problem we have trimmed the data of all outliers which makes a significant difference as one sees in the results above.

2.5 Basic Price Impact Analysis

The behavior of prices in response to new orders is significant factor when trying to understand the dynamics of a market. This is seen in many studies on price impact or stock price fluctuations [3, 5, 64, 8, 29, 30, 11, 12, 18, 19, 20, 21, 65]. The generalization is that if a trade causes a small change in price, we describe the market as liquid; otherwise it is illiquid.

A price impact function is a function which quantifies the effects of the volume of stock being traded on the price of that stock. There are several impact models documented in academic literature. A key consideration is whether impact is permanent or temporary are not unified. Bouchaud *et al* [1], argue that because of the auto-correlation of trade signs, impact has to be temporary. We review this model in the next chapter. Keim and Madhavan [66], and Almgren *et al* [67] argue that the price impact can be decomposed into permanent and temporary components.

Although most traders and academics believe that linearity is too simplistic an assumption, they disagree over what type of behavior price impact functions should follow. Some arguments have been made for convex and concave impact functions, but one still faces the question of what degree of

convexity or concavity the impact function takes. Any empirical findings as to the shape of the impact function, cannot be assumed to be universal, as impact can vary between different markets, across different sectors, or even between different liquidity brackets within a market.

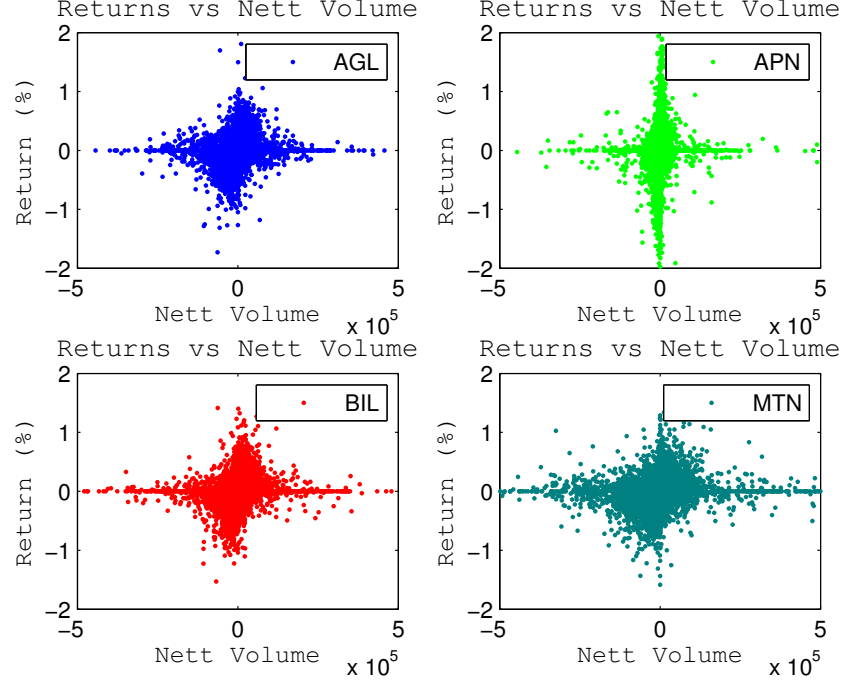


Figure 20: Five minute price movements plotted against the five minute nett volumes relating to that period for AGL, APN, BIL and MTN. Buys are represented by positive volumes and sells by negative volume.

One would expect liquidity to depend on a stock's properties, such as volume and market capitalization. Following the work done by Lillo *et al* [44], we take a look at the average effect a trade of a given size would have on the stock price.

In Figure (20) we can see that large volumes traded in a period seem to have less effect on the stock price than for smaller volumes. This is interesting as one would expect large volumes to have a large effect on the stock price and smaller volumes to have a negligible effect. One can see that this is clearly not the case. It is also important to note that the price seems to move with

as well as against the direction of trading. We note that the price seems to be affected more by the smaller volume trades. We also notice that the APN shows an almost vertical scatter, showing large returns for small volumes which is a clear indication that the price does not revert easily and the stock is therefore illiquid when compared to the others. On the other hand we see that MTN has many larger volumes which have had almost no effect on the price which implies that the stock is very. Thus, the liquid stocks namely AGL and BIL seem to have a star shape. One can also see that there seems to be a slightly hyperbolic shape to these figures which would suggest that the stock price is more likely to move up with a buy and down with a sell than the converse.

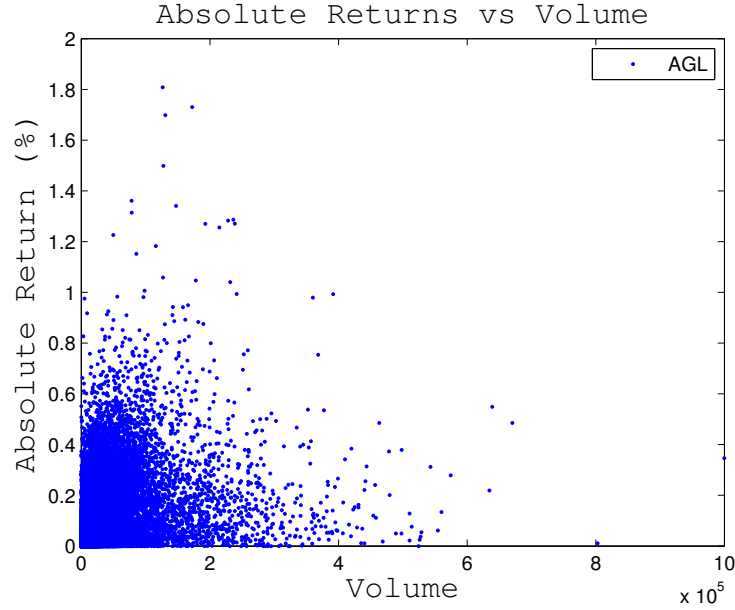


Figure 21: Five minute absolute returns plotted against the five minute volumes relating to that period for AGL. We see that for those few trades with a high return the volume of the trade is relatively low and for those trades which a high volume we see a relatively low. This is in contradiction with the common assumption that large volumes of trade have a greater effect on the stock price than that of lower volumes.

In Figure (21), where we plot the absolute returns against volume, we

find that for trades with a volume of less than 20 000 we see a vague but unquantifiable relation to the size of the return, between 0 and 0.6 percent. Any trades with larger volumes or returns seem to show an almost hyperbolic relationship between volume and size of returns.

Lillo *et al* [44] selected 1000 of the largest firms on the NYSE from 1995-98 and calculated a price impact for each firm. They found that the price impacts of all stocks showed similar power-law behavior. After scaling the price impacts using stock liquidity and market capitalization they all conformed to a single function that they called the master curve. In their study they took a buyer-initiated view. In our study we have looked at both sides of the order-book.

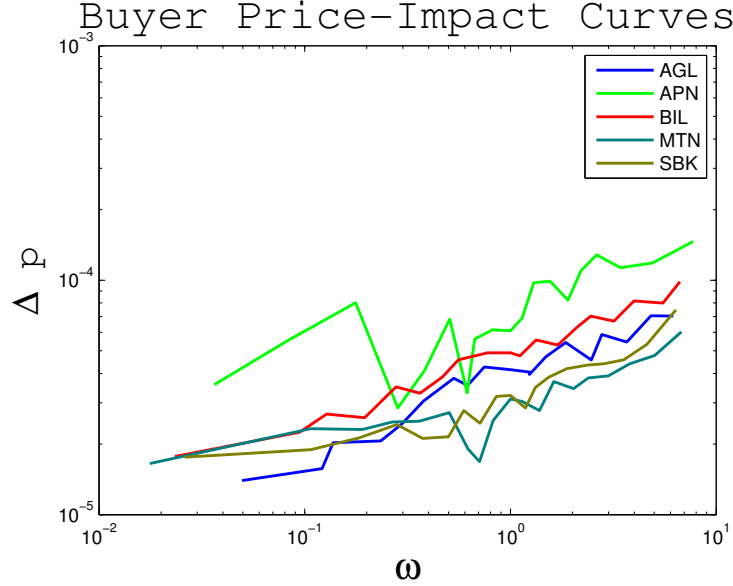


Figure 22: Buyer-initiated price impact for our five selected stocks. One sees that the impacts of stocks with similar liquidities have similar behaviors. We also note that only AGL and BIL display a power-law shaped impact similar to that found by Lillo *et al* [44].

In order to compare our price-impact results to those described above we have followed the same procedure as Lillo *et al* [44]. First we split our data into buy and sell orders then for each group we bin the transactions in terms

of order size. After sorting the data we took the logarithm of the mid-price before and after a trade has occurred to be $l(t_i)$ and $l(t_{i+1})$ respectively. Then our price shift becomes $\Delta l(t_{i+1}) = l(t_{i+1}) - l(t_i)$, for any trade that is followed by a consecutive trade we set $\Delta l(t_{i+1}) = 0$, since in this instance we are looking at a one lag impact on the mid-price consecutive trades would taint our results. To be able to compare the price-shifts we normalize the volumes by dividing through by the average volume per trade for each stock which gives us our ω .

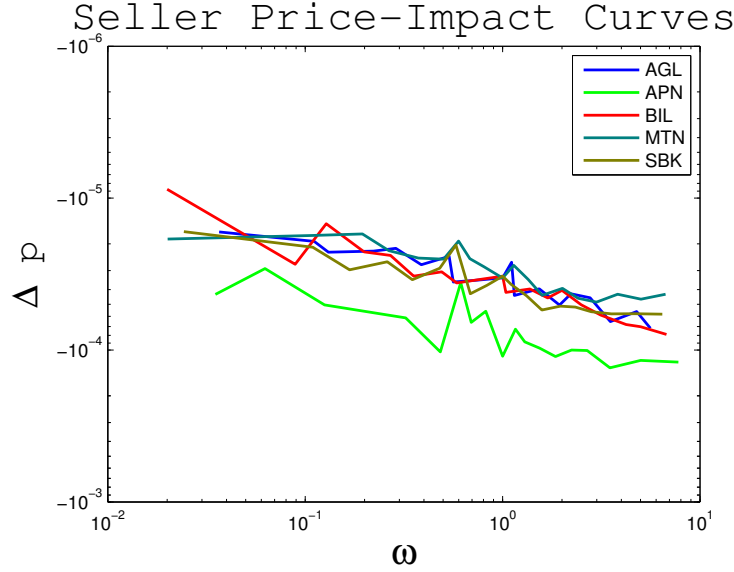


Figure 23: Seller-initiated price impact for our five selected stocks. As one would expect, the seller initiated price impact is seen to have a similar shape to a negative version of the buyer-initiated price impact. One sees that the impacts of stocks with similar liquidities have similar behaviors. We also note that only AGL and BIL display a power-law shaped impact similar to that found by Lillo *et al*[44].

In Figure (22) and (23) we have plotted the buyer-initiated and seller-initiated price impact for our five selected stocks. One sees that the impacts of our five stocks display a power-law shaped impact similar to that found by Lillo *et al* [44]. The price impact of APN is seen to be less smooth than the other four. This could be due to the fact that these are more liquid large cap stocks similar to those investigated in [44]. Since we find that our impact

functions do not all have a similar smooth shape we assume that the master curve for price impact is contingent on the stocks having a similar market cap and liquidity.

2.6 Summary and Conclusions

One problem that arose when averaging our intra-day data across days is that intra-day seasonal effects such as foreign markets opening were affected by daylight savings time. If one does not account for this, seasonal effects are lost due to over averaging, so for our study we chose to separate the periods of daylight savings and non-daylight savings time when averaging time dependent statistics.

When we look at our intra-day volume results we see that the opening of trade on foreign markets has a significant effect on the volume traded in the liquid as well as the more illiquid stocks. We see these effects more clearly when we compare the results for the period within daylight savings time with those outside of daylight savings time. From these results we realize that any time dependent intra-day computations should take this time shift into consideration in order to avoid over averaging and loss of important information. Our results on relative volumes show that there is a strong relationship between the relative volumes being traded for different stocks throughout the day. This suggests that there is some driving factor for the market as a whole when it comes to intra-day volumes.

In our five minute returns data we saw no patterns across the year and other than a few spikes, which we assume to be due to fundamental shocks or intra-day crashes, there was no information to be taken from the time series. Our average thirty minute returns were also very erratic and we only saw a few seasonal effects which were common among similar stocks. When we looked at thirty minute average absolute returns we found that, other than one spike in AGL and BIL, all five stocks displayed a similar behavior and the intra-day average absolute returns had a U-shape which seems to be common among intra-day statistics. We then found the average volatility across each day using five minute returns as well as thirty minute returns. The results showed that there seems to be a relationship between thirty minute return volatility and daily liquidity and when we use smaller intervals for our returns the volatility per period increases. We also saw

that there were seasonal effects present in all stocks when using five minute returns that weren't present when using thirty minute returns and vice versa. Thus we concluded that when one chooses only one interval size, as opposed to many, important information can be lost.

For our five selected stocks we find that their spreads behave similar to the results of previous studies by McNish and Wood [63] among others. We find as described our spreads drop rapidly coming out of the opening auction and rise slowly until the closing auction period, giving them a L-shape or what some describe as a crude J-shape. Using effective daily spreads to analyze liquidity we found that all of our stocks had a cluster of low liquidity days between the 24th and the 31st of December 2010, which is to be expected because of the holiday season and the broken trading week. Other than that there were only a few randomly scattered low liquidity days for each stock. We also found that our five selected stocks all had high liquidity days for around 90% of our 251 day period.

When looking at basic price-impact we found that there was a strong relationship between price-impact and liquidity. Out of the five stocks we chose, two were very liquid, two were liquid and one was relatively illiquid. The results supported our classification of the each stock's liquidity. We found in our data that a large trade volume does not necessarily cause a large price movement and a large price movement does not necessarily come from a large trade, in the results we found that quite often the inverse was true. This result has been noted in previous studies. When we looked at the one lag impact on the mid-price as done by Lillo *et al* [44], we found that the stocks in similar brackets of liquidity had similar shaped impact functions.

Upon closer inspection we found that markets are more complex than they appear. One of the more noticable achievements of the micro-structure literature is success in shedding a bit of light on the otherwise opaque world of stock price behavior in financial markets. The understanding of order flow and the fact that it could have a significant effect on the stock prices has some practical implications. The issues addressed in these studies should be relevant for exchange officials, trading system developers, regulators, and traders. Although certain anomalies and statistical rule are common in most markets, a one size fits all approach to strategic trading, regulation and policy making needs to be avoided.

3 Trade Sign Behavior

3.1 Auto-Correlation of Trade Signs

In our previous discussions we can see that the random-walk nature and the reaction of liquidity providers to the trading done by liquidity takers suggests that the idea of a "zero intelligence" trader market could help us understand the financial markets from a quantitative view. This model is however qualitatively incorrect: although the stock prices are weakly correlated at best, the correlations between the signs of trades are strong and fail to die out. This stems from the splitting of orders and spreading them over time to withhold demand information, or the implementation of optimal trading strategy that we discussed in section 1.4.

In our study we define our trade sign at time n $\epsilon_n = 1$ if the stock has been bought and $\epsilon_n = -1$ if the stock has been sold.

A basic function for the correlation of the signs over a lag of ℓ trades, starting with the n -th trade, can be given by:

$$\mathcal{C}_0(\ell) = E(\epsilon_n \epsilon_{n+\ell}), \quad (3)$$

where ϵ_n is the sign of the n -th trade.

In [1] it was shown that the auto-correlation in the sign of trades didn't die out for the stock studied, even after 1000 lags. Figure (24) shows that our analysis yields the same (truncated) power-law behavior for all 46 stocks.

As one can see in Figure (24) our results were similar to those found in [1], but in this study we only see results for a single stock. In order to compare our results as a market we have repeated these computations for 27 stocks on the BM&FBOVESPA exchange. The results are shown in Figure (25). We found that for these two very different markets that the results come out to be very similar to each other. We also found that all 73 stocks have the same power-law behavior that fails to die out that was found in [1].

Following [1], where the auto-correlations can be reasonably considered to have the form:

$$\mathcal{C}_0(\ell) \simeq \frac{\mathcal{C}_0}{\ell^\gamma}, \quad (4)$$

using the techniques described by Clauset, Shalizi and Newman [68] we find $\mathcal{C}_0 \approx 0.26 \pm 0.07^4$ and $\gamma \approx 0.32 \pm 0.06$. In [1], it was found that $\mathcal{C}_0 \approx 0.2$ but $\gamma \approx \frac{1}{5}$, and $\gamma \approx \frac{2}{3}$, for FT and Total respectively. The similar auto-correlation result seems to hold well for all 46 stocks; which could suggest that γ is market dependent. This is not surprising as the stocks we studied were all part of the basket of underlyings for the most liquid index future traded on SAFEX⁵: the ALSI TOP 40. Our results show that we consistently find $\gamma < 1$.

Let us now look at the two related correlation functions, described by Bouchaud *et al* [1], that should appear naturally in the stock price function, namely:

$$\mathcal{C}_1(\ell) = E(\epsilon_{n+\ell} \epsilon_n \ln V_n) \quad (5)$$

and

$$\mathcal{C}_2(\ell) = E(\epsilon_{n+\ell} \ln V_{n+\ell} \epsilon_n \ln V_n). \quad (6)$$

These “mixed” correlation functions can be found, empirically, to be proportional to $\mathcal{C}_0(\ell)$:

$$\mathcal{C}_1(\ell) \approx E(\ln V) \mathcal{C}_0(\ell) \quad \mathcal{C}_2(\ell) \approx E(\ln V)^2 \mathcal{C}_0(\ell). \quad (7)$$

There are systematic deviations, which indicate that long range correlations are affected more significantly by smaller volumes than larger volumes.

⁴The 1- σ error is from averaging over all 46 stocks measured auto-correlations.

⁵South African Futures Exchange

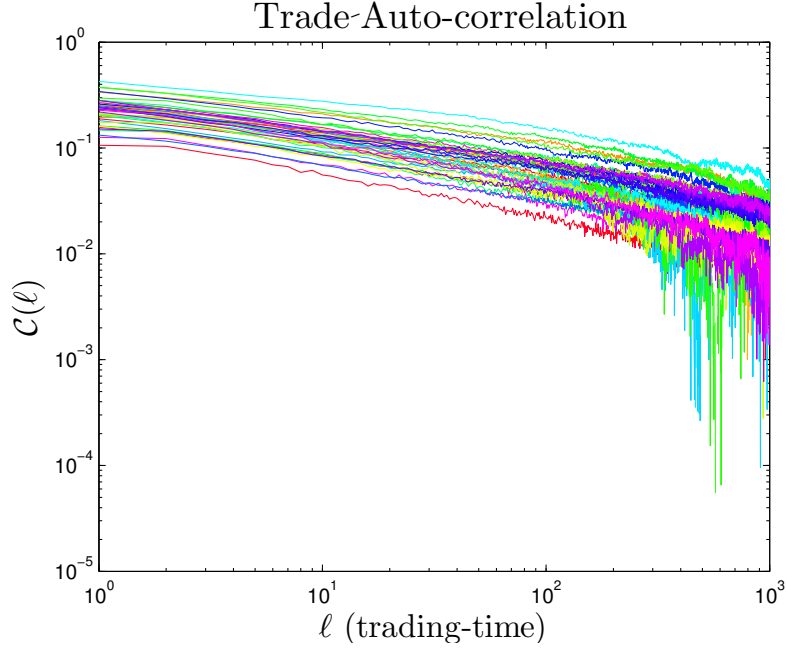


Figure 24: In this figure we plot the auto-correlation of the trade signs of the 46 stocks on the JSE. We note that for all 46 stocks the auto-correlation fails to die out even after a large number of trades. All of our stocks display a power-law behavior in their trade sign auto-correlations which is the same result found by Bouchaud *et al* [1], $\mathcal{C}_0(\ell) \simeq \frac{\mathcal{C}_0}{\ell^\gamma}$. We also found that $\mathcal{C}_0 \approx 0.26 \pm 0.07^3$ and $\gamma \approx 0.32 \pm 0.06$ for the JSE stocks (where [1] finds $\mathcal{C}_0 \approx 0.2$ but $\gamma \approx \frac{1}{5}$, and $\gamma \approx \frac{2}{3}$, for FT and Total respectively). The similar auto-correlation result seems to hold well for all 46 stocks; which could suggest that γ is market dependent.

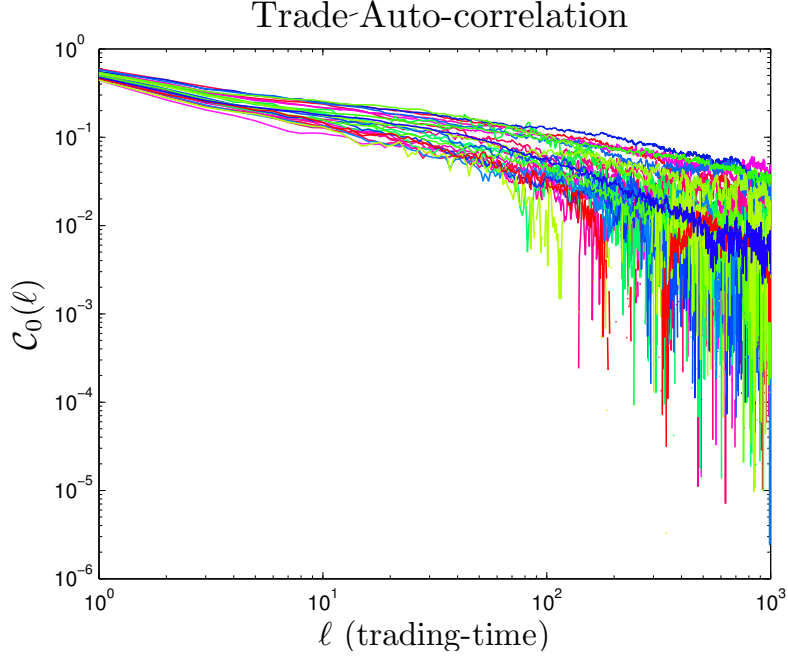


Figure 25: In this figure we plot the auto-correlation of the trade signs of the 27 stocks for the BM&FBOVESPA data. We note that for all 27stocks the auto-correlation fails to die out even after a large number of trades. One also sees that all of the display a power-law behavior in their trade sign auto-correlations which is the same result found in our JSE data and in [1], $C_0(\ell) \simeq \frac{C_0}{\ell^\gamma}$. Two of the auto-correlations seem to behave differently, but the others all have the same shape lie near each other which means that they would share a common γ . One thing that is different from our JSE results is the fact that these auto-correlations are significantly smoother.

3.2 Trade Sign Clustering

Taking into account the correlation of the trade signs over long lags and the idea of a optimal trading strategy and splitting up of market orders to withhold information, we went on to take a closer look at the behavior of the trade signs. On closer inspection one sees clustering of the same signs as is seen in Figure (26), this can be linked to this explanation of the correlation of the signs.

To investigate the clustering, we condensed the trade signs into runs of either positive or negative i.e. if there only are 3 positive signs followed by 6 negative signs our “runs” vector would read (3 6). After capturing the “runs” information we found that the “runs” data had a behavior similar to that of a geometrically distributed random variable. Upon closer inspection one finds that the distribution of the positive “runs” differs from that of the negative “runs”.

When using our runs statistics from our data to simulate trade signs we found that the auto-correlation of the simulated signs behaved similarly to the signs in the data for lags of up to 10, but then they would die off exponentially. This contradicts our findings in the data that auto-correlation of trade signs fails to die out even after large lags. This would suggest that there are other factors at play and that the trade sign clustering is not the sole cause of this behavior in our trade sign data.

<i>Statistics</i>	<i>Negative Mean</i>	<i>TradeSigns Volatility</i>	<i>Positive Mean</i>	<i>TradeSigns Volatility</i>
<i>AGL</i>	2.34	2.26	2.38	2.36
<i>APN</i>	2.55	3.14	2.53	3.21
<i>BIL</i>	2.35	2.33	2.41	2.43
<i>MTN</i>	2.25	2.21	2.26	2.27
<i>SBK</i>	2.32	2.41	2.33	2.58

Table 4: In this table we have some statistics regarding the trade signs of AGL, APN, BIL, MTN, SBK. One can see that the positive and negative runs for each stock have slightly different distributions. This suggests an overall dominance of one sign in each of the stocks. The dominance of one sign in each stock gives us a reasonable explanation for why trade signs are strongly correlated over a relatively long period.



Figure 26: In this figure we plot 100 trade signs of the AGL, APN and SBK stocks taken from an arbitrary day in our range of data. Note that for all 3 stocks there is a clear clustering of trade signs. In this short period one can see that traders favored buying AGL and APN while they were in favor of selling SBK. We also found, using the same method as Bouchaud *et al* [1], some of our trade signs came to be zero. This does not give us any information as to whether a buy or a sell transaction has occurred, taking a market order perspective. It does suggest that the transaction was executed at midprice, for example, as if it had taken place in a dark pool. This may suggest the presence of "book-over" type trades due to off-market agreement between participants but executed through the central order-book as normal continuous trading period transactions. In this study we disregard trades such as demarcated book-over trades as we are investigating trade statistics and price-impacts in a lit-market.

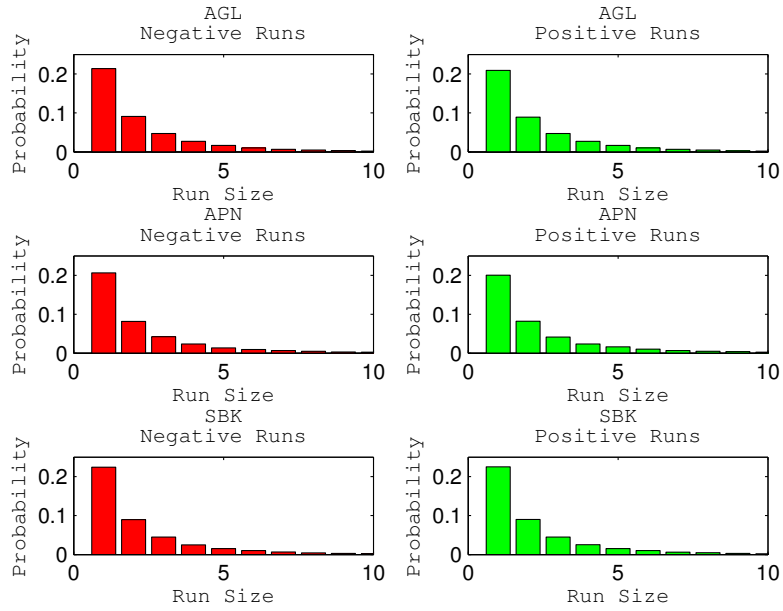


Figure 27: In this figure we have plotted the distributions for the positive as well as the negative signs of the AGL, APN and SBK stocks across the year. Note that all 3 stocks have a similar distribution when it comes to their trade sign clustering. Since these distributions are discrete and have similar means and volatilities, we have assumed that they follow a geometric distribution. One can see in Table (4) that the positive and negative runs for each stock have slightly different distributions. This suggests that overall one side is dominated by the other.

For the period under investigation the negative “runs” dominated the positive “runs” for 40 of the 46 stocks across the year. We found that if one looks at smaller periods of days or weeks, then positive “runs” are sometimes dominant. This behavior is indicative of a market consensus or a general reaction to stock price movements by participants. This may also have an effect on the correlation of the trade signs.

In Figure (28) we have plotted the cumulative trade signs for AGL, APN and SBK across the year so that this negative sign dominance and the smaller periods of positive sign dominance can be seen.

This long-term dominant behavior gives us a reasonable explanation for why the trade sign correlations are strong and fail to die out, even after a thousand trades.

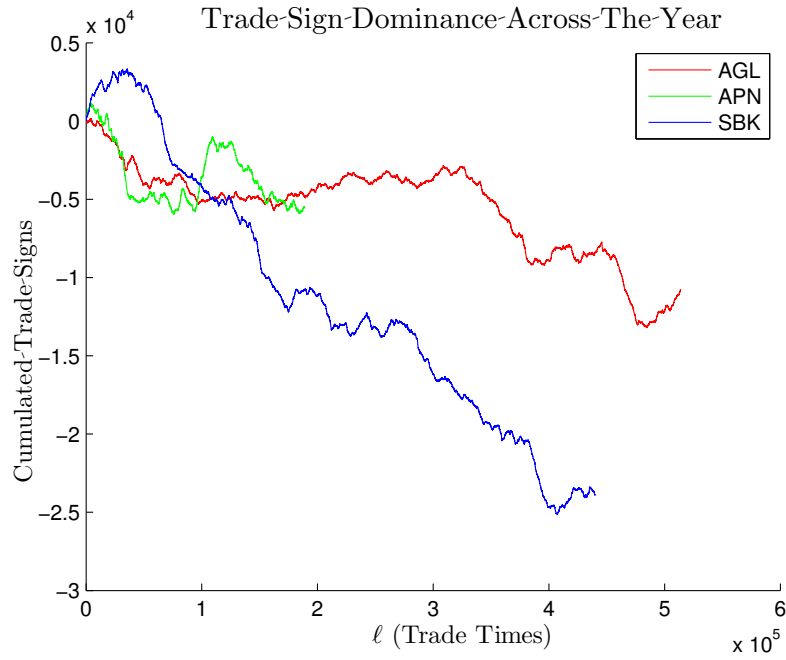


Figure 28: In this figure we plot the cumulative trade signs of the AGL, APN and SBK stocks across the year. This figure illustrates that there is a dominance of negative trade signs over the year in question. Apart from the overall dominance one sees a dominance of either positive or negative trade signs when one looks at smaller periods. The relatively long term auto-correlation of trade signs could be explained by the fact that the market moved in a negative or positive direction over small periods and thus the signs in these periods would be strongly correlated.

4 Cost of Trading Models

In this section, we review the concepts of cost of trading and examine the models for market impact suggested by Gatheral in 2009 [2] and Gatheral, Schied and Slynko in 2010 [69]. In these studies we see stock price, S_t at time t in the form:

$$S_t = S_0 + \int_0^t f(\dot{x}_s)G(t-s)ds + \int_0^t \sigma dZ_s \quad (8)$$

where $\dot{x}_s$ is our rate of trading in at time $s < t$, $f(\dot{x}_s)$ is the impact of trading on the stock price at time s and $G(t-s)$ is what they describe as a decay factor.

Z_s is a Brownian Motion, thus S_t follows an arithmetic random walk with a drift component which is dependent on the cumulative impacts of prior trades.

Gatheral *et al* [2, 69] refers to $f(\cdot)$ as the instantaneous market impact function and to $\mathcal{G}(\cdot)$ as the decay kernel. They show that the continuous time process Eq(8) can be viewed as a limit of the discrete time process:

$$S_t = \sum_{i < t} f(\delta x_i)G(t-i) + S_0 + Noise, \quad (9)$$

where $\delta x_i = \dot{x}_i \delta t$ is the volume traded in a small time interval δt , $f(\cdot)$ is the market impact function and S_0 is the initial price. $\delta x_i > 0$ would represent buying of stock and $\delta x_i < 0$ the sale of stock. The discrete representation is in line with the ideas of Bouchaud *et al* [1].

If the number of shares outstanding at time t is given by x_t . Gatheral *et al* [2, 69] find the expected cost $C[\Pi]$ associated with a given trading strategy $\Pi = X_t$ to be:

$$\begin{aligned} C[\Pi] &= E \left[\int_0^T \dot{x}_t (S_t - S_0) dt \right] \\ &= \int_0^T \dot{x}_t dt \int_0^t f(\dot{x}_s)G(t-s)ds. \end{aligned} \quad (10)$$

The $dx_t = \dot{x}_t dt$ shares traded at time t will then be traded at an expected price of

$$S_t = S_0 + \int_0^t f(\dot{x}_s)G(t-s)ds,$$

which will reflect the residual impacts of previous trades. The cost of trading is then given by the expected short-fall.

4.1 Special Cases

Below are some of the cost of trading functions described by Gatheral *et al* [2] for some established impact functions and how you would arrive at these functions using the cost of trading function he described.

Bouchaud *et al.* [1]

Bouchaud, Gefen, Potters, and Wyart (2004) [1] assume that we have $f(v) \propto \log(v)$ and

$$G(t-s) \propto \frac{l_0}{(l_0 + t-s)^\gamma},$$

with $\gamma \approx (1-\alpha)/2$ where α is the exponent for the trade sign autocorrelation power law (also see [27]). In Bouchaud *et al*'s model [1], market impact decays as a power law and the price impact becomes concave in the trading rate. The cost of trading becomes:

$$\begin{aligned} C[\Pi] &= \int_0^T \dot{x}_t dt \int_0^t f(\dot{x}_s)G(t-s)ds \\ &\propto \int_0^T \dot{x}_t dt \int_0^t \frac{\log(\dot{x}_s)}{(l_0 + t-s)^\gamma} ds. \end{aligned}$$

Almgren *et al.* [67]

The temporary component of the model of Almgren, Thum, Hauptmann, and Li (2005) [67] corresponds to setting $G(t-s) = \delta(t-s)$ and $f(v) = \eta\sigma v^\beta$ with $\beta = 0.6$. Here, σ is volatility and $v_t = \dot{x}_t/V$ is a dimensionless measure of the rate of trading, where V is the market volume per unit time (average

intra-day volume, say). In Almgren *et al*'s [67] model, the price impact dies out immediately. Any trading will only affect the price at the time of executions; other executions will not be affected. The cost of trading becomes:

$$\begin{aligned} C[\Pi] &= \int_0^T \dot{x}_t dt \int_0^t f(\dot{x}_s) G(t-s) ds \\ &= \eta \sigma \int_0^T \dot{x}_t^{1+\beta} dt. \end{aligned}$$

Obizhaeva and Wang [22]

In Obizhaeva and Wang (2005) [22], the form $G(\tau) = e^{-\rho\tau}$ and $f(v) \propto v$ was considered. In Obizhaeva and Wang's [22] model, the price impact decays exponentially and immediately price impact becomes linear in the trading rate. The cost of trading becomes:

$$\begin{aligned} C[\Pi] &= \int_0^T \dot{x}_t dt \int_0^t f(\dot{x}_s) G(t-s) ds \\ &\propto \int_0^T \dot{x}_t dt \int_0^t \dot{x}_s \exp\{-\rho(t-s)\} ds. \end{aligned}$$

Alfonsi, Schied, and Schulz (2007) [70] also assumed that the decay of price impact is exponential, but they assume that the price impact function is nonlinear.

As the shape and behavior of the price impact function changes the cost of a trade and consequently the optimal trading strategy will change. Although the strategy will change, the idea of order-book convergence, discussed in Section 1.3, where after a trade the order-book will fill until it reaches a new steady state, suggested by Obizhaeva and Wang [22] ensures that the strategy involving “eating off” incoming orders is relevant regardless of the shape or behavior of the price impact function. Thus the shape of the optimal trading strategy is similar for different impact functions, only the size of the trades would vary.

5 Price Impact Statistics

5.1 Introduction

Despite advances in the field, no standard approach for modeling price impact has emerged. Since the early contributions of Hasbrouck [50] and Bertsimas and Lo [71], price impact models have been used in the development of optimal trading strategies of large orders [22], its relation to other studied quantities such as liquidity [72] and the existence of arbitrage at high frequency [2, 73]. More realistic models of the price-discovery process consistent with economic theory have been developed [7].

For example, price impact is proxied by price change as a convex function of volume in [44] and by price change as a linear function of order flow imbalance in [74]. The measurement of autocorrelations [17, 75, 76, 77], in the signs of trades provides a robust, albeit blunt measure to detect of existence of some stylized fact. With the number of trades per day for JSE-listed stock in our investigation varying between 600 and 10 000, a more refined, yet equally easily measurable quantity, is the response function, $\mathcal{R}(\ell) = \mathbb{E}[(p_{n+\ell} - p_n)\epsilon_n]$, proposed in [1] which we will discuss later in this section.

Although many comments have been made about the relevance of Bouchaud *et al*'s model in current circumstances, with high frequency trading playing a large role in developed markets, we believe that it is still relevant in the South-African market.

For the analysis done in this paper we analyzed Thompson-Reuters Tick History data for 46 stocks⁶ listed on the JSE from 10 March 2010 to 9 March 2011. There are two possible methods involved in this analysis. One is to focus on each day's continuous trading period individually, with the assumption that the impact of trading is an intra-day effect we pursue this approach.

Our original data was made up of trade price, best bid, best ask and trade time for 251 full days of trading. The average number of trades, across all

⁶The tickers for these stocks are: ABL, AGL, AMS, ANG, APN, ARI, ASA, BIL, BTI, BVT, CFR, CSO, DSY, EXX, FSR, GFI, GRT, HAR, IMP, INP, IPL, KIO, LON, MNP, MSM, MTN, NED, NPN, NTC, OML, PPC, RDF, REI, REM, RMH, SAB, SAP, SBK, SHF, SHP, SLM, SOL, TBS, TRU, VOD and WHL. The associated company information can be found on the JSE website www.jse.co.za

46 stocks, was 802 trades per day which is equivalent to an average spacing 34.71 seconds between trades. Since we are focusing on intra-day continuous trading we removed any data from before 09h01 and after 16h45, to remove any effects of the opening and closing auctions.

We then used this data to calculate: ϵ_n , the sign of a trade executed at trade time n where $\epsilon_n = +1$ for a buy order and $\epsilon_n = -1$ for a sell order. We determine whether a trade is a buy or a sell by checking the trade price against the corresponding mid price, if the stock traded higher than the mid price we treat the trade as a buy and lower than the mid price we treat it as a sell. The lagged price difference, $p_{n+\ell} - p_n$, where p_n is the price before the n th trade and ℓ is the trade lag between the two.

In using the continuous trading period between opening and closing auction for our study, we faced the issue of a decreasing number of data points in each day as we increase our trade lag. Since we are not crossing days, at higher lags the less active days for a stock would be excluded from the data set. This leads to a loss of generality of our empirical study. In order to avoid deviations in our statistics due to lack of data points, we assumed that the median of trades per day, for each stock, would be a reasonable stopping point when calculating our statistics.

Since Bouchaud *et al* [1] have focused on one stock for the purposes of their study, below we have computed the diffusion and response functions for our 46 JSE stocks as well as 27 BM&FBOVESPA stocks.

We found that a problem of deciding on an appropriate cut-off point for the lags occurred when computing related statistics. When one looks at the amount of trading that occurs during each day, one sees large differences not only in the different stocks, but in different trading days of any particular stock. If one had to take the maximum trades occurring in a day for each stock as a cut-off point, the statistics one would obtain for the higher lags would be meaningless as they would be computed using a very small pool of data. We use the median of trades per day as a cut-off point, or ℓ_{max} , since in this case no less than half of the data would be used in computation of said statistics.

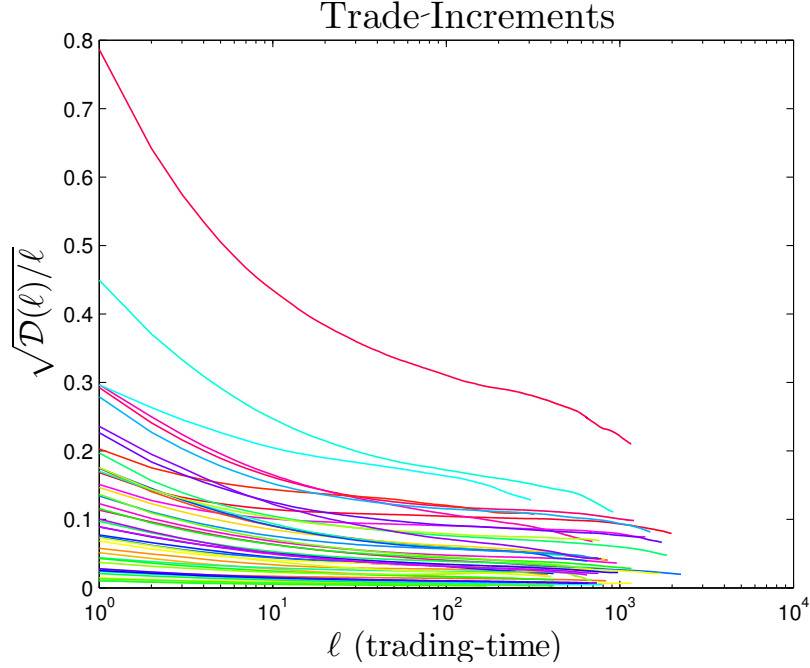


Figure 29: We have plotted $\sqrt{\mathcal{D}(\ell)}/\ell$ as a function of ℓ for all 46 stocks. One sees that the bigger tick stocks seem to have a sub-diffusive behavior while the smaller tick stocks have a relatively diffusive behavior.

5.2 Diffusion Function

If we take the price p_n as being the price of a stock before the n th trade and ϵ_n to be the sign attached to said trade. The diffusion function is given by:

$$\mathcal{D}(\ell) = \mathbb{E}[(p_{n+\ell} - p_n)^2]. \quad (11)$$

If we do not see any relationship between consecutive changes in price, $\mathcal{D}(\ell)$ has a purely diffusive behavior:

$$\mathcal{D}(\ell) = D\ell,$$

where D is a constant.

Figure (29) shows that for stocks with very small tick sizes $\sqrt{\mathcal{D}(\ell)}/\ell$ remains relatively constant even after large lags. This suggests diffusive

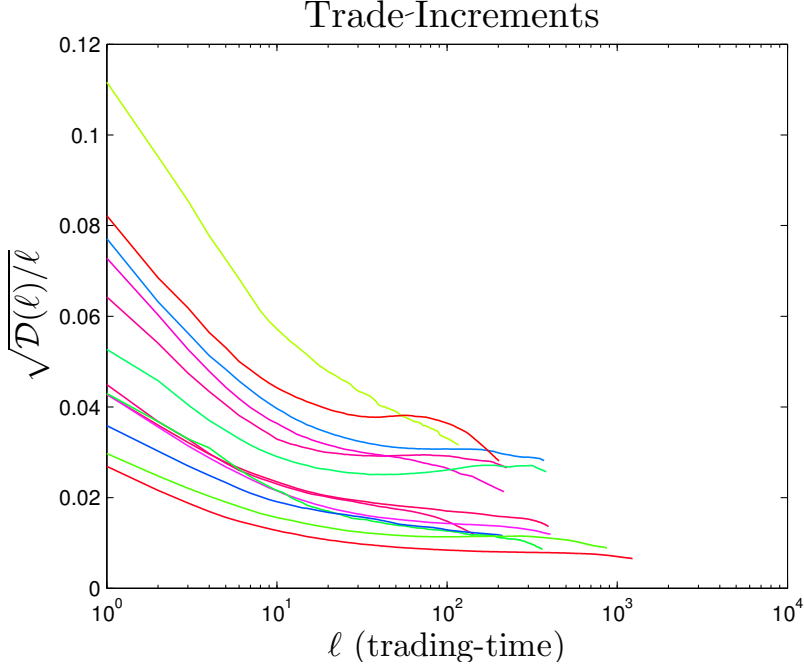


Figure 30: Here we plot $\sqrt{\mathcal{D}(\ell)}/\ell$ as a function of ℓ for all the bigger tick size stocks in our BM&FBOVESPA data. One sees that the bigger tick stocks seem to have a sub-diffusive behavior. This behavior suggests that mean reversion is at play in all of these stocks.

behavior of stock prices, even at a granular level, for the more liquid stocks. It is also noted that the $\mathcal{D}(\ell)/\ell$ for stocks with large tick sizes decays slowly. This sub-diffusive behavior corresponds to an anti-persistent effect.

To investigate this apparent difference between large tick and small tick stocks we studied these two groups separately and found that the small tick stocks also have a sub-diffusive behavior. In Figure (30) and (31) we have plotted the trade increments of the BM&FBOVESPA stocks separately for the big tick and small tick stocks to show that both groups have similar behavior on different scales.

We see that both groups of BM&FBOVESPA stocks have the same sub-diffusive behavior which suggests all the BM&FBOVESPA stocks in our study seem to be mean-reversive.

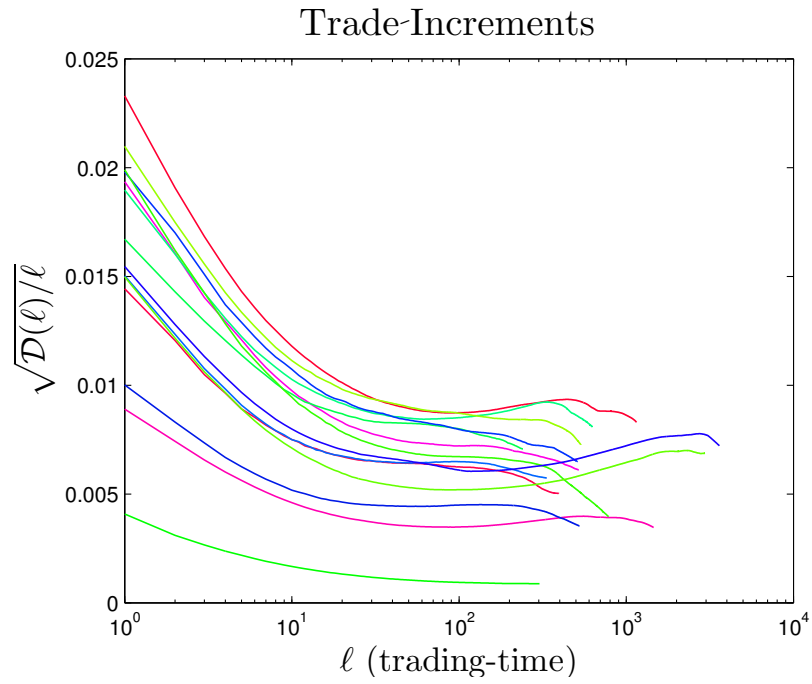


Figure 31: In this figure we have plotted $\sqrt{\mathcal{D}(\ell)}/\ell$ as a function of ℓ for all the smaller tick size stocks in our BM&FBOVESPA data. One sees that the small tick stocks seem to have a sub-diffusive behavior. This behavior suggests that mean reversion is at play in all of these stocks.

As one can see in Figure (29) when the stocks are not separated by their tick size, the data is misinterpreted. Thus we realize that the tick size of a stock is an important factor when doing a large scale study of stock markets.

5.3 Empirical Response Function and Market Impact

We follow Bouchaud *et al* [1] to better understand the impact that trading has on the change in stock price using the response function, $\mathcal{R}$:

$$\mathcal{R}(\ell) = E[(p_{n+\ell} - p_n)\epsilon_n], \quad (12)$$

where ϵ_n is the trade sign described the previous section and p_n is price before the trade executed at time n . The quantity $\mathcal{R}$ shows us the average price movement after ℓ ticks relative to an order at time n . This is an aggregate function and is not used to define the effect of a single trade, it also does not take the volume of the trade into account. Below we have the response functions of the 46 stocks:

In Figure (32) an almost constant behavior seems to appear in the “small tick” stocks, but when we looked at each stock separately we found that each empirical response function had its own shape which was not constant even in the short term. Similar to the results of Bouchaud *et al* [1], we find that our empirical response functions are positive. Since we are taking a market-order side view we assume that a significant portion of this response could be a ripple effect of the order-book moving in the direction of the trade. Like Bouchaud *et al* [1], we also found that the empirical response functions seem to misbehave after around 100 lags which would suggest that the price-impacts or our stocks are temporary.

In Figure (33) and (34) we have again plotted the response functions of the big tick and small tick BM&FBOVESPA stocks separately. One sees, for all the stocks, that the empirical response function is non-constant. Again we see a positive response in the BM&FBOVESPA stocks. We also see that the empirical responses of the small tick stocks tend to misbehave after around 100 lags, whereas the big tick stocks could misbehave sooner than 100 lags. One can see in Figure (32) that when the results of an empirical study are not viewed in the right scale, they can be misinterpreted.

A more detailed function can be defined by taking into account the volume V of the n -th trade:

$$\mathcal{R}(\ell, V) = \text{E}[(p_{n+\ell} - p_n)\epsilon_n / V_n = V]. \quad (13)$$

Prior to the work of Bouchuad *et al* [1] it was proposed that the dependence of the response function on volume should be logarithmic. A time dependent form of $\mathcal{R}(\ell, V)$ has not been investigated as thoroughly as the average response function. Results published in [31] reported that the conditional response function $\mathcal{R}(\ell, V)$ can be factorized:

$$\mathcal{R}(\ell, V) \approx \mathcal{R}(\ell)f(V),$$

where

$$f(V) \propto \ln V. \quad (14)$$

The average response function $\mathcal{R}(\ell)$ shows us the effect of the sign of a trade to the average price change. The variance of this effect is large and increases over time. A way to see this is to study the impact variables given by: $u_\ell = (p_{n+\ell} - p_n)\epsilon_n$. $\mathcal{R}(\ell)$ is the expected value of u_ℓ and $\mathcal{D}(\ell)$ is the expected value of u_ℓ^2 . We found that $\mathcal{R}(\ell)$ is fairly steady and $\mathcal{D}(\ell)$ grows over time. Thus we see that the effect and visibility of the impact of one trade is lost after a few lags.

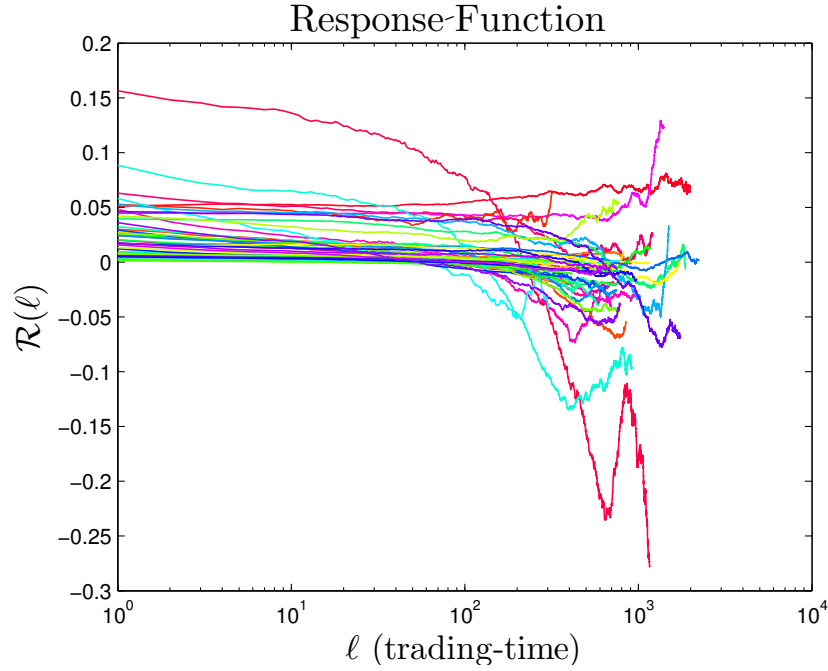


Figure 32: Here we plot $\mathcal{R}(\ell)$ as a function of ℓ for all 46 stocks. Similar to the results of Bouchaud *et al* [1], we find that our empirical response functions are positive. We also see that our response functions seem to misbehave after 100 lags for most of these stocks, this was also seen in the response function of the FT stock studied by Bouchaud *et al* [1]. When we plot the big tick and small tick stocks together, their response functions seem to be constant up to 100 lags. When we looked at each stock separately we found that the average responses were anything but constant. To illustrate this we have again plotted the response functions of the big tick and small tick BM&FBOVESPA stocks separately in Figure (33) and (34).

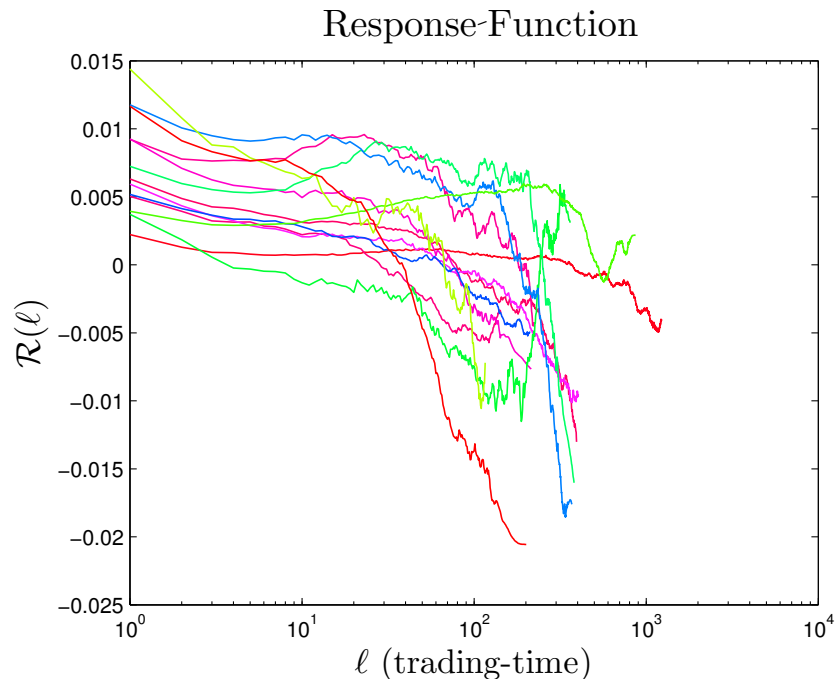


Figure 33: In this figure we have plotted $\mathcal{R}(\ell)$ as a function of ℓ for all the larger tick stocks in our BM&FBOVESPA data. For the big tick BM&FBOVESPA stocks we see, that their empirical response functions are positive and also misbehave after 100 lags.

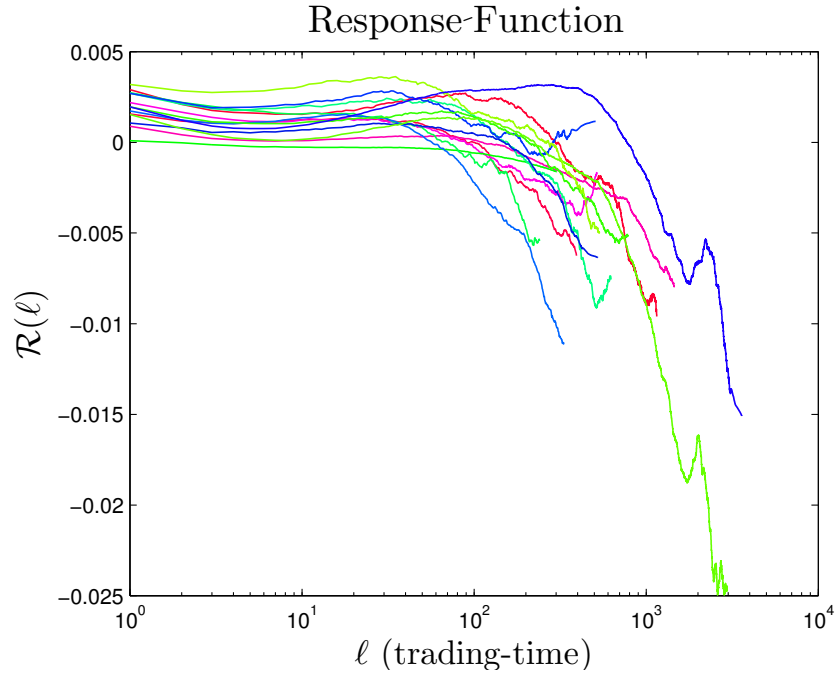


Figure 34: In this figure we have plotted $\mathcal{R}(\ell)$ as a function of ℓ for all the smaller tick size stocks in our BM&FBOVESPA data. For the smaller tick BM&FBOVESPA stocks we see, that the empirical response functions are positive and also misbehave after 100 lags.

5.4 Distribution of Impact Variables

Bouchaud *et al* analyzed the distribution of u_ℓ . When they fixed the lag at $\ell = 128$ for example they observed a thicker positive tail and when shifted 0.02 units to the left, the distribution became almost symmetric. They argued that if one had to take an EMH view point this shift could be seen as the proportion of properly informed agents in the market, whereas the “ZIA” model would suggest that this shift is caused by the bid-ask spread(cost of trading), they noted that the typical bid-ask spread on the stock was close to 0.02 euros.

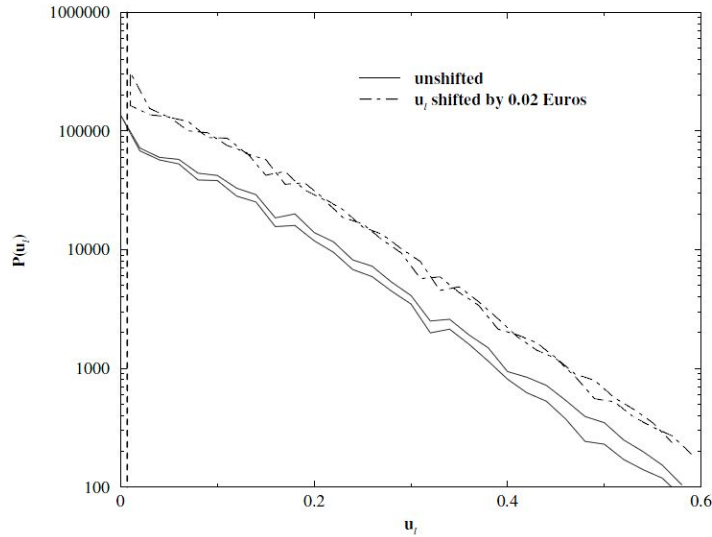


Figure 35: The distribution of the u_l at $l = 128$ as found by Bouchaud *et al* [1] for the FT stock. They found that the distribution of their impact variables computed for the FT stock had a thicker positive tail. One would expect this as the average response, which is derived from these impact variables, was found to be positive in their study.

Figure (35) depicts the distribution of the impact variable of the French-Telecom stock in [1]. In this study it was found that the distribution of the impact variables was slightly positively skewed. One would expect this as the empirical response was found to be positive in this study and this empirical response is merely the mean of the impact variables at different lags.

In our study we found that our u_ℓ variables had a positive mean as well as being positively skewed as one sees in Table (5) where we have a few statistics of the u_ℓ variables for three selected lag points for one of our stocks. These statistics are consistent with our results on the average response.

	u_{10}	u_{100}	u_{1000}
Mean	0.0526	0.0545	0.0652
Variance	0.1288	1.0903	8.7102
Skewness	0.1397	0.0500	0.0164
Kurtosis	5.6086	4.2369	3.2284

Table 5: AGL u_ℓ statistics for lags of 10, 100 and 1000.

In Table (5) we see that as one increases the lag, the variance of the impact variables increases greatly and the skewness and kurtosis tend towards zero, while the mean remains relatively constant. From this we conclude that impacts over longer lags tend to become noise and thus the impacts seem to be temporary. We see also that our mean and skewness are both positive in all three cases which is consistent with [1], and the positive response functions which we found for the JSE and BM&FBOVESPA data.

In Figure (36) we see the same positive skewness in the distribution of u_{100} for AGL. Like Bouchaud *et al* we also see that slight shift, 0.1 units to the left, in the negative tail would make the distribution symmetric.

We see the same skewness for BIL, MTN and SBK, with negative tail shifts of 0.1, 0.03 and 0.03 units respectively, required for symmetry. However we see the opposite behavior in the u_{100} 's for APN, which are negatively skewed and requires the positive tail to shift by 0.02 units to achieve symmetry.

This behavior was expected as we saw the empirical response function of APN become negative before 100 lags. As it is one of the less liquid stocks in the study one would expect its response function to misbehave earlier than others as it has less trades taking place, as well as lower volumes being traded each day than most of the other stocks.

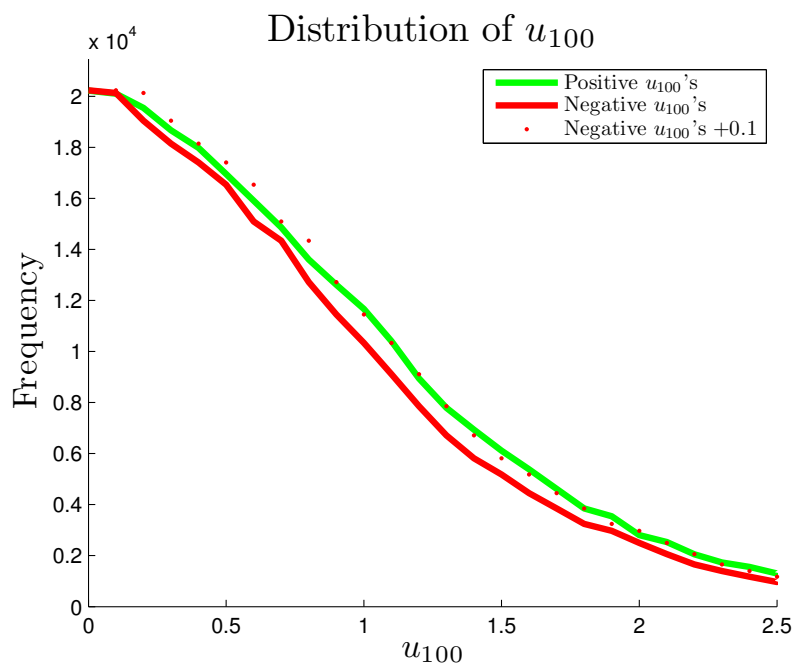


Figure 36: Plot of the distribution of the u_ℓ 's at a lag of 100 for AGL. We have plotted both the positive and negative tail on the same set of axes so that the skewness is visible. We have also plotted a shifted version of the negative tail which shows that if we move the negative tail 0.1 units to the left the distribution would be almost symmetric.

5.5 Stock Price Modelling for Price Impact

We consider a trade superposition model, used by Bouchaud *et al* [1], of the impact of one given trade propagated up to time n , where the price at time n is written as the sum of all past trades:

$$p_n = \sum_{n < n'} G_0(n - n') \epsilon_{n'} \ln V_{n'} + \sum_{n < n'} \eta_{n'}, \quad (15)$$

where $G_0(\cdot)$ is the ‘bare impact’ function for a trade on a subsequent price fluctuation. This ‘bare impact’ function is assumed to be a non random function that is only dependent on time lags. The η_n are random variables and are assumed to be independent of the ϵ_n . These η_n are residual values and are used to model any source of price change that cannot be described by the direct impact of trading e.g. changes in the bid-ask spread.

The ‘bare impact’ function $G_0(\ell)$ is defined as the average impact of a trade after ℓ trades. This could theoretically be fitted using empirical data, but it would cost too much to make it practical.

Using this representation, Bouchaud *et al* [1] defined the price increment between an arbitrarily chosen initial time 0 and time ℓ as:

$$\begin{aligned} p_\ell - p_0 = & \sum_{0 \leq n < \ell} G_0(\ell - n) \epsilon_n \ln V_n \\ & + \sum_{n < 0} [G_0(\ell - n) - G_0(-n)] \epsilon_n \ln V_n + \sum_{0 \leq n < \ell} \eta_n. \end{aligned} \quad (16)$$

If the trade signs ϵ_n were independent random variables, it would make the computation of the response and diffusion function a much simpler task. The response function would be given by:

$$\mathcal{R}(\ell) = E(\ln V) G_0(\ell), \quad (17)$$

and the bare impact function and the response function would be directly proportional.

In our study of trade sign auto-correlation we found, like Bouchaud *et al* [1], that in our ϵ ’s have long range correlations. Thus our average response functions should be computed as:

$$\begin{aligned}\mathcal{R}(\ell) &= E(\ln V)G_0(\ell) + \sum_{0 < n < \ell} G_0(\ell - n)\mathcal{C}_1(n) \\ &+ \sum_{n > 0} [G_0(\ell + n) - G_0(n)]\mathcal{C}_1(n).\end{aligned}\tag{18}$$

If the price impact of every trade was permanent, i.e. $G_0(\ell) = G_0$, according to Bouchaud *et al* [1], we would get:

$$\mathcal{R}(\ell) = E(\ln V)G_0 \left[1 + \sum_{0 < n < \ell} \mathcal{C}_0(n) \right],\tag{19}$$

using Eq (6). If $\mathcal{C}_0(n)$ dies down as a power-law with an exponent $\gamma < 1$, then $\mathcal{R}(\ell)$ will be amplified by a significantly large factor as ℓ increases. This is how Bouchaud *et al* [1] discovered that the bare impact function $G_0(\ell)$ should decay with time, as it would need to offset the amplification effect of the trade correlations. They conclude that price impact cannot be permanent.

Bouchaud *et al* [1] also argue that the integral of $\mathcal{C}_0(\ell)$ can be understood as the effective number of correlated successive trades i.e.

$$N_e \simeq 1 + \sum_{\ell=1}^n \mathcal{C}_0(\ell) \approx 1 + \frac{C_0}{1-\gamma} n^{1-\gamma},\tag{20}$$

consecutive trades.

For our case, we find $N_e \sim 35$, meaning that the effect of one trade should be amplified by a factor of 35. Which in turn would imply that the response and diffusion functions should increase by a factor of 35 in our empirical data - which we do not observe, nor did [1].

5.6 Fitting a Bare-Impact Function

From the empirical constraint that $\mathcal{D}(\ell)$ must be approximately linear in ℓ , Eq (6) Bouchaud *et al* made the ansatz that the bare impact function $G_0(\ell)$ also decays as a power-law:

$$G_0(\ell) = \frac{R_0}{(\ell_0 + \ell)^\beta} (\ell \geq 1). \quad (21)$$

It follows that one can estimate $\mathcal{D}(\ell)$ in the large ℓ limit.

For the purposes of our study, we have chosen to use 500 lags as a cut off point for $\mathcal{D}(\ell)$ and 200 lags for $\mathcal{R}(\ell)$ when fitting the empirical data to the model. This is due to the fact that the empirical $\mathcal{R}(\ell)$ functions tend to misbehave in the higher lags and whereas the the empirical $\mathcal{D}(\ell)$ functions tend to be smoother for most lag points. We also see that some of the more illiquid stocks have very few trades in a day, for these stocks the statistics in the higher lags would be tainted by the fact that we would not have a large enough data pool.

5.6.1 Fitting β

In [1] Bouchaud *et al* find that the general formula for the diffusion has the form:

$$\begin{aligned} \mathcal{D}(\ell) = E(\ln^2 V) & \left[\sum_{0 \leq n < \ell} G_0^2(\ell - n) + \sum_{n > 0} [G_0(\ell + n) - G_0(n)]^2 \right] \\ & + 2\Delta(\ell) + D_\eta \ell, \end{aligned} \quad (22)$$

where $\Delta(\ell)$ is the contribution of the effect of the correlations:

$$\begin{aligned} \Delta(\ell) = & \sum_{0 \leq n < n' < \ell} G_0(\ell - n) G_0(\ell - n') \mathcal{C}_2(n' - n) \\ & + \sum_{0 < n < n'} [G_0(\ell + n) - G_0(n)] [G_0(\ell + n') - G_0(n')] \mathcal{C}_2(n' - n) \\ & + \sum_{0 \leq n < \ell} \sum_{n' > 0} G_0(\ell - n) [G_0(\ell + n') - G_0(n')] \mathcal{C}_2(n' + n). \end{aligned} \quad (23)$$

When $\gamma < 1$, one finds that the $\Delta(\ell)$ is dominant, and all three terms scale a $\ell^{2-2\beta-\gamma}$, provided $\beta < 1$. Therefore, the Hurst exponent of price change becomes $2H = 2 - 2\beta - \gamma$. Thus for purely diffusive fluctuations,

$$\beta = \frac{1 - \gamma}{2}. \quad (24)$$

If the price is sub-diffusive

$$\beta > \frac{1 - \gamma}{2} \quad (25)$$

and if the price is super-diffusive

$$\beta < \frac{1 - \gamma}{2}. \quad (26)$$

The problem in Eq (20) was resolved in [1] using that $\beta \approx (1 - \gamma)/2$ such that the slowly decaying response function, $\mathcal{R}(\ell)$, with exponent β needs to exactly cancel the slow decaying auto-correlation of the trades with exponent γ in order to fit the diffusive nature of the data ⁷. Considering a power law in the bare impact function for their micro-structure model they found the Hurst exponent of the price fluctuations to be $2H = 2 - 2\beta - \gamma$. This implies that if price fluctuations are diffusive ($H = \frac{1}{2}$) over long-horizons that $\beta \approx (1 - \gamma)/2$. This means that if the auto-correlations in the sign of trades had long-memory⁸, then $\beta < \frac{1}{2}$ (as $0 < \gamma < 1$). This is consistent with our results.

Using the fact that $\mathcal{D}(\ell) \propto \ell^{2-2\beta-\gamma}$, we were able to fit β 's using our empirical $\mathcal{D}(\ell)$'s for AGL, APN and SBK. For the fitting procedure we used our empirical $\mathcal{D}(1)$ ($D(1)$) as a constant such that we get our function to be $\mathcal{D}(\ell) = D(1)\ell^{2-2\beta-\gamma}$. Thus at $\ell = 1$ the ℓ term in our function collapses and we are left with our empirical $\mathcal{D}(1)$.

One can see from Figure (37) and (38) that $\mathcal{D}(\ell) = D(1)\ell^{2-2\beta-\gamma}$ is a good fit to the diffusion functions of AGL, APN and SBK. In Table (5) we also see that our R^2 statistics for our fitting procedure show us that more than 95% of the variance in the empirical results of $\mathcal{D}(\ell)$ is explained by the model when using the statistics shown in the table.

Huberman and Stanzl [78] argued that there is only one value for which models of permanent market impact are free from arbitrage: when $\beta =$

⁷Using $\mathcal{R}(\ell) \propto \sum_{0 < n < \ell} \mathcal{C}_1(n)$ where $\mathcal{C}_1 \approx \mathcal{C}_0(\ell)E(\ln V)$ such that $\gamma < 1 \Rightarrow \mathcal{R}(\ell) \sim \ell^{1-\gamma}$

⁸A long-memory process has $H = 1 - \frac{\gamma}{2}$, a short-memory process is $H = \frac{1}{2}$

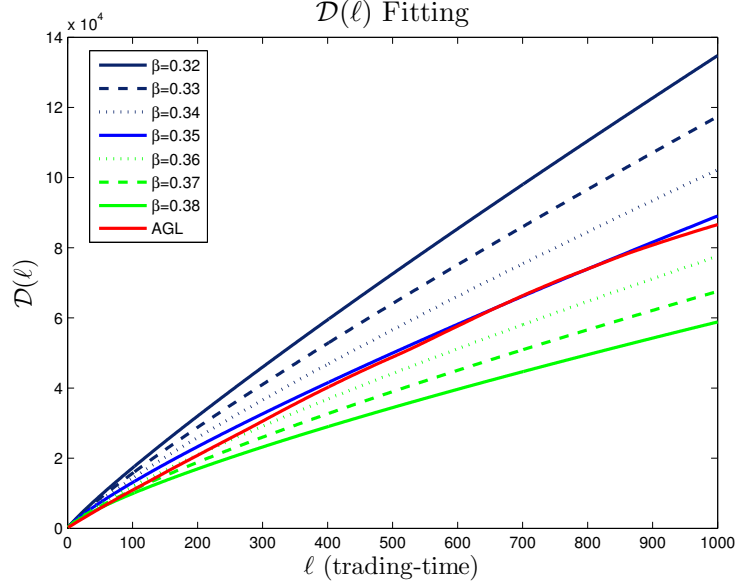


Figure 37: In this figure we show a brief graphical interpretation of the underlying procedure involved in fitting a β to the $\mathcal{D}(\ell)$ function of AGL. As one can see, when using $\gamma = 0.4666$ as we found when fitting the $\mathcal{C}_0(\ell)$ for AGL, we find a good fit using a β around 0.35.

1. Since our $\beta \neq 1$, it could either mean arbitrage opportunities exist or can be interpreted as excluding a permanent market impact effect from our dynamics.

Gatheral [2] argues that $\beta \geq 2 - \frac{\ln 3}{\ln 2} \approx 0.4$; faster decay is theoretically ruled out by no-dynamic arbitrage. This implies $\gamma \leq 1 - 2\beta \approx 0.2$. Our result appears to violate with the conditions of no-dynamic-arbitrage. The principle of no-dynamic arbitrage can be understood as the idea that price manipulation is not possible. No price manipulation [2] means that a round-trip trade [78] must not have a negative cost. The empirical deviations from the $\gamma \leq 1 - 2\beta$ condition, which empirically permits price manipulation, can be interpreted to imply that market agents do not just use a linear forecast of order flow when adapting their trading to market conditions [2].

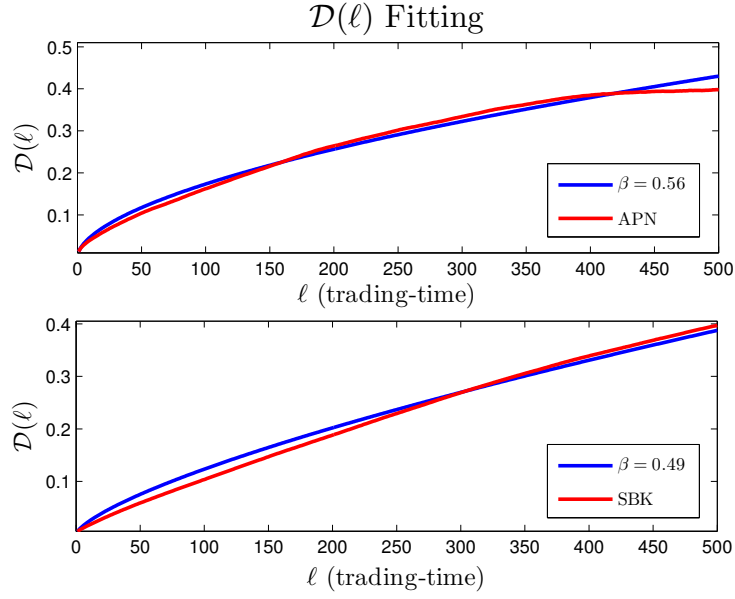


Figure 38: Here we have plotted our results of fitting β for APN and SBK. When using $\gamma = 0.3092$ and $\gamma = 0.3092$, found when fitting a power-law function to $\mathcal{C}_0(\ell)$, to fit our empirical $\mathcal{D}(\ell)$ for APN and SBK respectively both have a good fit. Our β for APN is found to be around 0.56, while for SBK we find $\beta \approx 0.49$ give us a good fit for the diffusion function.

5.6.2 Fitting ℓ_0 and R_0

After fitting a β for each stock we are left with two parameters in Eq (18) that we need to solve for. The procedure for fitting values to ℓ_0 and R_0 is more complicated than fitting β . This involves fitting our empirical $\mathcal{R}(\ell)$'s to the discrete sum Eq (18) which cannot be condensed into a simple form like we saw earlier that the $\mathcal{D}(\ell)$ could. This fitting procedure involved high performance computing techniques as well as the use of a miniature super computer. Even with these resource, the $\mathcal{R}(\ell)$ fitting procedure was a time consuming endeavor.

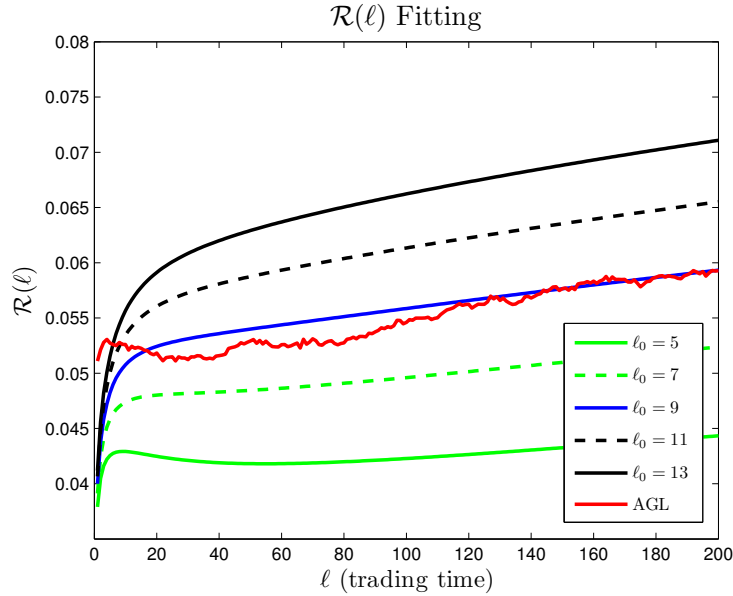


Figure 39: In this figure we have the empirical $\mathcal{R}(\ell)$ for AGL against 5 different response functions using the Eq (18), the best fit we achieved with $\ell_0 = 9$, $R_0 = 0.0144$ and $\beta = 0.35$. We found that AGL had a reasonable fit with an R^2 of 0.7.

In our section on the response function we noted the response function tends to misbehave at larger lags, to prevent this from affecting our fitting process we reduced our lags to a maximum of 200. To simplify the procedure we have decided to use the first point in our empirical response function to back solve for R_0 . This leaves us with one parameter to fit. We used a brute

force sum for different values of ℓ_0 to try and fit our empirical results to Eq (18).

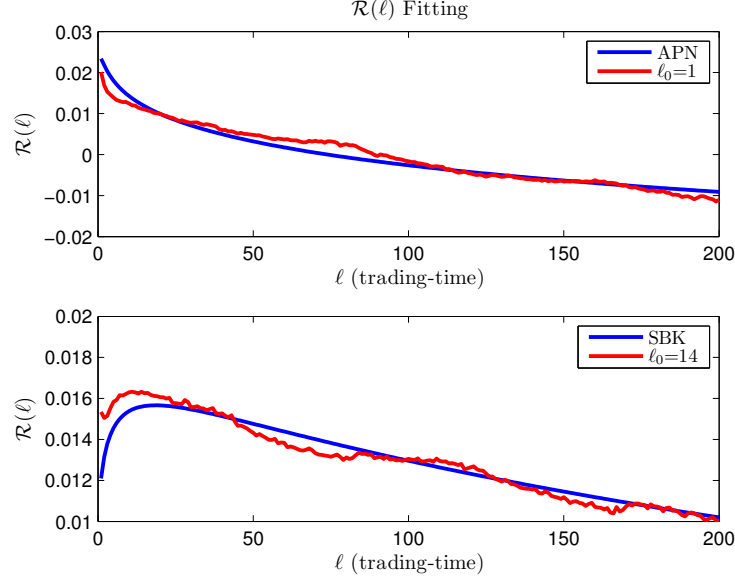


Figure 40: In this figure we show the empirical $\mathcal{R}(\ell)$ for APN and SBK plotted against their best fit response functions. We see that these two stocks fit even better to the model than AGL with both R^2 's around 0.95.

We found that for most of our stocks we could not find an a ℓ_0 that would allow for an acceptable fit, in fact the results of our fitting showed that the shape of our empirical response functions were not consistent with the possible shapes of the proposed model. This may suggest that most of our 46 stocks do not have a power-law bare impact function as described by Bouchaud *et al* [1].

5.6.3 BGPW Model Fitting Results

We have tabulated the statistics, for the five stocks, found during the fitting procedure below. These can be used to compute the $\mathcal{C}_0(\ell)$, $\mathcal{D}(\ell)$ and $\mathcal{R}(\ell)$ functions as described by the model.

Ticker	<i>AGL</i>	<i>APN</i>	<i>BIL</i>	<i>MTN</i>	<i>SBK</i>
C_0	0.15	0.26	0.15	0.15	0.17
$\mathcal{D}(1)$	0.0281	0.0129	0.0228	0.0057	0.0047
γ	0.4666	0.3092	0.3430	0.3295	0.3033
β	0.3543	0.5632	0.4247	0.4910	0.4922
$R^2(\mathcal{D}(\ell))$	0.9927	0.9877	0.9935	0.9818	0.9887
R_0	0.0144	0.0071	0.0166	0.0079	0.0075
ℓ_0	9	1	11	14	14
$R^2(\mathcal{R}(\ell))$	0.7003	0.9537	0.5665	0.8432	0.94874

Table 6: This table shows some of the statistics involved in fitting the diffusion function to empirical data for AGL, APN and SBK.

One can see in Table (6) that in all cases more than 95% of the variance in our empirical $\mathcal{D}(\ell)$ is explained by the fitted model. This shows us that $\mathcal{D}(\ell) = D(1)\ell^{2-2\beta-\gamma}$ is a good model for the diffusion function.

The model for $\mathcal{R}(\ell)$ is not as airtight as that of $\mathcal{D}(\ell)$ with only three of the stocks having an R^2 of more than 0.8 and only 56.65% of the variance of in the empirical response function of BIL can be explained by the fitted model.

6 Conclusion

In our study, we found that the intra-day information in the markets investigated is readily available. The availability of data provides for an exciting opportunity to study the price discovery process central to modern financial markets. The study of such data requires a sophisticated understanding of a market structures, regulation and micro-structure. Such studies require the ability to manage and analyze large inhomogeneously sampled datasets with heterogeneous data types and a variety of noise sources. As such, this type of analysis falls squarely in the domain of big-data, computational finance and market-micro-structure.

The importance of understanding market micro-structure and knowing the price-impact function for specific stocks is seen on a practical level when trying to minimize trading cost as a result of slippage. Although one can use Dark Pools to execute large trades with relative anonymity, such trades still rely on reference prices from Lit Markets and, as such, the processes of price discovery and information aggregation investigated in this project in terms of price impact still remain important.

Trading costs can be a significant factor in asset management and hedge funds where the company's livelihood is performance based. Thus, it becomes prudent to understand the dynamics at play. One of the important issues investigated by market micro-structure studies is market manipulation.

Gatheral [2], among others, suggested that price manipulation would be possible amongst low liquidity stocks. We see that this could be the case but that attempts to move the price would affect the liquidity dynamics of the stock and, as such, the feedbacks in the system would make it difficult to realize a profit without sophisticated feedback-control information into one's trading algorithms, in order to control the risk and return profile of such trading strategies.

In our examination of intra-day trade statistics, we found that for our five selected JSE stocks: the intra-day average volumes have a U-shape across the day, intra-day return distributions have a leptokurtic behavior, intra-day spreads have a L-shape peaking in the morning and becoming more steady across the day. When looking at price impact we found that lower trade volumes tend to be associated with higher returns and vice versa give

volumes and returns an almost hyperbolic relationship. We also saw, like Lillo *et al* [44], that the one lag price impact had similar behaviors for our five selected stocks.

Similar to the results of Bouchaud *et al* [1] we find that our trade signs have a long term auto-correlation; this suggests that the price impact of trading on the JSE should be temporary. In our empirical diffusion functions we find that the stock price, for our JSE stocks, has a sub diffusive behavior. We also find that the empirical response functions are positive and misbehave after about 100 lags. When we looked at the distributions of the u_ℓ 's we see that they are positively skewed and a slight shift could make the distributions more symmetric.

When fitting our empirical data to the BGPW [1] model we find that our trade sign auto-correlations can be fitted to $\mathcal{C}_0(\ell) = \frac{c_0}{\ell^\gamma}$, we found that for our JSE stocks $\gamma \approx 0.32$. Using $\mathcal{D}(\ell) = D(1)\ell^{2-2\beta-\gamma}$ to fit our diffusion functions we find that $\beta \approx 0.47$ for our JSE stocks. We note that our γ 's and β 's behave in such a way that $\gamma \geq 1 - 2\beta$, which break the no-dynamic arbitrage constraint discussed by Gatheral [2], this suggests that there may be an opportunity for price manipulation at high enough trading rates. When fitting the $\mathcal{R}(\ell)$ function we had mixed results, seeing strong fits for APN, MTN and SBK, and less acceptable fits for AGL and BIL which suggests that some stocks may require a different shape of 'bare impact' function.

After studying some of the key factors of price impact and the shapes and general behavior of price impact curves; we are of the opinion that, although there are certain circumstances where price manipulation is possible, in the context of the South African market it remains to be seen whether one can consistently profit from price manipulation without asymmetric access to either information, the markets involved or a significant commitment of capital in order to move prices in order to game other algorithms. It is becoming increasingly clear that there is a grey area between what is or is not considered price-manipulation from a market regulation perspective; for example, algorithms that are aiming to achieve best execution for client trades through the exploitation of sophisticated algorithms which can purposefully impact prices in order to achieve such objectives.

It is critical that the regulatory environment is cognisant of developments in market micro-structure as changes in market structure can inadvertently

provide asymmetric advantages to various participants. Having said that, there is a variety of features in modern markets to ensure that price dynamics are fair and transparent given that sophisticated market manipulation is in fact possible due to the lead-lag effects created by the dynamics between volume and prices, as we have documented in this study. The boundary between such price manipulations and what is or is not permitted in terms of the trading rules and regulation of a given trading venue, needs to be understood and monitored with significant care.

This study did not include an analysis of the auction period. Neither the opening nor closing auctions. In order to have a more complete understanding of the market structure an inclusion of the closing auction is necessary. This is important as many large market participants often prefer trading at the market close, both as an attempt to target the market closing price, which is important for portfolio attribute of pension funds, as well as providing large players improved volume and an environment where the risk of being gamed by other market participants is reduced because of the volumes involved.

One important reason for the study of market-micro structure would be the build of a catalog of stylized facts that could be used as a minimal set of required behavior that one would require from a simulated market. It could be useful to build a simulated South African market with the aim of recovering the stylized facts that our type of study catalogues. Such a development would allow us to carry-out numerical experiments towards understanding the impact of various changes in both the agents participating in the markets as well as changes due to market-structure and regulation, although this would be beyond the scope of this project. Understanding the phenomenology of the South African market is an important component of being able to realistically simulate such a market.

References

- [1] J-P. Bouchaud, Y. Gefen, M. Potters, and M. Wyart. Fluctuations and response in financial markets: the subtle nature of random price changes. *Quantitative Finance*, 4(2):176–190, 2004.
- [2] J. Gatheral. No-dynamic-arbitrage and market impact. *Quantitative Finance*, 10(7):749–759, September 2009.
- [3] D. Abbott, J. P. Bouchaud, X. Gabaix, and J. L. McCauley. Noise and fluctuations in econophysics and finance. 5848, 2005.
- [4] M. Aitken, A. Kua, P. Brown, T. Watter, and H. Y. Izan. An intraday analysis of the probability of trading on the asx at the asking price. *Australian Journal of Management*, 20(2):115–154, 1995.
- [5] A. A. Alzahrani. Four essays on return behaviour and market microstructures: Evidence from the saudi stock market. 2009.
- [6] J. P. Bouchaud. The subtle nature of financial random walks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 15(2):026104, 2005.
- [7] J-P. Bouchaud, J.D. Farmer, and Lillo F. How markets slowly digest changes in supply and demand. *Handbook of Financial Markets: Dynamics and Evolution.*, 1, 2009.
- [8] D. Bozdog, I. Florescu, K. Khashanah, and J. Wang. A study of persistence of price movement using high frequency financial data. *Handbook of Modeling High-Frequency Data in Finance*, 4:27, 2011.
- [9] L. K. C. Chan and J. Lakonishok. The behavior of stock prices around institutional trades. *The Journal of Finance*, 50(4):1147–1174, 1995.
- [10] K. H. Chung and Y. Kim. Volatility, market structure, and the bid-ask spread*. *Asia-Pacific Journal of Financial Studies*, 38(1):67–107, 2009.
- [11] Z. Eisler, J. P. Bouchaud, and J. Kockelkoren. The price impact of order book events: market orders, limit orders and cancellations. *Quantitative Finance*, 12(9):1395–1419, 2012.
- [12] J. D. Farmer, L. Gillemot, F. Lillo, S. Mike, and A. Sen. What really causes large price changes? *Quantitative finance*, 4(4):383–397, 2004.

- [13] J. D. Farmer, P. Patelli, and I. I. Zovko. The predictive power of zero intelligence in financial markets. *Proceedings of the National Academy of Sciences of the United States of America*, 102(6):2254–2259, 2005.
- [14] A. Gakhovich. The holiday effect in the central and eastern european financial markets. 2011.
- [15] N. Huth and F. Abergel. Some recent results on high frequency correlation. *Market Microstructure: Confronting Many Viewpoints*, pages 61–86.
- [16] B. Jacobsen, A. Mamun, and N. Visaltanachoti. Seasonal, size and value anomalies. 2005.
- [17] F. Lillo and J. D. Farmer. The long memory of the efficient market. *Studies in Nonlinear Dynamics & Econometrics*, 8(3):1–33, 2004.
- [18] F. Lillo, J. D. Farmer, and R. N. Mantegna. Price impact function of a single transaction. *The Complex Dynamics of Economic Interaction*, pages 153–160, 2004.
- [19] E. Moro, J. Vicente, L. G. Moyano, A. Gerig, J. D. Farmer, G. Vaglica, F. Lillo, and R. N. Mantegna. Market impact and trading profile of hidden orders in stock markets. *Physical Review E*, 80(6):66–102, 2009.
- [20] P. P. Tan, D. U. Galagedera, and E. A. Maharaj. A wavelet based investigation of long memory in stock returns. *Physica A: Statistical Mechanics and its Applications*, 391(7):2330–2341, 2012.
- [21] P. Weber. *Response functions, trading strategies, and random matrices: Analysis of large fluctuations and correlations in stock price diffusion*. PhD thesis, Universität zu Köln, 2006.
- [22] A. Obizhaeva and J. Wang. Optimal trading strategies and supply/demand dynamics. *Journal of Financial Markets*, 16(1):1–32, 2013.
- [23] R. Almgren and N. Chriss. Optimal execution of portfolio transactions. *Journal of Risk*, 3(2):5–39, 2000.
- [24] E. F. Fama. Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25(2):383–417, 1970.

- [25] P. A. Samuelson. Proof that properly anticipated prices fluctuate randomly. *Industrial management review*, 6(2), 1965.
- [26] P. Bak, M. Paczuski, and M. Shubik. Price variations in a stock market with many agents. *Physica A: Statistical Mechanics and its Applications*, 246(3):430–453, 1997.
- [27] J-P. Bouchaud, M. Mzard, and M. Potters. Statistical properties of stock order books: empirical results and models. *Quantitative Finance*, 2(4):251–256, 2002.
- [28] D. Challet and R. Stinchcombe. Analyzing and modeling limit order markets. *Physica A: Statistical Mechanics and its Applications*, 300(1):285–299, 2001.
- [29] D. Challet and R. Stinchcombe. Non-constant rates and over-diffusive prices in a simple model of limit order markets. *Quantitative Finance*, 3(3):155–162, 2003.
- [30] M. G. Daniels, J. D. Farmer, G. Iori, and E. Smith. Quantitative model of price diffusion and market friction based on trading as a mechanistic random process. *Physical Review*, 90(10), 2003.
- [31] S. Maslov. Simple model of a limit order-driven market. *Physica A: Statistical Mechanics and its Applications*, 278(3):571–578, 2000.
- [32] F. Slanina. Mean-field approximation for a limit order driven market model. *Physical Review*, 64(5), 2001.
- [33] K. Burdett and M. O’hara. Building blocks: An introduction to block trading. *Journal of Banking & Finance*, 11(2):193–212, 1987.
- [34] M.o Avellaneda and S. Stoikov. High-frequency trading in a limit order book. *Quantitative Finance*, 8(3):217–224, 2008.
- [35] G. Huberman and W. Stanzl. Optimal liquidity trading. *Review of Finance*, 9(2):165–200, 2005.
- [36] P. Lakner, J. Reed, and S. Stoikov. High frequency asymptotics for the limit order book. *Preprint available on the authors webpage*, 2013.

- [37] X. Li and A. Krause. An evolutionary multi-objective optimization of market structures using pbil. pages 78–85, 2010.
- [38] M. D. Mitton. *Derivative pricing and optimal execution of portfolio transactions in finitely liquid markets*. PhD thesis, University of Oxford, 2007.
- [39] C. C. Moallemi, B. Park, and B. Van Roy. Strategic execution in the presence of an uninformed arbitrageur. *Journal of Financial Markets*, 15(4):361–391, 2012.
- [40] F. Naujokat. *Stochastic control in limit order markets*. PhD thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II, 2011.
- [41] Y. Nevmyvaka, Y. Feng, and M. Kearns. Reinforcement learning for optimized trade execution. pages 673–680, 2006.
- [42] R. Velasco-Fuentes and W. L. Ng. Nonlinearities in stochastic clocks: trades and volume as subordinators of electronic markets. *Quantitative Finance*, 11(6):863–881, 2011.
- [43] A. Madhavan. Market microstructure: A survey. *Journal of Financial Markets*, 3:205–258, 2000.
- [44] F. Lillo, J. D. Farmer, and R. N. Mantegna. Master curve for price-impact function. *Nature*, 421:129–130, 2003.
- [45] M. O’Hara. Market microstructure theory. *Blackwell, Oxford*, 1995.
- [46] J. Hasbrouck. Empirical market microstructure. *Oxford Press, Oxford*, 2007.
- [47] T. Chordia, R. Roll, and A. Subrahmanyam. Commonality in liquidity. *Journal of Financial Economics*, 56(1):3–28, 2000.
- [48] T. Chordia, R. Roll, and A. Subrahmanyam. Market liquidity and trading activity. *Journal of Finance*, 56(2):501–530, 2001.
- [49] T. Chordia, R. Roll, and A. Subrahmanyam. Liquidity and market efficiency. *Journal of Financial Economics*, 87(2):249–268, 2008.

- [50] J. Hasbrouck. Measuring the information content of stock trades. *Journal of Finance*, 46:179–207, 1991.
- [51] J. Hasbrouck. Stalking the efficient price in market microstructure specifications: an overview. *Journal of Financial Markets*, 5(3):329–339, 2002.
- [52] G. Iori, M. G. Daniels, J. D. Farmer, Krishnamurthy S. Gillemot, L., and E. Smith. An analysis of price impact function in order-driven markets. *Physica A*, 324:146–151, 2003.
- [53] Y. Tse. Market microstructure of ft-se 100 index futures: An intraday empirical analysis. *Journal of Futures Markets*, 19(1):31–58, 1999.
- [54] F. Aparicio and J. Estrada. Empirical distributions of stock returns: Scandinavian securities markets, 1990-95. *Working Paper (Department of Statistics and Econometrics, Carlos III University)*, 2(4):251–256, 1997.
- [55] G. Booth, J. Hatem, I. Virtanen, and P. Yli-Olli. Stochastic modeling of security returns: Evidence from the helsinki stock exchange. *European Journal of Operational Research*, 56:98–106, 1992.
- [56] Y. Malevergne, V. Pisarenko, and D. Sornette. Empirical distributions of stock returns: between the stretched exponential and the power law? *Quantitative Finance*, 5(4):379–401, 2005.
- [57] A. Peiro. The distribution of stock returns: International evidence. *Applied Financial Economic*, 4:431–439, 1994.
- [58] J. Smith. The probability distribution of market returns: A logistic hypothesis. *Ph.D. dissertation, University of Utah.*, 1981.
- [59] D. Easley, M. Lopez de Prado, and M O’Hara. The volume clock: Insights into the high frequency paradigm. *The Journal of Portfolio Management*, 2012.
- [60] A. Madhavan. Market microstructure: A practitioner’s guide. *Financial Analysts Journal*, pages 28–42, 2002.

- [61] A. Abhyankar, D. Ghosh, E. Levin, and R. J. Limmack. Bid-ask spreads, trading volume and volatility: Intra-day evidence from the london stock exchange. *Journal of Business Finance & Accounting*, 24(3):343–362, 1997.
- [62] A. Sarr and T. Lybek. Measuring liquidity in financial markets. *International Monetary Fund*, 2:1–64, 2002.
- [63] T. H. McInish and R. A. Wood. An analysis of intraday patterns in bid/ask spreads for nyse stocks. *The Journal of Finance*, 47(2):753–764, 1992.
- [64] J. P. Bouchaud. The endogenous dynamics of markets: price impact and feedback loops. *arXiv preprint arXiv:1009.2928*, 2010.
- [65] W. X. Zhou. Universal price impact functions of individual trades in an order-driven market. *Quantitative Finance*, 12(8):1253–1263, 2012.
- [66] D. B. Keim and A. Madhavan. The upstairs market for large-block transactions: Analysis and measurement of price effects. *Review of Financial Studies*, 9(1):1–36, 1996.
- [67] R. Almgren, C. Thum, E. Hauptmann, and H. Li. Equity market impact. *Risk*, 2005.
- [68] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.
- [69] J. Gatheral, A. Schied, and A. Slynko. Exponential resilience and decay of market impact. *Econophysics of order-driven markets.*, pages 225–236, 2011.
- [70] A. Alfonsi, A. Schied, and A. Schulz. Optimal execution strategies in limit order books with general shape functions. *Discussion paper CER-MICS, projet MATHFI*, 2007.
- [71] D. Bertsimas and A. Lo. Optimal control of execution costs. *Journal of Financial Markets*, 1:1–50, 1998.
- [72] M. Avellaneda, J. Reed, and S. Stoikov. Forecasting prices from level-i quotes in the presence of hidden liquidity. *Algorithmic Finance*, forthcoming.

- [73] A. Alfonsi, A. Schied, and A. Slynko. Order book resilience, price manipulation and the positive portfolio problem. *Discussion paper CERMICS, projet MATHFI*, 2007.
- [74] R. Cont, A. Kukanov, and S. Stoikov. The price impact of order book events. *Journal of financial econometrics*, 12(1):47–88, 2013.
- [75] A. Lo. Long-term memory in stock market prices. *Econometrica*, 59:1279–1313, 1991.
- [76] W. Willinger, M.S. Taqqu, and V. Teverovsky. Stock market prices and long-range dependence. *Finance and Stochastics*, 3:1–13, 1999.
- [77] D. Wilcox and T. Gebbie. Serial correlation, periodicity and scaling of eigenmodes in an emerging market. *International Journal of Theoretical and Applied Finance*, 11(7):739–760, 2008.
- [78] G. Huberman and W. Stanzl. Price manipulation and quasi-arbitrage. *Econometrica*, 72(4):1247–1275, 2004.

A Stocks Used In Study

A.1 JSE Stocks

Equity Code	Company
ABL	African Bank Investments Ltd
AGL	Anglo American PLC
AMS	Anglo American Platinum Ltd
ANG	AngloGold Ashanti Ltd
APN	Aspen Pharmacare Holdings Ltd
ARI	African Rainbow Minerals Ltd
ASA	Barclays Africa Group Ltd
BIL	BHP Billiton PLC
BTI	British American Tobacco PLC
BVT	Bidvest Group Ltd
CFR	Cie Financiere Richemont SA
CSO	Intu Properties PLC
DSY	Discovery Ltd
EXX	Exxaro Resources Ltd
FSR	FirstRand Ltd
GFI	Gold Fields Ltd
GRT	Growthpoint Properties Ltd
HAR	Harmony Gold Mining Co Ltd
IMP	Impala Platinum Holdings Ltd
INP	Investec PLC
IPL	Imperial Holdings Ltd
KIO	Kumba Iron Ore Ltd
LON	Lonmin PLC
MNP	Mondi PLC
MSM	Massmart Holdings Ltd

Equity Code	Company
MTN	MTN Group Ltd
NED	Nedbank Group Ltd
NPN	Naspers Ltd
NTC	Netcare Ltd
OML	Old Mutual PLC
PPC	PPC Ltd
RDF	Redefine Properties Ltd
REI	Reinet Investments SCA
REM	Remgro Ltd
RMH	RMB Holdings Ltd
SAB	SABMiller PLC
SAP	Sappi Ltd
SBK	Standard Bank Group Ltd
SHF	Steinhoff International Holdings Ltd
SHP	Shoprite Holdings Ltd
SLM	Sanlam Ltd
SOL	Sasol Ltd
TBS	Tiger Brands Ltd
TRU	Truworths International Ltd
VOD	Vodacom Group Ltd
WHL	Woolworths Holdings Ltd/South Africa

A.2 BM&FBOVESPA Stocks

Equity Code	Company
BBAS3	Banco do Brasil SA
BBDC4	Banco Bradesco SA
BISA3	Brookfield Incorporacoes SA
BRFS3	BRF SA
BRML3	BR Malls Participacoes SA
BTOW3	B2W Cia Digital
BVMF3	BM&FBOVESPA SA
CCRO3	CCR SA
CIEL3	Cielo SA
CMIG4	Cia Energetica de Minas Gerais
ECOD3	Vanguarda Agro SA
ELPL4	Eletropaulo Metropolitana Eletricidade de Sao Paulo SA
EMBR3	Embraer SA
FIBR3	Fibria Celulose SA
GFSA3	Gafisa SA
JBSS3	JBS SA
LAME4	Lojas Americanas SA
LIGT3	Light SA
LLXL3	Prumo Logistica SA
LREN3	Lojas Renner SA
PDGR3	PDG Realty SA Empreendimentos e Participacoes
PETR3	Petroleo Brasileiro SA
PETR4	Petroleo Brasileiro SA
RSID3	Rossi Residencial SA
SBSP3	Cia de Saneamento Basico do Estado de Sao Paulo (SABESP)
USIM3	Usinas Siderurgicas de Minas Gerais SA
VALE5	Vale SA