

ALPHABET BIAS AND TRADING ACTIVITY: A SOUTH AFRICAN PERSPECTIVE

Master of Commerce (50% Research) in Finance

Brigitte Winterstein (2401298)



Supervisor: Mr. Douglas Mbululu

School of Economics and Finance

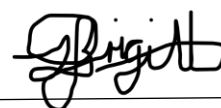
A research report submitted to the Faculty of Law, Commerce and Management, at the University of the Witwatersrand in partial fulfilment for the degree of Master of Commerce (50% Research) in Business Finance.

Abstract

This study examines the relationship between alphabetic position and trading activity for shares on the Johannesburg Stock Exchange (JSE) to determine if alphabet bias exists on the JSE. The study employs share turnover and illiquidity measures as proxies for trading activity, whilst incorporating controls for firm-level characteristics including market capitalisation, share price, and analyst coverage. Panel regressions from January 2000 to December 2023 do not reveal a statistically significant alphabetical bias on the JSE, in large contrast to evidence from studies from developed markets. This discrepancy is likely due to the relatively small universe of shares on the JSE, which can contribute to lower search costs, thus reducing the need for simplified methods such as alphabetic ordering. The empirical results confirm that market capitalisation, share price, analyst coverage, book-to-market ratios, leverage, and systematic risk maintain significant relationships with trading activity, consistent with established theoretical frameworks. Multiple robustness tests incorporating alternative measurement approaches and sensitivity analyses validate these conclusions. The findings contribute to the behavioural finance literature in South Africa and show that fundamental firm characteristics rather than alphabetical bias drives trading behaviour on the JSE. Robustness tests, which include alternative measures and sensitivity tests confirm these findings. Further research is suggested in the study to examine similar patterns across other emerging markets can be found in order to for the findings to be generalised.

Declaration

I, Brigitte Winterstein, declare that this research paper is my own work and that I have correctly acknowledged the work of others. It is submitted to fulfil the partial requirements for the degree of Master of Commerce in Business Finance at the University of the Witwatersrand, Johannesburg. I declare that this research paper has not been submitted for any other degree or examination in this or any other institution.



Brigitte Winterstein

July 2025

Contents

Abstract	ii
Declaration	iii
Contents	iv
List of Figures	vi
List of Tables	vii
Definitions and abbreviations	viii
1. Introduction	1
1.1. Background and overview	1
1.2. Motivation of the study	4
1.4. Research objective	5
1.5. Hypothesis	5
2. Literature review	5
2.1. Alphabet bias	6
2.2. Trading activity and its determinants	8
2.2.1. <i>Share price</i>	20
2.2.2. <i>Book-to-market</i>	20
2.2.3. <i>Number of analysts</i>	20
2.2.4. <i>Difference of opinions</i>	21
2.2.5. <i>Forecast dispersion</i>	21
2.2.6. <i>Positive return and negative return</i>	21
2.2.7. <i>Market age</i>	21
2.2.8. <i>Beta</i>	22
2.2.9. <i>Earnings surprise</i>	22
2.3. Alphabet bias in stock markets and its impact on trading activity	23
3. Methodology	27
3.1. Firm name position	27
3.2. Dependent Variables	28
3.4. Explanatory Variables	32
3.5. Rank effects on firms with lower visibility	36
3.6. Robustness Checks	36
3.7. Regression model	37
3.8. Data	37
3.9. Tools	39
4. Results	40
4.1. Descriptive statistics	40

4.2. Pre-regression testing	41
4.2.1. <i>Correlation</i>	41
4.2.2. <i>Multicollinearity</i>	44
4.3. Turnover	46
4.3.1. <i>Turnover and continuous position</i>	46
4.3.2. <i>Turnover and firm characteristics</i>	47
4.3.2. <i>Alternative regressions</i>	48
4.3.3. <i>Regressions with earnings data</i>	52
4.3.4. <i>Robustness checks</i>	54
4.4. Illiquidity	57
4.4.1. <i>Illiquidity and continuous position</i>	57
4.4.2. <i>Illiquidity and firm characteristics</i>	58
4.4.3. <i>Alternative regressions</i>	59
4.5. Discussion of results	62
4.6. Conclusions on hypotheses	64
5. Conclusion	64
References	68

List of Figures

<i>Figure 1 Number of active firms in the dataset</i>	<i>28</i>
<i>Figure 2 Amihud illiquidity.....</i>	<i>30</i>
<i>Figure 3 Log of share volume over time</i>	<i>40</i>
<i>Figure 4 Log of share turnover over time.....</i>	<i>41</i>
<i>Figure 5 Correlation matrix plot visualising the correlation between independent variables – generated with R code.....</i>	<i>43</i>
<i>Figure 6 Plot of key independent variables</i>	<i>44</i>
<i>Figure 7 Plot of logged turnover against continuous position</i>	<i>47</i>

List of Tables

<i>Table 1 List of variables</i>	<i>31</i>
<i>Table 2 Data obtained for the study.....</i>	<i>38</i>
<i>Table 3 Correlation between dependent and independent variables.....</i>	<i>43</i>
<i>Table 4 Multicollinearity</i>	<i>45</i>
<i>Table 5 Regression of logged share turnover on continuous position.....</i>	<i>46</i>
<i>Table 6 Regression of share turnover on continuous position and control variables</i>	<i>48</i>
<i>Table 7 Regression with share price.....</i>	<i>49</i>
<i>Table 8 Multicollinearity of updated model.....</i>	<i>49</i>
<i>Table 9 Regression with market age.....</i>	<i>50</i>
<i>Table 10 Regression with firm size</i>	<i>51</i>
<i>Table 11 Regressions with earnings data</i>	<i>53</i>
<i>Table 12 Regression of logged rand share volume on continuous position.....</i>	<i>54</i>
<i>Table 13 Multivariate regression of rand share volume on continuous position</i>	<i>54</i>
<i>Table 14 Regression of logged trading volume on continuous position</i>	<i>55</i>
<i>Table 15 Regression of turnover on position dummies</i>	<i>56</i>
<i>Table 16 Regression of liquidity on continuous position</i>	<i>58</i>
<i>Table 17 Regression of liquidity on continuous position and control variables.....</i>	<i>59</i>
<i>Table 18 Regression of illiquidity on continuous position and control variables (excluding market age and firm size).....</i>	<i>60</i>
<i>Table 19 Regression of liquidity on continuous position and control variables (excluding share price and firm size).....</i>	<i>61</i>
<i>Table 20 Regression of illiquidity on continuous position and control variables (excluding share price and market age)</i>	<i>62</i>

Definitions and abbreviations

AIC Akaike's Information Criteria

ALPHABET BIAS The tendency to give priority to items appearing earlier in an alphabetically sorted list

AMEX American Stock Exchange

BEHAVIOURAL FINANCE A field of finance studying how human biases affect prices and trading activity in financial markets

BETA The measure of a firm's market risk that cannot be diversified away

BOOK TO MARKET Book value of equity divided by the market value of equity

EMH Efficient Market Hypothesis

HAC Heteroskedasticity and Autocorrelation-Consistent

JSE Johannesburg Stock Exchange

NASDAQ Second largest stock exchange in the United States of America

NYSE New York Stock Exchange

TRADING ACTIVITY A measure of how often a share is bought and sold

VIF Variance Inflation Factor

1. Introduction

1.1. Background and overview

Psychologically, lists often imply a hierarchy, suggesting that items that are listed first are superior to items listed further down. This form of superiority in listing would make sense when an explicit hierarchy is used to sort names, one would not expect this in a list sorted by alphabetical order or indeed an arbitrary order. Even so, appearing at the top of a list tends to provide many benefits. For instance, politicians that are at the top of the ballot get more votes (Ho & Imai, 2008). A similar phenomenon occurs in academia as well, as shown by Brogaard, Engelberg, and Parsons (2014), who find that the ordering of articles gives undue attention to articles appearing in the beginning of the journal. A paper that appears first gets 50% more citations than papers appearing at the end of the journal. This extends beyond just first placed papers as second placed papers get 26% more citations and papers placed third get 17% more citations. The implication here is that people are biased towards the items that come first regardless of the sorting mechanism.

Papers published in economics that have multiple authors tend to list the names by alphabetical order of the surnames. Researchers with names that are near the earlier portions of the alphabet get more citations. Einav and Yariv (2006) as well as van Praag and van Praag (2008) show that the benefits of alphabetic listing extend beyond citations as researchers with names at the beginning of the alphabet are successful in other areas in academia, such as being more likely to receive tenure at top schools in economics or a highly coveted award.

The order of letters in the alphabet is arbitrary and yet it appears that individuals who are fortunate enough to have inherited a name at the beginning of the alphabet receive many benefits. Richardson (2008) shows that alphabetical ordering influences academia in the medical field. The author shows that editors of a medical journal provided with an alphabetically sorted list of referees contact the people at the top of the list twice as much as people at the bottom of the list.

Reasons for this type of bias are unclear. Simon (1957) hypothesizes that given the choice among a large pool of choices, people tend to select the first option that satisfies their criteria, instead of choosing the best option. Therefore, when given an alphabetically sorted list, it is human nature to select the first options at the top of the list, naturally items that start with letters

near and around the letter A. Another example of this is demonstrated by Meer and Rosen (2011), who show that university alumnus with last names close to A donate more to their alumni than alumnus with names towards the end of the alphabet. The explanation for this is that calling lists are ranked alphabetically, so people with names near to A are more likely to get called first.

Alphabetical sorting of lists is ubiquitous across various fields. This type of list sorting occurs in financial markets as well, such as a list of shares trading on an exchange. However, biases due to alphabetical listing should not affect financial markets. This is because according to the strong form of the Efficient Market Hypothesis (EMH), share prices reflect all available information, and therefore share prices are not affected by anything else other than new information (Chau & Vayanos, 2008). Consequently, it is not possible to make riskless profits in an efficient market. The implication here is that trading activity only intensifies to incorporate new information, therefore, alphabet bias should not affect trading activity in an efficient market.

On the other hand, the EMH, has not held up well as several studies have found results contrary to the EMH. For instance, Rosenberg, Reid, and Lanstein (1985) find that buying shares with a high book value to market value and selling shares with a low book value to market value produces statistically significant abnormal returns, which contradicts the semi-strong version of the EMH, which states that publicly available information cannot be used to predict share prices. Worse still, Jegadeesh and Titman (1993) find that shares that do well in the past 3, 6, 9, and 12 months continue to do well for up to 12 months, whilst shares that perform poorly continue to drop. This is contrary to the weak form of the EMH, which states that past information cannot be used to predict future prices (Young & Auret, 2018).

The explanations usually given for deviations from market efficiency is that market participants do not display rational behaviour (Shiller, 2003), at least not in the classical economic sense of the term. Market participants have human biases that they carry into financial markets that affect share prices and trading activity. The impact of biases on trading activity has received less attention in comparison to impact of biases on share prices. If the EMH holds, then trades should only occur when there is new information. The differences in trading should be entirely explained by arrival of new information.

However, human limitations may create additional reasons to trade. For instance, Merton

(1987) develops a model for investors with incomplete information, where investors can only trade shares that they know. As a result, two similar firms will have massively different trading volumes.

Another reason for trading activity would be the need to reset the weights of shares in a portfolio following price changes. Others may trade to act on private information that they may possess, to earn abnormal returns. Even in the case where all investors have equal access to information, they may not all come to the same conclusions about the firms and as a result, trading activity amongst firms becomes massively different.

Chordia, Huh and Subrahmanyam (2007) conduct a major study on the determinants of trading activity. The study tests the relationship between trading activity and liquidity, number of informed traders, and the degree of uncertainty and disagreement on fundamental value. In addition to this, Chordia *et al.*, (2007) implement Merton (1987) rationale that investors trade on stocks that are visible to them. The study uses firm size, age, price, and the book-to-market ratio to measure the level of visibility of a firm. The rationale for using book-to-market as a proxy for firm visibility is that low book-to-market shares are growth firms, typically technology firms, and are therefore have a greater chance of having higher visibility. The number of analysts covering a share is used as a proxy for the number of informed agents. The difference of opinion in a firm's underlying value is proxied by leverage. Chordia *et al.*, (2007) employ share turnover as the dependent variable and is used as a measure of trading activity. The data period covered by the study is almost 40 years long, spanning from July 1963 to December 2002.

The results of Chordia *et al.*, (2007) first show that the level of trading activity has increased over time. This is attributed to the explosion of mutual funds, reduced transaction costs and algorithmic trading. The results further demonstrate that trading activity differs substantially between shares. Trading activity is affected by the previous month's share return as shares with good past performance have higher number of trades in the following trading period. Also significant is the difference in analysts' forecasts, earnings surprises, earnings volatility, price, and firm size. Share turnover in the technology sector is found to be higher than share turnover in other sectors.

The study by Jacobs and Hillert (2015) is the first to investigate the impact of alphabetical listing in financial markets. Financial markets have many alphabetically sorted lists for shares

traded on the exchange. Since there is a great number of options to choose from, there are high search costs involved in share selection. Therefore, it is likely that ordering has an impact on investors' decision making. To test this, Jacobs and Hillert (2015) apply cross-sectional regressions of trading activity and liquidity on firm characteristics, among which alphabetical listing is the main variable of interest. The findings indicate that firm position in an alphabetically sorted list influences both turnover and liquidity. Therefore, the authors conclude that alphabetical ordering affects trading activity and liquidity, controlling for other factors.

Itzkowitz, Itzkowitz, and Rothbort (2015) comment that alphabetical listing of shares is likely to have a significant effect on trading activity. The authors investigate this using data on US shares from 1985 to 2012. The study tests the relationship between liquidity and alphabetic listing of shares. The univariate tests show that shares with names starting with letters in the beginning of the alphabet have trade volumes greater than shares with initials later down in the alphabet. The result remains true even after controlling for other firm characteristics.

Itzkowitz *et al.*, (2015) also test the impact of alphabet bias on market values of firms. Since shares with names starting with early alphabet letters have higher liquidity, they should also have higher valuations. The results show that firms with early alphabet names have higher market-to-book ratios than firms with names starting later in the alphabet. Interestingly, the effects become stronger after the onset of internet and online trading. Itzkowitz *et al.*, (2015) confirm the findings of Jacobs and Hillert (2015), further showing that shares with names starting with letters close to A trade more often and are often more valuable than shares that start with a letter further down the alphabet.

1.2. Motivation of the study

This study is motivated by the results of Jacobs and Hillert (2015) who find that US traded shares at the top of an alphabetically sorted list have trading activity that is between 5% and 15% higher than stocks that appear near the bottom of an alphabetically sorted list. A study of a similar nature has yet to be conducted in South Africa.

1.3. Problem statement

Research has shown that alphabetical listing has a profound effect on the behaviour of people using that list, most notably that the people at the top of an alphabetically sorted list receive more benefits than people at the bottom of the list. This should not be the case in financial

markets as Fama (1970) has hypothesised that markets are efficient, therefore the only factor that affects share prices is new information, which is random and unpredictable. However, the study of Jacobs and Hillert (2015) has shown that alphabet bias exists in US and international financial markets.

In South Africa, little is known about the impact of alphabetical bias on trading behaviour. Worse still, not many papers exist that tackle the determinants of trading activity on the JSE. The results of Jacobs and Hillert (2015) suggests that shares listed at the top of an alphabetical list in larger markets have higher trading activity due to biases exhibited by investors and high search costs. However, the JSE's smaller share universe may reduce these cognitive biases. Therefore, this study seeks to directly address the knowledge gap on alphabet bias and determinants of trading activity on the JSE.

1.4. Research objective

The aim of this study is to determine if alphabetic bias exists on the JSE and whether it has a relationship with trading activity. A secondary aim of the research is to investigate which firm-specific factors drive trading activity. Trading activity is used in this study to describe quantifiable measures such as turnover and illiquidity, whereas trading behaviour refers to investor decisions and trading patterns, as inferred from the trading activity measures.

This is the first study in South Africa to investigate the alphabet bias on the JSE. The benefit of this study is to inform academics and market participants of the level of biases existent on the JSE, which will inform future studies and trading strategies. This study adds to the thin body of work on the determinants of trading activity on the JSE.

1.5. Hypothesis

H₀: Alphabetic listing of shares has no effect on trading activity for shares on the JSE

H₁: Alphabetic listing of shares influences trading activity for shares on the JSE

H₀: Firm-level characteristics do not have a significant relationship with trading activity

H₁: Firm-level characteristics have a significant relationship with trading activity

2. Literature review

This chapter aims to offer a discussion on the literature involving behavioural biases in market,

market efficiency and the specific bias of alphabetical lists and its effect on stock markets. The purpose of this chapter is to build a conceptual understanding for the empirical tests described in Chapter 3. The chapter is structured as follows: Chapter 2.1 provides an overview of the alphabet bias, Chapter 2.2 is a development of literature on trading activity and its determinants, and Chapter 2.3 covers alphabet bias in financial markets.

2.1. Alphabet bias

Simon (1957) argues that when choosing an option or options from a list, human beings search the list in the order that they appear until they have a satisfactory selection is made. This is done even if a criterion of suitability is available, as soon as a person comes across the first suitable option, this option is selected. Given that options are searched sequentially, this gives an advantage to items that appear at the beginning of a list over items towards the end of the list. Carney and Banaji (2012) state that human beings are biased towards items that occur first in a list. Furthermore, primacy (that is, being first) is a well-known, almost undisputed fact according to Carney and Banaji (2012). Primacy has a strong cognitive effect on human beings, inducing stronger memory, greater attachment, stronger impressions, and higher levels of persuasion (Carney & Banaji, 2012).

Being put at a top of a list, therefore, gives numerous advantages, regardless of how the list is sorted. For example, Ho and Imai (2008) conduct an experiment to determine the effect of ballot order on the number of votes received. Analysing elections across California from 1978 to 2002, the authors find that ballot order has a much greater impact on lesser-known candidates than on major candidates. Specifically, only minor candidates benefit from ballot order in general elections, well known candidates do not. However, in primaries, every candidate gains extra points by appearing earlier in the list. This shows that being highly visible pre-selection seems to undo much of the benefit of ordering.

Nonetheless, the benefits still occur, first placed items are more likely to be chosen in any type of list sorting. This occurs in academia as well. For instance, Brogaard *et al.*, (2014) show that *ceteris paribus*, articles that appear first in a journal receive 50% more citations than articles published in the same journal. Second place articles receive 26% more citations while third place articles are cited 17% more frequently. The authors conclude that article placement has a significant relationship with number of citations. However, the reason for this not clear. Articles may be placed first by editors because they know *a priori* which article is the best and therefore present it first. Alternatively, this may be due to bias, where authors simply use the

first articles they see for citations. In addition, firms also pay to appear first on google search and people rarely go the next page when searching on google.

Alphabetically sorted lists order the items in ascending order from A to Z. This puts items starting with letters A, B, C at an advantage because they are closer the list. Alphabetically sorted lists are ubiquitous in life, implying that many people potentially benefit from this. Several studies have been conducted on the effect of alphabetical ordering. As an example, Jurajda and München (2009) show that applicants at hard-to-get-into prestigious schools, who have a high position in the alphabetically sorted lists of applicants, have a higher chance of being accepted into the school.

Alphabetically sorted lists are popular in economic publications. Published articles with multiple authors have the names of the authors ordered alphabetically. Einav and Yariv (2006) investigate how this impacts the career prospects for economists. Focusing on university staff from the top 35 economics departments in US universities, the authors find that individuals with surnames starting with letters closer to A are more likely to receive a fellowship at the econometric society and are more likely to become tenured at a top ten school in economics. These results are significant. Less significant is the finding that these individuals are more likely to win the Clark Medal award and the Nobel Prize in Economics.

The findings remain statistically significant when controlling for religion, ethnicity, country of origin or departmental fixed effects. In comparison, there is no statistically significant relationship between alphabetical position and tenure for psychologists at top 35 US schools for psychology. This is because in psychology, there is no alphabetical ordering of names on co-authored papers (Einav & Yarin, 2006). The result is intuitive. If authors are consistently ordered according to alphabetic order, names beginning with letters close to A will consistently appear first and will therefore be more likely to be chosen in a satisficing selection criterion. Whereas in psychology, there is no persistent order of names such that the benefits are spread out across names over time, rather than accruing for a select group at the top of the alphabetical list, as is the case in economics.

Similarly, van Praag and van Praag (2008) find that the practice of sorting names according to alphabetical order in economics, gives advantages to authors with last names that start with letters at the beginning of the alphabet. A professor with a last name starting with the letter A will be listed as a first author on a co-authored article more often than a professor whose last

name begins with Z. Professor A will have a higher number of articles published in journals and as a result of the increased visibility from appearing first, will also experience career growth prospects greater than Professor Z. Interestingly, economists are aware of this and actively choose to go against the alphabetic ordering, knowing that appearing first in the list is important for one's career (van Praag and van Praag, 2008).

2.2. Trading activity and its determinants

Fama (1970) provides a formal framework to define movement of share prices. Fama (1970) defines an efficient market as one where prices “fully reflect” available information. The EMH comes in three forms, each differing in the definition of fully available information. The weak form of the EMH states that only past information is incorporated into share prices. The semi-strong version states that all public information is incorporated into share prices. Lastly, the strong form market efficiency includes information available only to insiders.

The implication of the EMH is that prices only react to new information, rather than existing information (Shiller, 2003). Therefore, prices only change when new information arrives. This implies that trading activity is affected by the arrival of new information, since market participants must trade on news for prices to adjust. Given that new information is unknowable, it is not possible to earn abnormal returns reliably and consistently in an efficient market.

Market efficiency is an important concept here. If markets are efficient, one should not expect human biases, such as the alphabet bias, to creep into markets. Market participants only trade on new information. However, Shiller (1981) argued that share prices are far too volatile. Present value of a firm's future cashflows does not change anywhere near as much as prices do. Therefore, there must be other reasons why market participants trade shares. Similarly, De Bondt and Thaler (1985) state that the share price volatility is too high and cannot be justified by changes in fundamental value. In addition, DeBont and Thaler (1985) discover an apparent seasonality effect in returns. Every year, share returns tended to be higher in January than they were compared to the rest of the months in the year. Clearly something is inducing extraordinary trading behaviour in the month of January.

Jegadeesh and Titman (1993) show that the shares with the highest positive return in the prior 3, 6, 9 and 12 months continue to have high positive returns in the following 3, 6, 9, and 12 months. Meanwhile, shares that have had the lowest negative returns over the same period continue to demonstrate negative returns for periods up to 12 months. The authors conclude

that the cause of this differential is human bias.

The studies finding evidence against the EMH all point to one thing, investor's reaction to information is just as important, if not more, than the information. Given that it has been sufficiently shown that markets are not efficient, it is no surprise that alphabet bias can exist in financial markets as well.

Market efficiency is generally tackled through the lens of share prices, whereas trading activity is not given as much attention. Early studies generally investigate returns or prices and only in passing mention differences in trading activity. An early example of this is Beaver (1968) which examines if investors react to earnings announcements by testing for return predictability of earnings announcements, which ultimately shows that earnings announcements contain new information about the future of the firm that investors react to, according to the EMH.

In addition to this, Beaver (1968) also investigates the impact of earnings announcements on volume, positing that even in an efficient market, earnings announcement would lead to changes in volume trading with no changes in price. This would be due to the differences in preferences in investors as although everyone agrees on the price, others would not want to hold that share at that price. Therefore, trades will occur due to individual portfolio balancing but not to input new information into the price.

Beaver (1968) finds that volume increases substantially in the week that earnings are announced, increasing by 33% on average. This shows that investors adjust their holdings based on the earnings report. Furthermore, trading volume is significantly below the average values during the eight weeks leading up to the earnings announcement. This shows that investors wait for earnings are announced before making trades. This result appears to be contradictory to the EMH at first glance, since the EMH predicts that investors forecast future earnings therefore any announcement should not impact prices or trading volumes. On the other hand, earnings announcements can be considered as new information, particularly under the weak or semi-strong version, where only past and public information respectively, is already factored into the price.

Kiger (1972) performs a similar study using NYSE listed entities between 1966 and 1969, provided they had issued quarterly earnings reports during that period, did not issue a dividend or a share split or any other important trade and price influencing information. Additionally, ideal firms for inclusion in the sample were ones that issued earnings reports as close as

possible to the end of the accounting period. This is because the larger the gap between the two, the more likely it is that earnings will be leaked to the market. The price and volume data were de-trended to isolate the impact of the earnings report on the data.

Kiger (1972) assumes that investors use the earnings report to make decisions on trades. Therefore, one would expect that trading volume would be highest on or immediately after the announcement date. To test this hypothesis, the amount of trading volume during the announcement period was compared to the trading volume during a non-announcement period.

According to the EMH, prices react instantaneously to new information, so the period used to evaluate the impact of trading volume is only three days. An alternative measure is used with a period of just five trading days. Volume data is adjusted to remove trades that were booked prior to the earnings announcement. The tests performed on the data is a sign test and Wilcoxon's Matched-Pairs Signed-Rank Test.

The results show that trading volume is significantly different during the announcement period compared to the non-announcement period. The results are the same on both the three-day trading period and the five-day period (Kiger, 1972). However, the five-day period is less significant and has smaller coefficients. This suggests that the reaction to news is relatively slow, which is in opposition to the EMH proposition that reaction to new information is instantaneous.

The EMH implicitly supposes that gathering information is costless, as argued by Grossman and Stiglitz (1980). This is not true when arbitrageurs search activities that are costly, as it would be in financial markets. Therefore, because gathering information is costly, arbitrageurs will need to be compensated for their private information. As a result, markets are never fully efficient, and prices will reflect information to the extent that it remains profitable to gather information (Grossman & Stiglitz, 1980).

Morse (1980) proposes that the costs to information gathering can potentially result in differences in trading activity. For example, privately informed investors will trade on their information before uninformed traders in the market. This suggests that volume and price changes are related to access to private information. Morse (1980) tests this proposition directly by inspecting autocorrelation of returns in high volume trading sessions. The results of the paper show that there is serial dependence in returns during periods with above average trading volume. Morse (1980) further showed that trading volume is significantly above average on

the day before a large price change. This shows that investors with access to private information trade on this information, subsequently affecting future volume and prices.

Amihud and Mendelson (1987) evaluate the phenomenon where the opening price series has different returns and variance than the closing price returns. In the paper, the authors state that share volume was consistently higher for the open than for the prior day closing volume, demonstrating differences in trading activity during daily transactions. This finding demonstrates what is referred to as the U or V shaped nature of trading volume.

The U-shaped nature of mean daily trading volume motivates the study of Admati and Pfleiderer (1988). The focus of the study is on two areas, namely trading volume and return variation. The objective of the study is to address three problem areas. Firstly, the study investigates the reason for large trading volumes at the beginning and at the end of the trading day. Secondly, the paper aims to examine the differences in return variation between periods. The third objective of the paper is to explore the rationale for increase in trading volume during periods when return variation is high. Admati and Pfleiderer (1988) model trading behaviour of investors in a market must make decisions on whether they should trade or search for differential private information.

Admati and Pfleiderer (1988) explain that investors trade primarily two reasons, which are liquidity and information, both of which are almost universally accepted explanations for trading behaviour. Traders with access to differential private information that other market participants do not have will make trades based on this information unbeknownst to other traders. Investors who make trades for liquidity are not concerned about future profits; they have other needs that must be attended to.

Admati and Pfleiderer (1988) state that a widely assumed fact of most studies is that liquidity traders are not concerned with when they make their trades. Admati and Pfleiderer (1988) relax this assumption by proposing that some of the liquidity traders prefer to and do time their trades, even if it is just to liquidate their positions. This assumption is relevant to large institutional investors who can easily move prices of a share if large trades are executed. This has an important impact on the market. The rationale of this assumption is that, when combined with trades made by privately informed investors, the trading behaviour exhibited by investors in markets can be better explained.

Admati and Pfleiderer (1988) show that liquidity traders prefer to trade when the market has the greatest number of traders, to reduce the price impact of their large trades. This results in concentrated trading from liquidity traders. Interestingly, Admati and Pfleiderer (1988) show that informed traders prefer to trade during periods when liquidity trading is at its highest. As a result, prices will have greater informational content during these periods. Admati and Pfleiderer (1988) state the combination of these two factors will result in consistent trading patterns over time, such as the U-shaped nature of mean daily trading volume.

Subsequent studies arrive at similar conclusions as Admati and Pfleiderer (1988). For instance, Foster and Viswanathan (1993) investigate for differences in volume between trading days as well as between hours in a trading day. Ideally, in an efficient market, trading volume should be the same each trading hour and on each day.

However, Foster and Viswanathan (1993) postulate that some market participants have more information than others. This then forces other less informed traders to wait until these market participants to trade on their information. Since the traders with an information advantage have far fewer uninformed traders to deal with, market makers can much more easily decipher why they want to trade, which then leads to the information known to these traders being made to the public, at which point the formerly uninformed traders join and thus increase liquidity (Foster & Viswanathan, 1993).

The implication of this is that there will be hours of the day, and days of the week, which will have less trading activity in comparison to others. The two days of no trading between Friday and Monday gives sophisticated investors time to build an information advantage, which other less informed investors are aware of. As a result, they will avoid trading on Mondays, leaving Monday with much less trading volume than other days of the week.

Gallant, Rossi and Tauchen (1992) notice that large drops in price generally come on days with large trading volumes. For instance, on the crash of 1987, the S&P 500 lost almost 23% on the day with the second largest trading volume at the time. The next day had the largest trading day ever of 608 million shares, with an increase of 5.2% on the S&P 500. In the same week, a 450 million share trading day ended with a nearly 9% increase in the S&P 500. The 7% decrease in the index on Friday, 13 October 1989 was followed by a 250% rise in trades on the subsequent Monday (Gallant *et al.*, 1992).

Gallant *et al.*, (1992) state that these events imply a relationship between price changes and trading volume. Using nonparametric and semi-nonparametric techniques on data between

1928 and 1987 for shares traded on the New York Stock Exchange (NYSE), the study's main objective is to test this relationship directly, aiming to determine a significant statistical relationship between trading volume and price. The study finds that large changes in prices are strongly related to larger future trading volumes, with the effect being mostly the same regardless of whether the price change is positive or negative.

Morse (1981) examines the fluctuations in prices and trading volume focusing on annual and quarterly announcements published in the Wall Street Journal. Morse (1981) argues that investors trade before an announcement because of divergent views of how the market will react to the information released. Another explanation offered by Morse (1981) is that market participants will trade prior to the announcement date because they have private information. Others will trade to close positions taken to speculate on the information released in the announcement.

Morse (1981) defines volume number of shares traded divided by the number of shares outstanding. The tests performed examine if there is a statistically significant difference in the price and volume during the announcement period compared to the average day. The results show that the most significant trading occurs on the day before the announcement and on the announcement day. Interestingly, Morse (1981) notes that the level of reaction to information observed through volume differs to the level of reaction observed through a change in price.

Bamber (1986) conducts an extensive study on trading volume, covering multiple hypotheses. Firstly, Bamber (1986) re-tests the positive relationship between volume and earnings announcements on extended dataset. Relatedly, Bamber (1986) tests the hypothesis that the unexpected portion of trading volume is positively related to the surprise in earnings. Bamber also investigates the relationship between firm size and trading volume, hypothesizing that firm size will have a negative relationship with unexpected trading volume. Bamber (1986) further hypothesises that the firm size and the magnitude of the earnings surprise will also be negatively related.

The results of Bamber (1986) showed that trading volume after announcement of earnings was significantly larger than the median value of trading volume in that year. Furthermore, trading volume was found to have a statistically significant positive relationship with unexpected earnings.

Bamber (1986) finds a negative relationship between trading volume and firm size, showing that the increase in trading volume was larger for small firms. Bamber (1986) states that this

might be due to the increased difficulty of forecasting earnings for smaller firms, therefore the earnings surprise ends up being much larger for smaller firms than for large firms. This conjecture is supported by the findings of the paper, which shows that smaller firms earnings tend to be more surprising than larger firms and the earnings surprise changes each year. Moreover, trading volume is higher for smaller firms even when controlling for the surprise in earnings. The explanation provided for this is that earnings announcements for small firms are a major source of information for small firms that would have less coverage from analysts and the media.

Karpoff (1986) shows that trading volume is positively correlated with the total number of shares outstanding but is negatively correlated with the bid-ask spread. Karpoff (1986) empirical tests are unable to provide evidence for the conjecture that differences of opinion to new information results in differences in trading volume. Karpoff (1986) explains this stating that investors may view the information identically but would have had different beliefs prior to the release of the new information.

An example provided by Karpoff (1986) is that investors are aware that new information will be released and enter speculative trades about what this information might contain. The increase in trading volume is therefore to exit the trades once the content of the information is known, rather than due to differences in opinion of the information.

However, Karpoff (1986) states that the surprising content of the information released may result in increased trading, therefore trading volume is more likely to be positively related to information surprise rather than information releases.

Merton (1987) proposes a model that in a world with incomplete information, it is only possible for investors to trade the securities that are known to them. Therefore, this leads into different levels of trade activity even for shares that are otherwise similar. Similarly, Goetzmann and Kumar (2008) in showing how undiversified most investors are, also demonstrate that most unsophisticated market participants only trade amongst a very small group of well-known shares. Varian (1985) argues that differences of opinion amongst market agents leads to differences in trading volume, even when agents are privately informed.

Kim and Verrecchia (1991) conduct a study to establish theoretical explanations for the movements in both price and volume following announcement of new information. The focus for the explanations of changes in volume during announcements is on the quality of the

information that exists prior to the announcement. Kim and Verrecchia (1991) assume that traders form their portfolios using information known to themselves and others before the announcement is made. When new information is released to the public, the beliefs of the traders will be updated, thus causing traders to trade for a second time. Additionally, Kim and Verrecchia (1991) assume that the levels and quality of information known by individual traders is not identical. This causes a diverse spectrum of reactions among traders, which in turn results in positive volume.

The model developed by Kim and Verrecchia (1991) shows that there is a direct relationship between the change in price and the new information as well as the quality of the information held prior to the announcement. Furthermore, the model shows that trading volume is directly related to change in price as well as the quality of the trader's preannouncement information.

As mentioned earlier, Beaver (1968) showed that changes in price are due to the consensus agreement to the new information whereas changes in volume are due to individual trader needs. Kim and Verrecchia (1991) posit that the reason for the unique motivations for trading is due to the difference in quality of the information that they had prior to the announcement. For those traders with low quality information, new information is of greater importance to them and consequently, the impact on their understanding of the firm is correspondingly greater. Therefore, trading volume shows totality of the differing reactions from traders. This differs from price changes, which shows traders mean reaction (Kim & Verrecchia, 1991).

The theories put forward by Kim and Verrecchia (1991) motivate the study of Atiase and Bamber (1992), who venture to conduct a study on the impact of earnings announcements on trading volume. Atiase and Bamber (1992) proxy for the quality of preannouncement information using the spread and range of analyst's forecasts, since quality of preannouncement information is not directly observable. The study utilises a data sample containing 5,282 annual earnings releases.

Atiase and Bamber (1994) find that trading volume increases with the scale of the price reaction. For instance, if the absolute change in price is 20%, the trading volume is larger than if the absolute change in price is 10%. As stated before, this finding aligns with prior research that shows a positive relation between trading volume and absolute value of abnormal returns.

An additional finding from the Atiase and Bamber (1994) study is that the proxies for quality of preannouncement information have a statistically significant relationship with changes in

trading volume during earnings announcements periods. This is true even when the study controls for the change in price. This shows that the quality of information held by investors prior to release of new information does have an impact on trading volume. Furthermore, it provides an explanation for the relationship between trading volume and changes in prices following release of new information.

It has been shown in the literature that both prices and trading volume react to new information such as earnings announcements. As a response to this, Harris and Raviv (1993) develop a model aimed at providing explanations for the noted differences in trading volume across the literature. Harris and Raviv (1993) state that trading volume is induced by sales for tax, liquidity, rebalancing and speculative purposes, although the focus of the explanations is on speculative trading. Harris and Raviv (1993) postulate that investors speculate on information because they have differing views on the impact of the new information on the future of the firm. Similarly, speculative trading sometimes occurs because of information asymmetry. Harris and Raviv (1993) focus on the former in the explanations, stating that investors have strong convictions about their interpretation of the data, which is referred to as differences of opinion.

Harris and Raviv (1993) assume that investors have identical a priori views on the prospects of the firm. Once information on the firm is released, investors will change their opinion on the firm, at which point, differences of opinion arise. Harris and Raviv (1993) assume two types of traders, with both types trading for speculative purposes and being risk neutral. In the model, both investors agree on the assessment on the rating of the information (which is positive or negative) but do not agree on whether it will impact the firm. When information is positive, investors who believe the information impacts the firm will have a higher return expectation than the other investors who do not. As a result, they will end up owning all the shares following consecutive positive news. The converse is true, such that trades will only happen when previously cumulated information moves from positive to negative (Harris & Raviv, 1993).

Harris and Raviv (1993) findings show that there is a positive relationship between the changes in price and trading volume and that changes in the average forecast of the returns and trading volume. Harris and Raviv (1993) also show that average trading volume is higher on market openings than at any other point.

Kandel and Pearson (1995) tackle the commonly held assumption of heterogeneity in financial markets. Typically, economic theory assumes that market participants arrive at the same

conclusions in their interpretations of information. To test this assumption, the study examines trade volume and changes to forecasts by analysts during the periods when earnings are announced. The study is performed using data on trade volume and share returns during periods of earnings announcements.

The results of the study show that the trading volume during the announcement period are statistically significantly different from the trading volume during the non-announcement period. Kandel and Pearson (1995) argue that the difference in trading volume cannot be attributed to just the price changes. Therefore, Kandel and Pearson (1995) conclude that these findings do not align with the assumption of homogeneous beliefs.

Furthermore, Kandel and Pearson (1995) argue that the changes to forecasts made by analysts following earnings announcements shows that market participants do not hold identical beliefs. Alternatively, Kandel and Pearson (1995) state that homogeneous expectations is possible if private information is only produced during earnings announcements. This then suggests that trading activity is influenced by the number of analysts producing information on a firm. Therefore, the conclusion to be drawn from these findings is that the number of analysts has a significant impact on trading activity. This is further evidenced by Brennan and Subrahmanyam (1995), which shows that the use of analyst coverage as a proxy for the number of informed traders is accurate.

Brennan, Jegadeesh and Swaminathan (1993) examine the relationship between financial analysts and how quickly prices change when new information is released. This is motivated by theoretical developments postulating that the speed in the change in share price will increase with the increase in analysts covering the firm. Brennan *et al.*, (1993) employ Granger causality tests to determine if the returns on portfolios containing shares with large analyst following Granger cause the returns on the portfolios containing shares with low analyst following (and vice-versa).

Lee and Swaminathan (2000) examine the predictive power of trading volume for momentum returns, using the turnover ratio rather than direct trading volume. Lee and Swaminathan (2000) demonstrate that the momentum of Jegadeesh and Titman (1993) experiences long-term reversal. Following this, Lee and Swaminathan (2000) show that past trading volume has predictive power for the duration and significance of momentum returns. The study finds that winners with high historical volume reverse quicker than winners with low historical volume.

This finding makes sense intuitively, as more trades will lead to quicker corrections to the initial underreaction. Some of the explanations found in Lee and Swaminathan (2000) are in the affirmative to this intuition. For instance, Lee and Swaminathan (2000) state that trading volume increases the flow of information on winner shares, which slows down the momentum.

Lee and Swaminathan (2000) show that trading volume can be used to enhance momentum portfolios by forming momentum on shares with low trading volumes, which results in slower reversals. Consequently, it can be argued that momentum has an impact on trading volume. Lee and Swaminathan (2000) state that the findings of the study shows that volume is not just related to liquidity but is also related to some investment strategies. Lee and Swaminathan (2000) state the surprising findings indicate that relatively little is known about trading volume.

Lee and Swaminathan (2000) suggest that the results can be explained by concluding that trading volume contains useful information on the market's uncertain estimations of future earnings and thus fundamental value. This is demonstrated by the finding that low volume shares tend to be under-valued in comparison to the opposite pairing. This finding remains consistent even when valuing firms using multiples, current operating performance, past operating performance and earnings surprise.

Lee and Swaminathan (2000) conclude that although volume is clearly related to momentum, it can be complex to describe the relationship. Lee and Swaminathan (2000) close by stating that the most puzzling finding is that volume can be used to enhance momentum portfolios. This is because it provides even stronger evidence against market efficiency, as publicly available information is not fully reflected in prices.

Grinblatt and Keloharju (2001) show that investors have a high propensity to trade shares of firms listed in Finland that are within their vicinity, communicate in their native language and are headed by individuals with similar cultures. As expected, this finding is strongest among retail investors, who tend to be less sophisticated and have less access to information. Additionally, retail investors do not have access to massive teams, which makes their search costs higher. This result shows that visibility is important to investors for selecting shares to trade.

Loughran and Schultz (2005) show similar results in the US. The paper investigates of the effect that a firm's location in the country has on its liquidity. Firms in rural areas have much lower trading activity, receive less analyst coverage, and have lower levels of institutional

ownership in comparison to firms in urban areas.

The findings of Loughran and Schultz (2005) are related to visibility, which in turn is related to attention. Barber and Odean (2008) hypothesise that given the search costs involved in share selection, investors will often select shares that grab their attention. These are shares that are reported about in the news, have experienced a large return in a single trading day, or have had an abnormally large trading day. Barber and Odean (2008) further hypothesise that since retail investors tend to only sell shares that they own, they should be net buyers of shares that are attention grabbing.

Barber and Odean (2008) formally test their hypothesis using a share sorting methodology. Shares are sorted using events that are potentially attention grabbing for investors. These events are extraordinary trading volume, extraordinary returns on a single trading day, and firms being reported in the news. The results show that retail investors are susceptible to purchasing shares that are attention grabbing. Investors buy more shares than they sell on days when trading volume is high and when single day returns are extreme in either direction.

Therefore, it has been shown that trading volumes are higher than would be expected given a rational expectations model. However, the trading activity between shares that are almost identical. For instance, Chordia *et al.*, (2007) report two firms with virtually identical market capitalisation, which had a difference in share turnover by a factor of greater than eight. The aim of this study was to investigate such differences. This study is one of the most influential papers in trading activity and forms the basis of many papers to follow.

Firstly, Chordia *et al.*, (2007) define the variables that are used to test their research question. The authors identify share turnover as the key dependent variable, where share turnover is the number of shares traded divided by the number of shares outstanding. The reason for this choice is based on Lo and Wang (2000), who show that share turnover is the best measure for trading activity because share turnover definition is the same for all investors in a market since it is a relative measure. For instance, two investors in a market holding the same two in the same proportion of shares, 50% in each, but with different number of shares held. If both investors traded divested out of one of the shares, the number of shares traded would be different, the value of shares traded would be different, and although the number of trades for each would be the same, the total number of trades for the two would be different than the individuals. Share turnover would be consistent for each investor and for both investors and hence is the best

proxy for trading activity.

The independent variables are based on various theories on trading activity. Each of the determinants are discussed in the following sections.

2.2.1. Share price

As mentioned by Merton (1987), investors largely trade shares that they have sight of. To proxy for the visibility of a share, Chordia *et al.*, (2007) utilise the log of the share price. The inclusion of price as a proxy for visibility is motivated by the proposition of Brennan and Hughes (1991), which stated that brokers advertise cheap shares more so than shares with high prices to increase their commissions.

2.2.2. Book-to-market

Another measure of visibility employed by Chordia *et al.*, (2007) is the book-to-market ratio, which is defined as the book value of equity divided by the market value of equity. The rationale for this inclusion is the perception of low book-to-market shares have of being growth firms, such as technology companies, which have higher visibility due to investors wanting to earn outsized returns from firms with high growth potential. Also included is the log of market capitalisation, motivated by the study by Lo and Wang (2006), which proposes that both volume and price impact returns.

2.2.3. Number of analysts

An additional reason to trade shares would be to act on superior information in comparison to other market participants and even other companies. To proxy for this, Chordia *et al.*, (2007) utilise the number of analysts covering a firm. This choice is reasonable considering that the quality of information on a firm with 1000 analysts would be higher than for a firm with no analysts. Consequently, investors feel more comfortable investing in a firm where they have enough information to make an investment decision. Furthermore, the content of analyst reports can have an impact on trading. For example, if more than half of the analysts covering a share release a recommendation to sell the company's shares, this will result in market participants to begin selling the firms shares. This is because they believe that analysts have access to superior information. In regressions, Chordia *et al.*, (2007) use the log of one plus the number of analysts covering the firm.

2.2.4. *Difference of opinions*

Related to analyst coverage is the diverse set of opinions between investors on the market. Investors will likely have differing views on the prospect of the firm, consequently impacting how they trade. Chordia *et al.*, (2007) speculate that the dispersion of views is at its widest on firms with high debt levels. The authors of the paper base this argument on agency theory, where it has been postulated that owners of a business will take excessive risks to salvage their equity in the firm. The increase in risk of such firms then results in investors having substantially different views on the future of the firm. As a results, this leads to higher trading activity. It is difficult to argue against this viewpoint. Naturally, leverage is proxied by the debt-to-asset ratio, which is total debt divided by total assets. Total debt includes both short-term and long-term debt (Chordia *et al.*, 2007).

2.2.5. *Forecast dispersion*

An added measure of differing opinions used by Chordia *et al.*, (2007) is analyst forecast dispersion. This is motivated by the findings of Diether, Malloy, and Scherbina (2003) who show that firms with vast differences in analyst forecasts earn lower subsequent returns in comparison to similar shares with lower amounts of forecast dispersion. Forecast dispersion is calculated as the standard deviation of future earnings forecasted by analysts (Chordia *et al.*, 2007).

2.2.6. *Positive return and negative return*

Investment theory dictates that investors also trade because they need to realign their portfolios to their original weights due to price changes. Chordia *et al.*, (2007) speculate that this need to rebalance is based on past share returns. Therefore, past share returns themselves also have an impact on the trading activity of shares. To account for this, Chordia *et al.*, (2007) use two variables, namely positive return and negative return. The positive return variable is defined as the return in the month if the return is positive, or zero if it is not. Conversely, negative return is the return in the month if the return is negative, or zero if it is not.

2.2.7. *Market age*

Chordia *et al.*, (2007) speculate that the age of the firm has an impact on how much of its shares get traded. The argument is that new entrants to a stock exchange have large amounts of publicity, which could result in higher share turnover. This seems counterintuitive as one would expect a firm that is older and more established to have higher trading activity than new firms

that investors are unsure of. The argument of Chordia *et al.*, (2007) might be true in the scenario of a unicorn company such as Google compared to a company with minimal growth prospects like General Electric, for example. These two examples are obtuse, as they are two outlier firms on opposite ends of the growth spectrum. Many new listings do not materialise, as evidenced by the long run underperformance of IPOs (Ritter, 1991).

Either way, the inclusion of firm age is logical and justified. The variable is calculated as the log of one plus the number of months the firm has been listed (Chordia *et al.*, 2007). However, given that this is not a proxy for firm age as companies exist before they list, the variable name is changed in this study to market age. It is, however, constructed exactly as it was by Chordia *et al.*, (2007).

2.2.8. Beta

Chordia *et al.*, (2007) further include betas in the regression. The reason for this is that for shares where information is sparse, investors are more likely to make mistakes in forecasting future performance. As a result, they will trade more of these shares to correct their previous mistakes. This means that firms with high betas will have higher trading activity (Chordia *et al.*, 2007).

2.2.9. Earnings surprise

Another proxy for the difficulty of making future forecasts on firms with low quality information used by Chordia *et al.*, (2007) is earnings surprise, which is measured as the absolute difference between the most recent quarter's earnings and the earnings from the previous quarter. An additional variable employed is the earnings volatility, which is the standard deviation of earnings from the eight previous quarters (Chordia *et al.*, 2007). Both variables are included as proxies for changes in fundamental values. If earnings surprise or earnings volatility is high, then this means that the fundamental value of the company is changing substantially between each quarter. Therefore, investors will trade in and out of these shares, with the magnitude of trading increasing depending on the size of the earnings surprise.

Following this, Chordia *et al.*, (2007) regress trading activity on each of the variables. The regressions are performed on monthly panel data. Chordia *et al.*, (2007) use Newey and West (1987) t-statistics to gauge the significance of the coefficients in the presence of autocorrelation among the variables.

The results of the regression show that positive return, negative return, leverage, price and market age all have significant explanatory power for trading activity at the 1% level. The explanatory power of the model increases significantly when earnings surprise, earnings volatility and number of analysts are included in the model. According to Chordia *et al.*, (2007) this shows these three variables have significantly more impact on trading activity than the other variables. This result is intuitive based on direct observations of stock markets and how investors react to missed earnings forecasts or a drop in earnings from a previous period.

The significance of the other variables is not surprising either, given the theoretical background. For instance, large positive or negative returns will induce investors to buy or sell more of the shares. The statistical significance of both positive and negative return confirms this theoretical understanding. Likewise, the effect of leverage on trading activity is as hypothesized, that is there is a positive relationship between leverage and trading activity. Firms with high amounts of leverage will have higher amounts of trading activity (Chordia *et al.*, 2007). Therefore, the paper has strong findings for the effect of firm visibility on trading activity.

An important factor that affects firm visibility is advertising. Many firms focus on advertising to gain market share and increase sales. However, Grullon, Kanatas and Weston (2004) show that firms that spend more on advertising have greater visibility not just to customers but also to investors. Firms with a large advertising spend have more investors, both retail and institutional. In addition, greater advertising spend is associated with more liquidity. Therefore, advertising spend has an impact on a firm's market value (Grullon *et al.*, 2004).

Brandt, Brav, Graham and Kumar (2009) investigate the increase in idiosyncratic volatility that occurred 1962 to 1997. The study finds that shares with low prices are mostly traded by retail investors. Shares that have a larger portion of its trading volume coming from retail investors have higher idiosyncratic volatility. The study also finds that during high profile events such as share splits, both idiosyncratic volatility and retail trading activity increases.

2.3. Alphabet bias in stock markets and its impact on trading activity

As noted in Section 2.1, humans have been shown to exhibit alphabetic bias particularly when scanning through alphabetically sorted lists, with more attention being given to the items appearing first. Jacobs and Hillert (2015) postulate that this same bias possibly exists on stock markets as well. This is because it is common practice in finance and investment to list shares

in alphabetical order in financial publications. Jacobs and Hillert (2015) state that since the universe of investable companies is massive and investors have limited time, resources and attention to evaluate them all, the order in which a company in a list is presented is likely to have an impact on investment decisions. The study therefore sets out to test for the existence of the alphabetic bias on stock markets and is the first study to do so. Specifically, the study tests the impact of alphabetic bias on trading activity and liquidity.

Jacobs and Hillert (2015) agree with Merton (1987) that investors can only invest in shares that are well known to them. Therefore, visibility is an important factor for the trading activity of a share. One of the markers of visibility being at the top of a list of shares. Consequently, shares that are highly visible, tend to have more liquidity and higher amounts of trading activity. Jacobs and Hillert (2015) argue that studying the determinants of trading activity and liquidity is economically important because these characteristics are inherent to markets. Jacobs and Hillert (2015) state that liquidity is important as it enables information dissemination through the trades of agents with private information, which results in prices reflecting information. This argument is relevant for South Africa as well considering that Young and Auret (2018) show that the increase in liquidity on the JSE has coincided with reduced return predictability.

Jacobs and Hillert (2015) rely on the methodology of Chordia et al., (2007). So just as in Chordia et al., (2007), the key dependent variable is trading activity, proxied by share turnover, dollar volume and share volume. However, Jacobs and Hillert (2015) also investigate liquidity, in addition to trading activity. Just as in Chordia et al., (2007), share turnover is calculated by dividing the number of shares traded and the number of outstanding shares. Dollar volume is the number of shares traded multiplied by the price of the share, whereas share volume is simply the number of shares traded. Liquidity is based on the formula developed by Amihud (2002) illiquidity ratio, which is given in Equation 3 of this study, in the Methodology section. All the dependent variables used in the regressions are logged to negate the effect of outliers, which can be significantly large with volume and liquidity data.

The key independent variable is the position of the firm in a list sorted by firm name. The alphabetic position is measured by two variables. The main one is called continuous position. To build the continuous position variable, firms are sorted alphabetically and ranked based on position in the list. The variable is then obtained by dividing the rank by the total number of firms in the list. This will then result in values between 0 and 1. The construction of the

continuous position variable is performed for each month in the sample (Jacobs & Hillert, 2015).

The alternative definition for alphabetic position is five dummy variables. For the first position dummy, it is calculated as equal to 1 if the firm is in the top 5% of firms for that month, else it is equal to zero. Similarly, the second position dummy is set to 1 if the firm is between the top 5% and top 25% of firms, or else it is set to zero. The same is done for the next two position dummies, that is between 25%, 50% and 75%. This is done to consider the diminishing returns of being at the top of the list as the position reduces. For example, in a list of 1000 firms, a firm listed at 150 is high up the list but does not get the same benefit of a firm at 50.

For the control variables, Jacobs and Hillert (2015) rely on all the variables used by Chordia et al., (2007) described in detail above. Since these have been sufficiently covered from Chapter 2.2.1 to 2.2.9, they will not be repeated here. The rationale for including these as control variables is that they have been shown to have an impact on trading activity and therefore it is important to control for these variables.

Jacobs and Hillert (2015) also include variables that do not appear in the regressions of Chordia et al., (2007). For instance, the study includes the twelve-month momentum return based on the reasoning that shares that have done well in the past have higher visibility to investors. Similarly, the six-month cumulative return and the highest price value in the past year is included for similar reasons.

Also included is idiosyncratic volatility, which is calculated by obtaining the standard deviation of the residuals from regressing excess share return on a factor model. Jacobs and Hillert (2015) also use dummy variables for firms that are members of a share index because inclusion in an index may result in heightened visibility. Another aspect that increases visibility is advertising spend, as shown by Grullon et al., (2004). Based on this, Jacobs and Hillert (2015) include the ratio of advertising spend to total sales.

Also related to visibility is location, as shown by Loughran and Schultz (2005) who find that firms in urban areas have greater trading activity than firms in rural areas. Due to this finding, Jacobs and Hillert (2015) include dummy variables based on the classification of the firm's location. Similarly, media coverage increases firm visibility and as discovered by Hillert, Jacobs and Müller (2014), firms with high media coverage have greater share turnover than other firms with less media coverage. Based on this finding, Jacobs and Hillert (2015) include a dummy variable for media coverage is also added. Lastly, Jacobs and Hillert (2015) control

for industry effects using a dummy variable based on the industry classifications of Fama and French (1997).

The study uses panel regression method on data from the New York Stock Exchange (NYSE) and the NASDAQ between January 1983 and December 2011. As with Chordia et al., (2007), standard errors are calculated using the approach of Newey and West (1987) to negate serial correlation.

The results of the regressions show continuous position has an impact on the differences turnover and liquidity. The results show that firms at the top of the list have 10% more share turnover than firms at the bottom. Likewise for liquidity, firms at the top of the list are roughly 8% more liquid than firms at the bottom of the list (Jacobs & Hillert, 2015).

The results from the position dummies are similar as well. The firms in the top 5% of the list had 12% more share turnover than the firms in the other position groupings. Similarly, liquidity is 13% higher for firms in the top 5% than the other firms (Jacobs & Hillert, 2015).

Itzkowitz *et al.*, (2015) state that given the large volume of information on stock markets investors cannot possibly select shares using this data and will result to simple heuristics. Itzkowitz *et al.*, (2015) hypothesise that since shares are typically presented alphabetically, there will be a strong alphabet bias in trading activity as investors will select shares from alphabetically sorted lists. Given that the list is ordered alphabetically, investors are more likely to arrive at shares to include in portfolio from shares at the top of the list.

Itzkowitz *et al.*, (2015) explain that their two explanations for this from behavioural finance, which are the satisficing and the bias referred to as status quo. Satisficing is where individuals will select the first acceptable option, rather than the best option. When applied to stock markets, investors select the most easily available acceptable share to buy. The status quo basis is the case where investors will choose to go with the default listing rather than reorder the list, even when it is trivial to do so such as in online lists.

Itzkowitz *et al.*, (2015) test the relationship between alphabetically sorted shares and liquidity. The results show that shares at the top of the alphabetically sorted list have greater trading volume than shares lower down the list. Furthermore, shares at the top of the list have greater turnover and dollar volume than shares at the bottom of the list. Further to this, shares at the top of the alphabetically sorted list have lower illiquidity, as measured by Amihud (2002) illiquidity measure. These results confirm the findings of Jacobs and Hillert (2015).

Itzkowitz *et al.*, (2015) further show that the difference in trading activity between shares at the top of the list increases over time. This reflects the increase in information brought by the internet, which has increased search costs for investors. In addition, there is no difference in trading activity for shares before 1999.

The literature as reviewed above identifies multiple determinants for trading activity. A key critique of the literature reviewed is that the paper's mostly examine financial markets of developed countries, with emerging markets given far less attention. Jacobs and Hillert (2015) for example largely focus on the US market, where search costs can indeed be significant due to the large number of shares. However, the applicability of these findings to smaller markets such as the JSE, is limited due to lower search costs. In a small market like the JSE, it is much more likely that the traditional markers of firm visibility will be much more likely to drive trading activity.

The key drivers of trading activity highlighted by Chordia *et al.* (2007) such as firm size, share price, and analyst coverage. However, the paper is largely focused on the US market, and it is not clear if this would apply in an emerging market as well. A survey of the available literature reveals that these studies, while robust, rarely address behavioural biases from smaller markets. Therefore, this study adds to the literature by examining the alphabet bias on the JSE.

3. Methodology

This chapter outlines the data sources, variable construction, and empirical approach used to test the objectives of the study. The objectives of this study are to determine whether alphabetic bias and firm specific characteristics have a significant relationship with trading activity.

3.1. Firm name position

This study uses the methodology from the study by Jacobs and Hillert (2015). Two variables are defined to measure the firms position in the list of companies called continuous position and position dummy. Continuous position is the company's rank in an alphabetically sorted list divided by the total number of shares in a month. This value is therefore between 0 and 1.

Following Jacobs and Hillert (2015) to construct the continuous position variable, a list of firm names in each month was obtained. To list firm names for the study, any share that did not have valid observations for both number of shares outstanding and number of shares traded in a particular month was excluded.

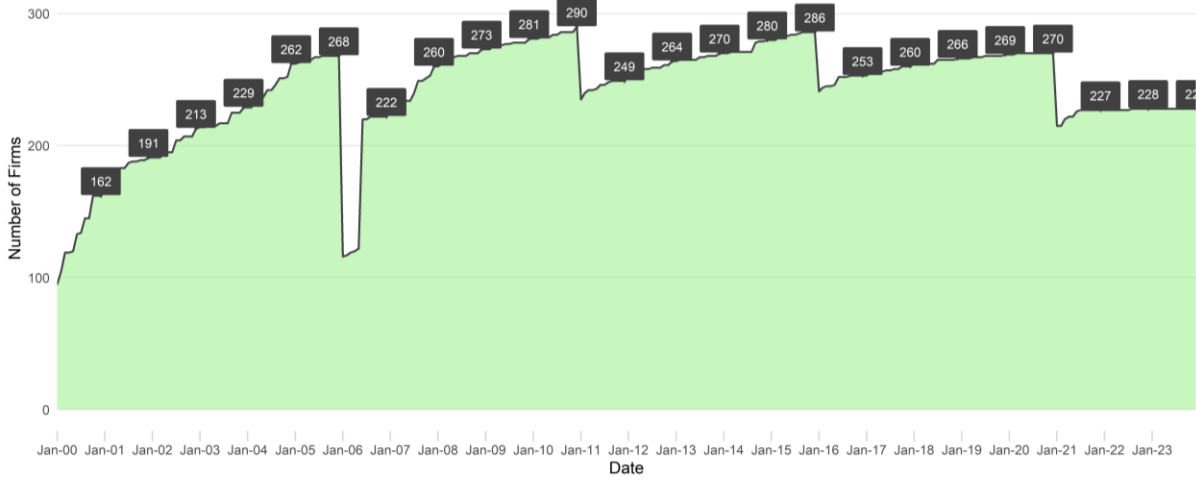


Figure 1 Number of active firms in the dataset

The figure above shows the number of firms for which data is available in each month. In December 2000, 162 firms of the 483 firms in the dataset have valid data for both volume and shares outstanding. In December 2023, the number of firms in the dataset is 223.

The position dummy is a dummy variable designed to account for the intuition that the effect of ranking deteriorates as one progresses down the list. Four position dummies are created. Position dummy 5 is set to either 1 if the company name is within the top 5% of names in a month or 0 if not. Position dummy 25 is equal to 1 if the company name is between 5% and 25% of firms listed alphabetically. Position dummy 50 is equal to 1 if the firm name is between 25% and 50%. Lastly, position dummy 75 is 1 if the company name is between 50% and 75%. Panel regressions are run on firm characteristics defined in Chapter 3.2, which is done to identify the relationship between a firm's position in an alphabetically sorted list and firm characteristics. Continuous position and positions dummies are also separately included as explanatory variables in the main regression model.

3.2. Dependent Variables

Following Jacobs and Hillert (2015), two measures of trading activity are employed, namely share turnover and liquidity. Share turnover is the ratio of volume of shares traded for company i at time t to the number of shares outstanding at time t :

$$Share\ Turnover_{i,t} = \frac{Shares\ Traded_{i,t}}{Shares\ Outstanding_{i,t}} \quad (1)$$

Share turnover tends to be excessively large for some firms, whilst smaller for others. This would make it difficult to capture the differences between firms. To avoid this, share turnover

is logged in regressions. To circumvent taking the log of zero, the turnover variable used for the regressions is the log of the sum of turnover and an arbitrary constant of 0.0000025. This follows Jacobs and Hillert (2015), but alternative constants such as 0.0001 and 0.001 were tested and yielded consistent results. The turnover variable is therefore:

$$Turnover_{i,t} = \log(Share\ Turnover_{i,t} + 0.0000025) \quad (2)$$

In addition to this definition of turnover, raw values of turnover are also used in robustness checks.

Jacobs and Hillert (2015) proxy for illiquidity using the definition of Amihud (2002). Amihud (2002) defines illiquidity as the annual average ratio of the absolute share return to the trading volume (in dollars). For the purposes of this study, the ratio is modified to a monthly measure and trading volume is quantified in Rands naturally, since this study uses South African data. The illiquidity measure as defined by Amihud (2002) is given by:

$$ILLIQUID_{i,t} = \frac{1}{D_{i,t}} \cdot \sum_{d=1}^{D_{i,t}} \frac{|R_{i,td}|}{VOLR_{i,td}} \quad (3)$$

Where $D_{i,t}$ is the total number of days with available data for share i in month t , $R_{i,td}$ is the return of share i on day d of month t and $VOLR_{i,td}$ is the volume in Rands on day d . This measure results in extremely small observations, typically less than 0.005, as shown in Figure 2 below.

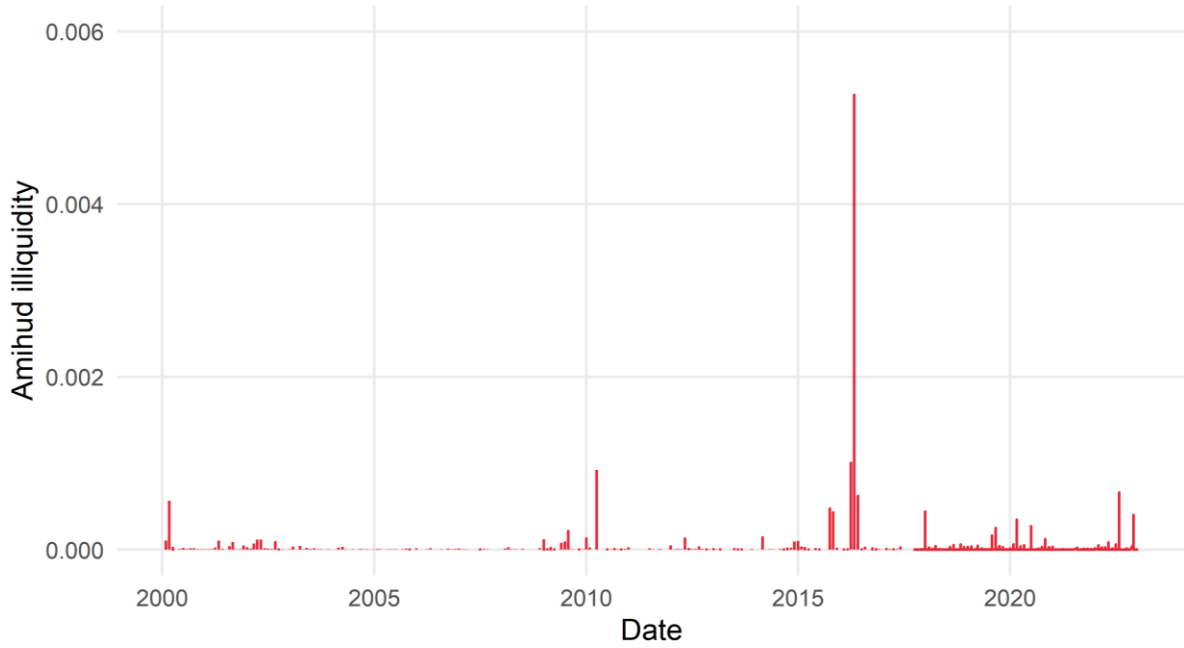


Figure 2 Amihud illiquidity

As a result, it may be difficult to capture the impact of much larger variables on illiquidity. To combat this, the illiquidity measure is logged in the main regressions performed. Equation 4 below provides the definition for the illiquidity dependent variable:

$$illiquidity_{i,t} = \log(ILLIQUID_{i,t}) \quad (4)$$

For robustness, the dependent variables are applied in two forms. The first is as defined in Equations 2 and 4 above. These two forms log the turnover and illiquidity measures. As stated by Jacobs and Hillert (2015), the log of zero is undefined. Therefore, cases where turnover is zero is dealt with by adding 0.0000025 prior to logging the variables. Jacobs and Hillert (2015) note that taking logs is beneficial as it reduces the effect of large outliers. The other two forms are the untransformed variables, as defined in Equations 1 and 3 above. Results using both forms of the response variables are reported.

3.3. Regression formula

As in Jacobs and Hillert (2015), trading activity is regressed on a vector of firm characteristics, including alphabetical share listing. The model used to test the effect of alphabetical ordering is taken directly from Jacobs and Hillert (2015) and is given below:

$$Trading_{i,t+1} = \gamma_{0,t} + \sum_{k=1}^n \gamma_{k,t} Firm\ Characteristic_{k,i,t} + \varepsilon_{i,t+1} \quad (5)$$

Where $Trading_{i,t+1}$ is the measure of trading activity, $Firm\ Characteristic_{k,i,t}$ is the k^{th} explanatory variable for share i in month t . As mentioned by Chordia *et al.*, (2007), the challenge with this approach that is difficult to interpret the results, since, unlike share returns, the dependent variable is not independently and identically distributed. To solve this problem, heteroskedasticity and autocorrelation-consistent (HAC) t -statistics are calculated using the Newey and West (1987) procedure.

The table below lists the variables to be included in the regression.

Table 1 List of variables

VARIABLE	DESCRIPTION	Role
Turnover	Turnover is calculated by dividing the number of shares traded by the number of shares outstanding for the month. Because turnover tends to be large, differences between firms may be difficult to capture. For this reason, turnover is logged.	Dependent variable – measure of trading activity
Amihud illiquidity	Annual average ratio of the absolute share returns to the trading volume	Dependent variable – measure of trading activity
Continuous position	A value between 0 and 1 based on the firm's relative position in an alphabetically ranked list of firms in month t .	Independent variable – measures a firm's position in an alphabetical list.
Positive return	Return value if return is positive, zero otherwise	Independent variable – captures the positive return for separate testing.
Negative return	Return value if return is negative, zero otherwise	Independent variable – captures the negative return for separate testing.
Book-to-market	Book-to-market ratio	Independent variable – a firm characteristic
Age	Logged value of market age. Market age is proxied as the number of months from the first observation in the price data obtained from Bloomberg.	Independent variable – a firm characteristic
Market capitalization	Logged value of market capitalization	Independent variable – a firm characteristic
Share price	Nominal share price	Independent variable – a firm characteristic
Number of analysts	Log of number of analysts covering a firm	Independent variable – a firm characteristic

Leverage	Firm total debt to total assets	Independent variable – a firm characteristic
Beta	Dimson (1979) betas	Independent variable – a firm characteristic
Earnings surprise	Difference between reported earnings per share in quarter t and the earnings from four quarters prior, scaled by the share price in quarter t .	Independent variable – a firm characteristic
Earnings volatility	Standard deviation of quarterly earnings in the last eight quarters.	Independent variable – a firm characteristic
Position dummy	A dummy variable which is equal to 1 if the firm is in the top 5, 25, 50 and 75 percent of firms ranked on alphabetical order and equal to 0 otherwise. Five dummy variables are constructed for each of 5%, 25%, 50% and 75%.	Dummy variable

The following section gives a detailed explanation of how these variables are constructed.

3.4. Explanatory Variables

As mentioned in the literature review, there are other factors related to firm visibility that influence the trading activity of a share. These variables are considered among the explanatory variables. The variables included in this section are largely based on these findings and their inclusion in the study by Chordia *et al.*, (2007). Firm-level characteristics were selected based on their relevance to visibility and investor attention.

Since many participants do not invest in the market portfolio, but rather buy individual shares, it is possible that they would need to sell shares based on the return of the stock. A positive or negative return could potentially trigger a shareholder to transact, thus impacting the trading activity. To control for this, two dummy variables are inserted. One is the value of the positive return of the share and zero if negative while the other has this reversed.

The book-to-market ratio is included to account for the observation that share turnover for growth shares is often greater than share turnover for value shares (Hong & Stein, 2007). Additionally, Chordia *et al.*, (2007) include the book-to-market ratio as a measure of the visibility of the firm. This is because growth shares are more likely to be visible over value shares.

The book-to-market ratio is defined as the accounting value of common equity divided by market value of common equity (Fama & French, 1992). Following Fama and French (1992), the book-to-market ratio is every 12 months at the end of December. This aligns with the

release of most firms' year-end financial information. The study excludes shares with negative book values. The book-to-market ratio is winsorised at the 99% level.

Included in the model are logged values of market age, firm size, and nominal share price. These values are potentially correlated, therefore alternative models with just one and two of the variables included are also run. Firm size is proxied by the log of market capitalisation in ZAR, computed in month $t-1$. Firm size is included in the model of Chordia *et al.*, (2007) because of its relationship with trading volume, and therefore included in this study for the same reason.

Market age is the log of the firm's age in months, also computed in month $t-1$, where age is the number of months from IPO date. Chordia *et al.*, (2007) posit that age is negatively related to firm visibility. The authors theorise that with recent IPOs are more interesting to investors who seek high returns from the publicity from the IPO. Therefore, market age is included as a measure of firm visibility.

The date of IPO for every share in the South African data obtained is not readily available. As a result, market age is proxied as the number of months from the first observation in the price data obtained from Bloomberg.

Nominal share price is the log of the absolute value of the share price in month $t-1$, shown in Equation 6 below:

$$PRICE_{t-1} = \log(Share Price_{t-1}) \quad (6)$$

Chordia *et al.*, (2007) employ the log of the firm's share price to proxy for the firm's visibility. The reason for the inclusion of price is due to the finding of Brennan and Hughes (1991) that brokers are likely to advertise firms with a high share price since price and commissions are inversely related.

Another proxy for firm visibility used by Chordia *et al.*, (2007) is the level of information dissemination and the quantity of informed traders for a specific share. Analysts will often have different estimates for earnings, referred to as forecast dispersion, which can motivate investors to trade. Similarly, the level of leverage held by a firm might be analysed differently depending on the individual leading to investors trading a share for different reasons. As in Chordia *et al.*, (2007), the number of analysts covering a firm is used to proxy for trading activity arising from

informed traders. As discussed in Section 2, analyst coverage is linked to the production of information, which in turn leads to informed trading. This study includes analyst coverage as a variable for the same reason.

Analyst coverage is calculated as:

$$Analyst\ Coverage_{t-1} = \log(1 + analyst_{t-1}) \quad (7)$$

Where $analyst_{t-1}$ is the number of analysts covering a company in month $t-1$. Analysts typically release estimates quarterly or semi-annually to align with the release of financial performance by a firm. The data obtained from Bloomberg is annual. The annual observation is carried forward to the following 12 months as a monthly observation to align with the monthly data for regressions. Firms with missing data are not excluded but rather treated as not having analyst coverage.

Related to analyst coverage is forecast dispersion, which is a measure of the differences in opinion of informed traders. As mentioned in the literature review, difference of opinion has been shown to have an impact on trading activity. Chordia *et al.*, (2007) use forecast dispersion to a measure of the difference in opinion. Likewise, this study employs forecast dispersion as a measure of the difference in opinion.

Forecast dispersion is the standard deviation of forecasted earnings per share scaled by the absolute value of the average forecast. If there are fewer than two analysts providing estimates for a company in a month, then the forecast dispersion is zero. Missing values are filled with the most recent value. Forecast dispersion is included in the regression as a lagged variable (one month, that is, $t-1$) and is winsorised at the 99% level.

According to Chordia *et al.*, (2007), market participants perceive firms with high debt levels as being riskier than firms with low amounts of debt. This is due to the higher probability of default and the associated costs bankruptcy costs. Chordia *et al.*, (2007) also note that agency costs are higher for debt laden companies as owners take on a lottery mindset and prefer to take on extremely risky projects in attempts to rescue their equity. As a result, the dispersion in opinions among high debt firms would be wider than for low debt firms. Therefore, leverage proxies for differences in opinion.

Leverage is defined by the firm's debt as per the company annual financial statements, scaled

by total assets. For this study, firm debt is the total of current liabilities and long-term debt. As with book value, leverage is computed once every 12 months at the end of December. Leverage is also winsorised at the 99% level.

As explained by Chordia *et al.*, (2007), estimation uncertainty of fundamental value can trigger trading activity due to new information that requires investors to update their positions. To control for this, lagged values of beta, earnings surprises and earnings volatility are included as variables in the regression. Betas are calculated using Dimson (1979) methodology, which is done to factor in the fact that shares do not trade synchronously. Dimson (1979) proposes the approach below:

$$R_t = \alpha + \sum_{i=-k}^k \beta_i R_{m,t+i} + \varepsilon_i \quad (8)$$

Where k is the number of lags.

Beta is then given by:

$$\beta = \sum_{i=-k}^k \beta_i \quad (9)$$

Using a value of 1 for k gives:

$$R_t = \alpha + \beta^{-1} R_{m,t-1} + \beta^0 R_{m,t} + \beta^{+1} R_{m,t+1} + \varepsilon_t \quad (10)$$

Beta is then calculated using:

$$\beta = \alpha + \beta^{-1} + \beta^0 + \beta^{+1} + \varepsilon_t \quad (11)$$

This means regressing the excess return for each individual firm excess return on the contemporaneous excess market return, the lagged market return, and the lead excess market return. This is done on a rolling three-year window, effectively providing three-year betas. The three-year window used to calculate rolling beta begins in January 2000.

In addition to Dimson betas, Jacobs and Hillert (2015) consider earnings surprise as a proxy for the difficulty of estimating fundamental value. Earnings surprise is the difference between reported earnings per share in quarter t and the earnings from four quarters prior. This value is scaled by the share price in quarter t . Since there earnings are only announced quarterly and the regressions are monthly, there's a mismatch in timing. In this case, the most recent value is

used for the regression. However, the maximum allowable mismatch is set to 125 days. The earnings surprises are winsorised at the 99% level.

Furthermore, Jacobs and Hillert (2015) utilise earnings volatility to proxy for the uncertainty in estimation of fundamental value. Earnings volatility is defined as the standard deviation of quarterly earnings in the last eight quarters. The last available lagged data point is the value inputted into the regression. For this measure, the maximum number of days between the earnings report date and the date of the regression should be 250 days. The earnings volatility is winsorised at the 99% level.

3.5. Rank effects on firms with lower visibility

The effect of higher ranking in an alphabetically sorted list might be higher for firms that are otherwise less visible. To test for this, market capitalisation, residual analyst coverage, and lagged market share are used as proxies for firm visibility. Residual analyst coverage is taken from the residuals produced by cross-sectional regressions in the form of:

$$\log(1 + \text{number of analysts providing fiscal year one earnings estimates in } t - 1) = \log(\text{firm size in } t - 1) \quad (12)$$

The sample is split into visible and less visible firms using the median of the market capitalisation, residual analyst coverage and market share. These three separate measures give three different classifications of low and high visibility firms. The main model is then be run on each grouping.

3.6. Robustness Checks

To ensure results are robust, alternative model specifications are used. Firstly, alternative definitions for illiquidity and turnover are used. These are Rand share volume, rate of zero trading days, and number of trades. Rand share volume is defined as:

$$\text{Rand Share Volume}_t = \log(\text{number of shares traded} \cdot \text{share price} + 0.0000025) \quad (13)$$

The rate of zero trading days is simply the number of days with a return of zero divided by the number of trading days in that month. This requires that a firm has a minimum of 10 days of non-zero return days in a month. The reason for this is to ensure that there is a consistent minimum level of trading activity in order for the calculations of turnover and illiquidity to be

accurate. This approach follows the logic used by Amihud (2002) in measuring illiquidity in emerging markets.

The number of trades is defined as:

$$\text{Number of trades}_t = \log(1 + \text{trades in a month} + 0.0000025) \quad (14)$$

Additionally, the number values of turnover and illiquidity measures are used instead of the logged values, which are winsorised at the 95% level. Another model specification used is yearly data for the regression, rather than monthly.

Shares may sometimes be sorted by share symbol rather than by the firm name. To account for this, the list of firm names is replaced by the list of share symbols.

3.7. Regression model

The data structure of the dataset used for the model is panel data. This means that in each month, there are observations on each variable for each firm. Testing the research question requires determining differences in turnover and illiquidity between each firm in each month. The methodology that will accurately test for this difference in firms across time is a panel regression. Where dependent variables are logged, coefficient estimates are interpreted as percentage changes in the original variable. Variables that have frequencies below a month, such as analyst coverage or book-to-market ratios, are carried forward until the next updated observation in the data.

Given that the study uses panel data and hence panel regressions, there are multiple methods that are available. This study focuses primarily on two methods for data analysis. The first is the fixed effects or random effects model. The choice between the two is made by running the Hausman test. The second model is the Generalised Method of Moments (GMM). The use of both models is motivated by the multicollinearity concerns from variables that are known to be correlated. The GMM model deals with multicollinearity better than the fixed effects or random effects models.

3.8. Data

The data for the regressions was obtained from Bloomberg and IRESS. The variables used in the regression are dependent on data availability. Bloomberg is the source for data on share price, trading volume, number of shares outstanding, market capitalisation, IPO date, earnings

per share, analyst's earnings estimates, industry, and market share. IRESS is the source for analyst coverage and accounting data such as total debt, current liabilities, total debt.

Table 2 below gives a description of the data collected for the study. The number of firms in the dataset is 230 covering a period from January 2000 to December 2023. This period is chosen as data is readily available on Bloomberg for this period.

Daily observations were collected for price, trading volume, number of shares outstanding, market capitalisation, and earnings per share. Only annual observations were available for number of analysts covering a firm.

Some of the data is not readily available. For instance, there is no list of firms on the JSE in each month, a key variable. A workaround is used, which is to use the names of each firm in the dataset for that month. The workaround is described in the column titled "alternative".

Table 2 Data obtained for the study

Dataset	Source	Alternative	Period	Frequency
<i>Share prices</i>	Bloomberg	N/A	January 2000 to December 2023	Daily
<i>Market capitalisation</i>	Bloomberg	N/A	January 2000 to December 2023	Daily
<i>Trading volume</i>	Bloomberg	N/A	January 2000 to December 2023	Daily
<i>Price-to-book ratio</i>	Bloomberg	N/A	January 2000 to December 2023	Daily
<i>Earnings per share</i>	Bloomberg	N/A	January 2000 to December 2023	Quarterly
<i>Shares outstanding</i>	Bloomberg	N/A	January 2000 to December 2023	Daily
<i>Total assets</i>	Bloomberg	N/A	January 2000 to December 2023	Daily
<i>Short- and long-term debt</i>	Bloomberg	N/A	January 2000 to December 2023	Daily
<i>Short-term borrowings</i>	Bloomberg	N/A	January 2000 to December 2023	Daily

<i>Long-term borrowings</i>	Bloomberg	N/A	January 2000 to December 2023	Daily
<i>Total-debt-to-total-assets</i>	Bloomberg	N/A	January 2000 to December 2023	Daily
<i>Number of analysts</i>	IRESS with many missing entries	N/A	January 2000 to December 2023	Annual
<i>IPO Date</i>	Bloomberg with many missing entries	Date of first non-blank observation in price dataset	N/A	N/A
<i>Firm names</i>	Not available	List of firms with non-blank price data in the month	January 2000 to December 2023	Monthly
<i>Market age</i>	Not available	Number of months since first non-blank observation in price data.	January 2000 to December 2023	Monthly
<i>Name changes</i>	Not available	None – excluded	N/A	N/A
<i>Advertising spend</i>	Not available	None – excluded	N/A	N/A
<i>JSE Top 40 constituents</i>	Monthly historical data not available	Top 40 shares by market capitalisation in each month	January 2000 to December 2022	Monthly
<i>Risk-free rate</i>	SARB	N/A	January 2000 to December 2023	Monthly

3.9. Tools

The research tests are implemented using the free statistical software R. The research tests are designed in R with roughly 1600 lines of code. All of the output presented in Chapter 4, including charts and tables, is produced from R code written by the author of this study. The R code and data used for the study are available on request.

4. Results

This section presents the results of the research tests, beginning with pre-regression testing in Section 4.1 and 4.2. After that, the main results are presented in two parts, with the first part of being the results of the regressions with turnover as the measure of trading activity, presented in Section 4.3. The second set of results are presented in Section 4.4, which are results from regression models with illiquidity as the dependent variable. After the results are presented, the section closes out with a conclusion on the hypotheses set out in Chapter 1.

4.1. Descriptive statistics

The graph below shows the progression of the monthly average of logged share volume of all shares on the JSE from 2000 to 2023. As the graph shows, share volume has increased steadily over time. There is a drop-off in 2021, which can be attributed to the COVID-19 crisis.

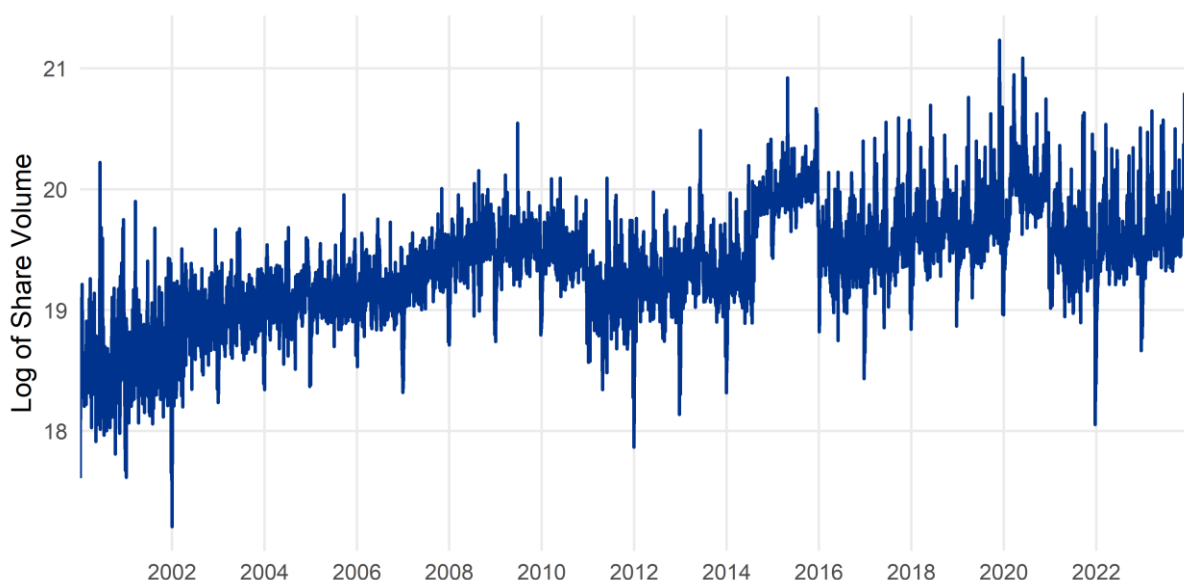


Figure 3 Log of share volume over time

However, share turnover depicts a different picture, as can be seen in Figure 4 below, which plots the monthly average of logged share turnover for all JSE listed shares across time. The graph shows that share turnover has trended downwards over time on the JSE. Share turnover is approximately 9.5 in the beginning of the millennium but drops significantly to 8 between 2006 and 2008, which roughly corresponds to the global financial crisis. Similarly, share turnover drops below 8 in 2020 during the COVID-19 crisis.

Since the COVID-19 crisis, share turnover continues to decrease, reaching a low of nearly 7 around 2023. This cannot be due to reduced trading activity as share volume has increased

during the same period. The potential explanation for this is the increase in institutional investors, who tend to trade less frequently but will trade large amounts when they do. In addition, institutional investors hold more companies in their portfolios than retail investors, which means that even though they buy and hold, the high number of companies in their portfolios would create trading opportunities through portfolio rebalancing. This explanation is consistent with increased trading volume and lower share turnover.

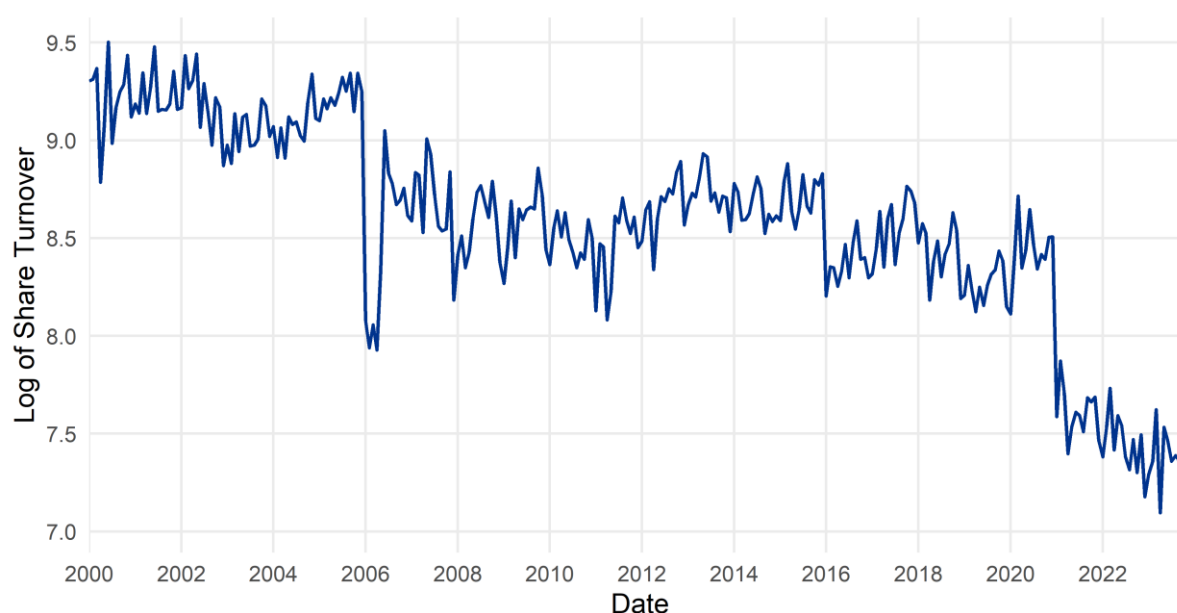


Figure 4 Log of share turnover over time

4.2. Pre-regression testing

Prior to running the regressions, data checks are necessary to inform the types of regressions to be run, what variables to include and whether the results of the regressions can be interpreted accurately.

4.2.1. Correlation

The first data check is a test for correlation between the explanatory variables. Several of the variables in Table 3 have been known to be correlated with one another. For instance, share price and market capitalisation are correlated, since market capitalisation is constructed using the share price and number of shares. Another less obvious example is market age and share price, since the older the firm is, the more time it has had to build up its share price.

Highly correlated variables present a challenge because independent variables should not be affected by other explanatory variables that have been included in the model. Otherwise, it is

difficult to interpret results of significant variables that are highly correlated. Figure 5 shows the correlation between the explanatory variables included in Table 3.

Figure 5 shows that market capitalisation, share price and number of analysts are all highly correlated with each other. Market age is also correlated with the variables, albeit with slightly weaker correlation. In addition, market age is correlated with number of analysts, by a factor of 0.11. This is intuitive, as older firms are more likely to have analysts covering due to the firm building up its profile over time.

According to loss aversion theory, losses have a greater impact on investors than wins. Therefore, negative return should have a greater impact on trading activity. To capture these potential asymmetries in investor reaction to positive and negative returns, the monthly return is split into positive return and negative return. The positive return variable retains the observed return value if the return is greater than zero and is set to zero otherwise. The negative return variable retains the return value if it is negative and is set to zero otherwise.

The positive and negative return variables are not expected to be highly correlated, as they are mutually exclusive by definition. In a particular month, only one of the variables can be non-zero. In the event of a zero return, both variables will take on the value of zero, which is likely the source of the small positive correlation of 0.11. With this approach, the regression is able to separate the effect of positive and negative returns on trading activity.

The key takeaway from the correlations is that price, age, size, and earnings ideally should not all be present in the same model. The other problem presented by high correlations is multicollinearity. To circumvent this, alternate regressions that exclude these variables are performed.

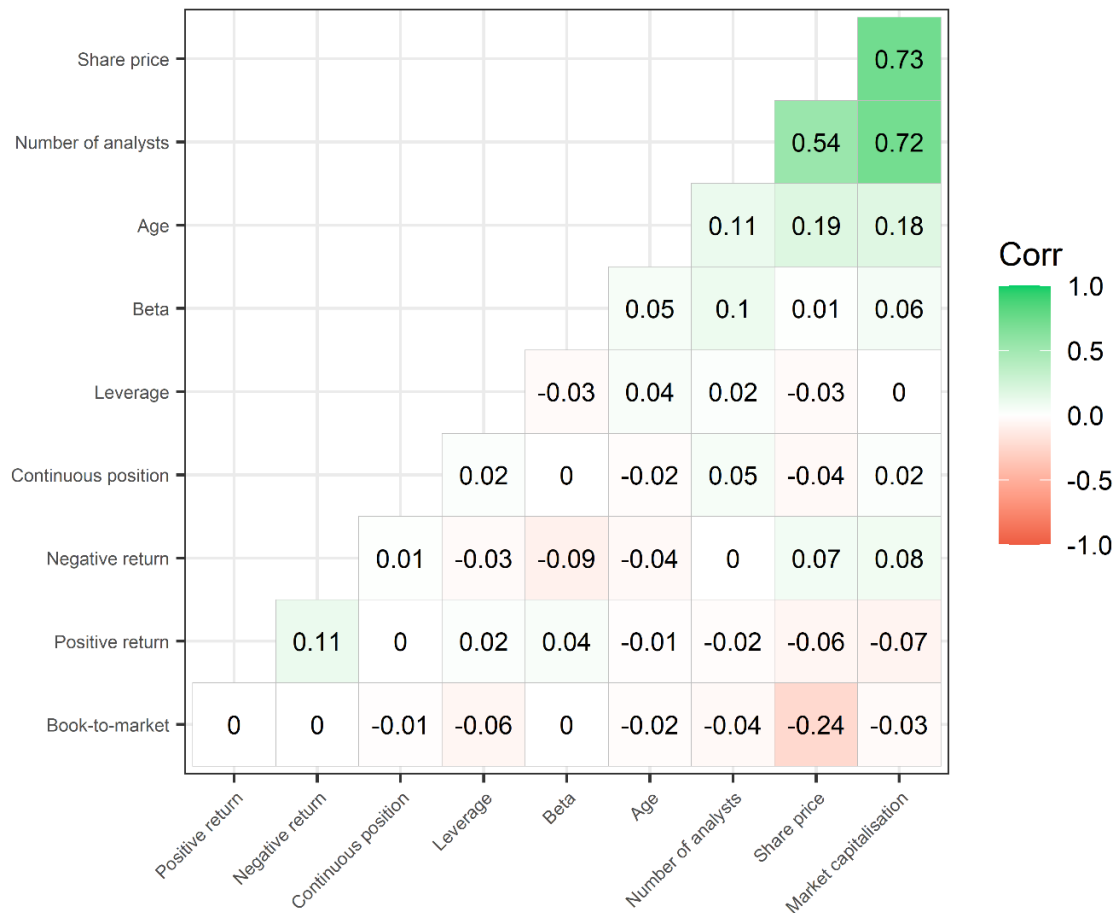


Figure 5 Correlation matrix plot visualising the correlation between independent variables – generated with R code

In addition to correlations between independent variables, the correlation between the dependent variables and independent variables is performed. Table 3 below shows the correlation between turnover, illiquidity and the independent variables.

The table below shows that market capitalisation, share price and number of analysts are highly correlated with trading activity. Larger firms tend to attract more attention from investors. The high correlations means that inclusion of these variables in one regression could lead to a spurious regression. Therefore, this simply means it is not prudent to put these highly correlated variables in the same regression equation. For this reason, the variables must be separated, and hence multiple regression models are run.

Table 3 Correlation between dependent and independent variables

Variable	Turnover	Illiquidity
Continuous position	-0.00303	0.006557
Positive return	-0.01737	0.101382
Negative return	-0.04474	-0.09451

Variable	Turnover	Illiquidity
Book-to-market	-0.0027	0.253137
Market age	0.048629	-0.00604
Firm size	0.356323	-0.37935
Share price	0.325221	-0.38948
Leverage	0.002956	-0.00119
Beta	0.074853	-0.03638
Number of analysts	0.311854	-0.21855

The figure below shows a plot of turnover against the four variables with the highest correlation with turnover. The plot shows that an increase in share price and market capitalisation is associated with an increase in turnover. However, the positive relationship is not completely strong as the points are scattered around the regression line rather than being perfectly plotted along the line. This suggests that there are other aspects affecting the variation in turnover.

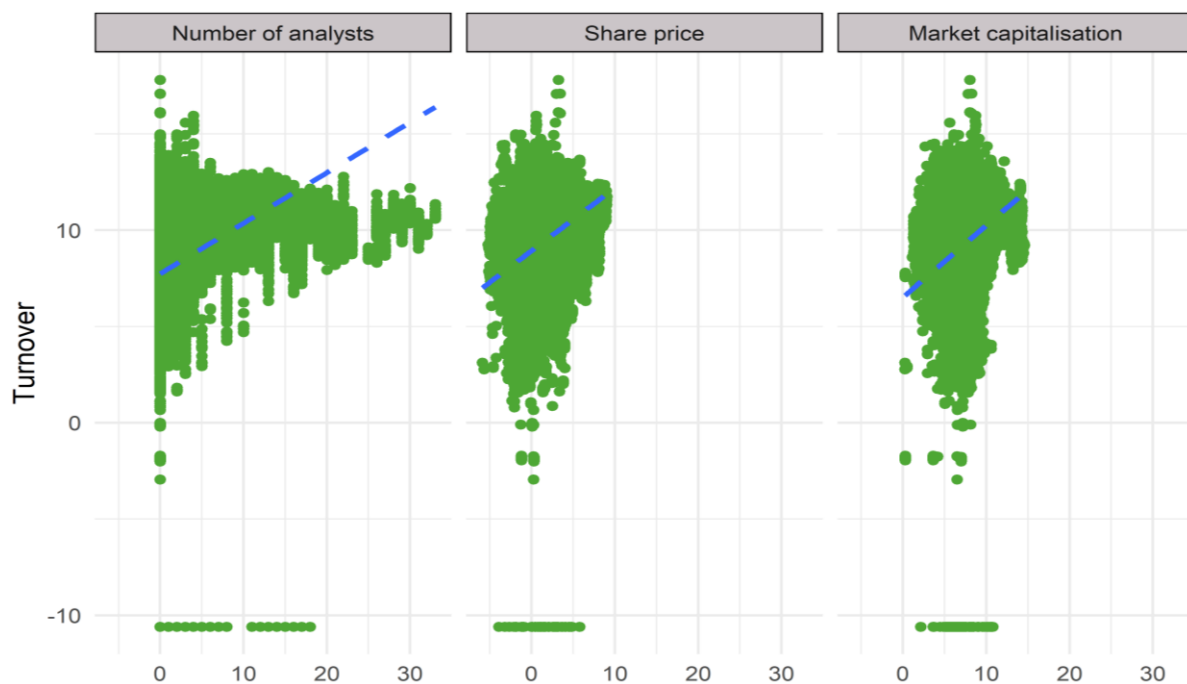


Figure 6 Plot of key independent variables

4.2.2. Multicollinearity

To ensure robustness, variance inflation factors (VIF) were computed before the main regression analysis. No variable exceeded a VIF of 5, indicating multicollinearity was not a concern.

As was stated in 4.1.1, some of the independent variables are highly correlated with each other, which may potentially give rise to multicollinearity. For instance, from Figure 5 share price is highly correlated with three variables and is therefore likely to exhibit multicollinearity. The danger of multicollinearity is that the statistical significance of the estimated coefficient is less reliable. The significance may be due to the correlation with another included variable, which should not be the case for an independent variable.

As a result, it is important to check for multicollinearity in the main regression model. The test performed is known as the Variance Inflation Factor (VIF) test. The rule of thumb is that a value between 1 and 5 is acceptable. Coefficients with values greater than 5 should be excluded from the model.

Table 4 shows the results of the VIF test performed on the main model presented in Table 5. As expected, share price has a high value, its VIF is 2.465. This VIF is lower than the threshold for multicollinearity, which is set at 5. Therefore, the model selected appears to not suffer from multicollinearity. This could be because market capitalisation is the only variable to have a correlation coefficient greater than 0.7 with more than one variable (share price and number of analysts). Even so, market capitalisation (firm size in the table below) has a VIF of 3.35, which is below the value of 5. All the variables included are within the generally accepted limits. Based on this it seems unlikely that the model selected can be improved upon. Nonetheless, models with some of the variables are reported on in the following section.

Table 4 Multicollinearity

Variable	VIF
Continuous position	1.010422
Positive return	1.024548
Negative return	1.042242
Book-to-market	1.118909
Share price	2.465006
Firm size	3.351571
Market age	1.053196
Leverage	1.012911
Beta	1.027265
Number of analysts	2.097

4.3. Turnover

4.3.1. Turnover and continuous position

Table 5 shows the results of univariate regression of turnover on continuous position variable, without any control variables included. The number to the right of the variable is the coefficient.

The level of statistical significance is based on t-statistics calculated using the Newey and West (1987) method. As explained in Chapter 3, these t-statistics are used because the dependent variables are non-stationary, which makes interpretation of the results complicated.

Continuous position is the main variable of interest as it represents the alphabetical ranking of firms. By itself, continuous position has a significant relationship with turnover, with the coefficient being statistically significant at the 1 percent level.

The R squared value of the model represented in Table 4 above is 0.01, while the adjusted R-squared value is 0.006.

Table 5 Regression of logged share turnover on continuous position

Variable	Statistic	p-value
Continuous position	0.302	<0.001
Model information		
Num of Observations	69684	
R ²	0.010	
R2 Adj.	0.006	
R2 Within	0.000	
R2 Within Adj.	0.000	
AIC	421228.4	
BIC	423873.2	
RMSE	4.95	

Figure 7 illustrates this quite well. The figure is a plot of turnover against continuous position. If the hypothesised relationship were true, the expectation would be that as continuous position increases from 0 to 1, turnover would steadily decrease. However, as the plot shows, there is no increase or decrease as continuous position increases. The dashed line shows the fitted linear model. The line appears to be slightly upwards sloping, indicating a slight positive relationship, as seen from the model output.

The significance of continuous position in the model may be due to omitted variables. There are several firm characteristics that have been known to have an impact of the number of shares traded such as firm size and market age to name a few. As explained in Chapter 3, these variables are included to capture the effect that these variables would have on trading activity.

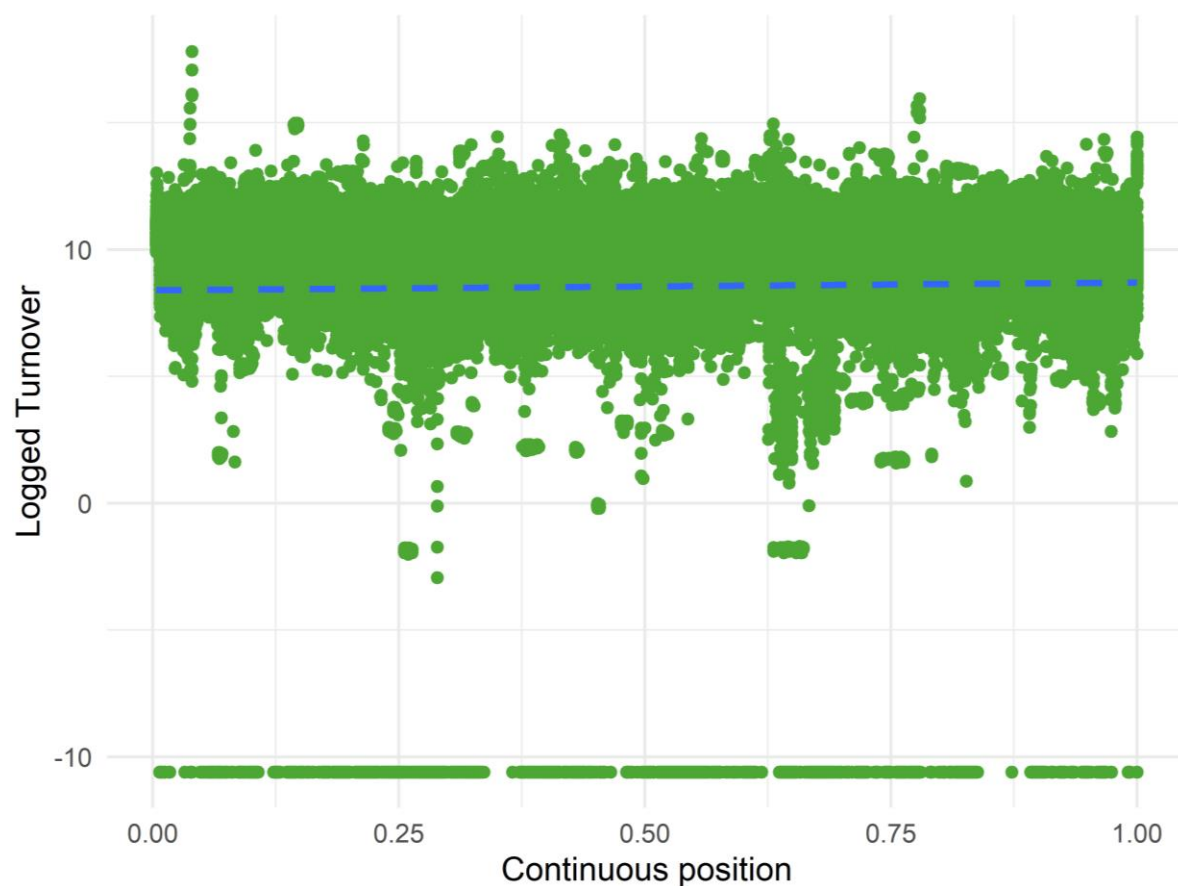


Figure 7 Plot of logged turnover against continuous position

4.3.2. Turnover and firm characteristics

Table 6 presents the results of multivariate panel regressions of logged turnover on continuous position as the main variable of interest and eleven control variables, all of which are lagged by a minimum of a month, taken from various firm characteristics. The reason for lagging the firm characteristics is that investors trade on past information, that is in month t investors would be using information from month $t-1$. These control variables are described in Chapter 3 and Table 1. In the regressions reported in this section, the main independent variable that is of interest to this study is continuous position, as it directly answers the research question. However, other known influencers of trading activity are also of interest as this has not been

sufficiently addressed in the South African literature. Table 6 is thus the key finding of this study. The period of the data is from January 2000 to December 2023.

Table 6 Regression of share turnover on continuous position and control variables

Variable	Statistic	p-value
Continuous position	-0.029	0.278
Positive return	0.161	0.238
Negative return	-2.526	<0.001
Book-to-market	0.001	<0.001
Market age	0.001	<0.001
Share price	0.176	<0.001
Firm size	0.196	<0.001
Leverage	0.002	0.033
Beta	0.085	<0.001
Number of analysts	0.043	<0.001
<i>Model information</i>		
Number of observations	57601	
R ²	0.162	
R ² Adj.	0.158	
R ² Within	0.151	
R ² Within Adj.	0.150	
AIC	243748.4	
BIC	246096.2	
RMSE	2.00	

The results presented in this section suggests that alphabetic position of shares does not impact trading activity on the JSE. This conclusion can be made based on the improved model fit when one considers the much higher r-squared value and much lower AIC. This means that the added variables explain the variation in logged turnover better than the single model of continuous position. Essentially, the result obtained in Table 4 could potentially have occurred by chance.

4.3.2. Alternative regressions

Table 7 below shows results of multivariate regressions that exclude firm size and market age, both of which are correlated with the share price. This is rational considering that firms that

have been listed for longer would be more likely to have grown over time. There is also likely to be a bias here as older firms that are still listed are likely to be large, stable mature firms.

Removing market age and firm size from the model does not impact the statistical significance of the continuous position variable. Negative return, book-to-market, share price, market beta and number of analysts remain statistically significant at the one percent level.

The r-squared value drops to 14.7% from 16.1% and the AIC of the model is 246246.8, which is up from 244505.1 in the main model. This indicates that the main model reported in Table 5 is has a better fit on the data.

Table 7 Regression with share price

Variable	Statistic	p-value
Continuous position	-0.015	0.584
Positive return	0.075	0.500
Negative return	-2.080	<0.001
Book-to-market	0.002	<0.001
Share price	0.275	<0.001
Leverage	0.001	0.052
Beta	0.087	<0.001
Number of analysts	0.080	<0.001
<i>Model information</i>		
Number of Observations	57816	
R ²	0.148	
R ² Adj.	0.144	
R ² Within	0.137	
R ² Within Adj.	0.137	
AIC	245503.4	
BIC	247834.3	
RMSE	2.01	

The table below shows the VIF for the coefficients in the updated model. The VIFs are all less than 2, indicating that multicollinearity is not an issue with this model.

Table 8 Multicollinearity of updated model

Variable	VIF
Continuous position	1.009477
Positive return	1.020366
Negative return	1.03104
Book-to-market	1.077342
Share price	1.55346

Leverage	1.009622
Beta	1.026113
Number of analysts	1.47247

Table 9 reports results from regressions of continuous position on positive return, negative return, book-to-market, market age, leverage, market beta and number of analysts. This is the model differs from the model in Table 8 in that it includes market age over share price.

The result of this model is not different from the results of Table 7 as all the same variables are statistically significant at the same levels of significance. As before, continuous position is not statistically significant, and neither is positive return.

The r-squared value has reduced from 16.1% in the main model and 14.7% in the model with share price to 10.8% in this model. The AIC has also increased to 285059.9, which is higher than the previous two models. This indicates that model fit is not as good as the model fit of the previous two models. This may suggest that share price accounts for more variation in logged turnover than market age.

Table 9 Regression with market age

Variable	Statistic	p-value
Continuous position	-0.015	0.664
Position return	0.235	0.198
Negative return	-2.437	<0.001
Book-to-market	0.000	<0.001
Market age	0.001	<0.001
Leverage	-0.002	0.034
Beta	0.165	<0.001
Number of Analysts	0.163	<0.001
<i>Model information</i>		
Number of observations.	58368	
R ²	0.108	
R ² Adj.	0.104	
R ² Within	0.084	
R ² Within Adj.	0.084	
AIC	284664.6	

BIC	286998.0
RMSE	2.76

The model below is a regression of continuous position on the independent variables excluding share price and market age. This model, with firm size only, has a statistically significant coefficient, which is significant at the five percent level. The coefficient has the correct hypothesised sign.

The sign of the coefficient means that turnover reduces as continuous position increases, which indicates that there is an alphabetic bias on the JSE. However, given that this can only be determined once share price and market age are excluded from the model, share price and market age have a greater impact.

The r-squared value of this model is 14.3% is lower than the main model, and the AIC is higher, which indicates that the model fit is not as good as the main model. Based on this, the main model should be the one to derive conclusions from. However, given that the results are similar with this model, it can be used to validate the results.

Table 10 Regression with firm size

Variable	Statistic	p-value
Continuous position	-0.065	0.016
Positive return	0.181	0.222
Negative return	-2.589	<0.001
Book-to-market	0.000	<0.001
Firm size	0.313	<0.001
Leverage	0.000	0.726
Beta	0.081	<0.001
Number of analysts	0.048	<0.001
<i>Model information</i>		
Number of observations	57660	
R ²	0.144	
R ² Adj.	0.140	
R ² Within	0.133	
R ² Within Adj.	0.133	

AIC	250004.3	
BIC	252334.5	
RMSE	2.11	

Interesting to note at this juncture is that when turnover is the dependent variable, negative return has been consistently statistically significant in the regressions while positive return has not. The coefficient on negative return has a negative sign, which indicates an inverse relationship between negative return and trading activity. What this means is that when returns are negative, trading activity increases, which seems to suggest increased selloffs during market downturns. An interesting result here is that positive return is not statistically significant, which means that markets react stronger to negative returns than they do to positive returns.

This result is in line with prospect theory put forward by Kahneman and Tversky (1979), which states that wins and losses are evaluated differently by humans. The pain from losing is greater than the joy from winning, as a result, humans irrationally go to great lengths to avoid losing. Conversely, with liquidity both positive return and negative return are statistically significant across the models. This suggests that positive returns and negative returns both impact liquidity but only negative returns impact trading activity.

For the turnover models, the number of analysts is statistically significant across all models. The coefficient is positive, indicating that turnover increases as number of analysts increases. This result matches the hypothesis, that having analysts covering the firm increases visibility of the firm, which results in greater trading activity. For the liquidity models, the coefficients are all negative and statistically significant. This means that liquidity reduces as liquidity reduces, which is the expected result.

4.3.3. Regressions with earnings data

In this section, results with earnings surprise included as a variable are reported. It should be noted that earnings data starts from 2003 and runs to 2022 and is annual. This is because quarterly data could not be sourced. Therefore, the regressions are performed on annual data. Furthermore, not all firms have earnings data available in each period.

Despite the data issues, earnings surprise turns out to be significant in the regressions. The coefficient on earnings surprise is positive and statistically significant at the one percent level.

This is in line with the intuition that a surprise in the earnings earned by a company will trigger increased trading activity. However, the coefficient is small, suggesting that the economic impact of earnings surprise is small.

Table 11 Regressions with earnings data

Variable	Statistic	p-value
Continuous position	0.215	0.044
Positive return	0.284	0.010
Negative return	-0.998	<0.001
Book-to-market	0.001	0.032
Market age	0.019	(0.071)
Share price	0.142	0.011
Firm Size	0.230	<0.001
Leverage	0.001	0.504
Beta	0.000	0.823
Earnings surprise	0.000	<0.001
<i>Model information</i>		
Number of Observations	2581	
R ²	0.200	
R ² Adj.	0.192	
R ² Within	0.191	
R ² Within Adj.	0.188	
AIC	9057.0	
BIC	9203.4	
RMSE	1.39	
Std. Errors	by: Time	
FE: Time	X	

After running multiple regressions, continuous position is still not statistically significant. However, share price, market age, market age, book-to-market, positive return, negative return and earnings surprise consistently remain statistically significant. These results agree with the intuition that investors do not typically select shares purely on rational reasons but apply biases, such as selecting shares that are the most visible to them. On the JSE, that would be large firms with high share prices and a several analysts following them.

4.3.4. Robustness checks

This section reports results from robustness checks described in Section 3.5. Table 12 shows the findings from equation 13. In this model, logged turnover is replaced with the log of the rand value of total share volume. This is done to determine if the alternative use of trading activity influences the result. The finding here is that the p-value of the coefficient on continuous position is 0.226, which is above five percent. Therefore, continuous position is not statistically significant at the five percent level.

The r-squared value of the single variable model is low at just 4.1%, indicating that this model explains very little of the variation in rand share volume.

Table 12 Regression of logged rand share volume on continuous position

Variable	Statistic	p-value
Continuous position	0.030	0.225
<i>Model information</i>		
Number of Observations	65979	
R ²	0.041	
R ² Adj.	0.037	
R ² Within	0.000	
R ² Within Adj.	0.000	
AIC	356630.0	
BIC	359259.1	
RMSE	3.59	

Table 13 shows the results of the regression of rand share volume on continuous position. Just as before, continuous position is not statistically significant. However, the following variables yield statistically significant coefficients: negative return, book-to-market, firm size, market age, share price, market beta and number of analysts. As was the case in previous models, positive return and leverage are the only two control variables that are not statistically significant.

Table 13 Multivariate regression of rand share volume on continuous position

Variable	Statistic	p-value
Continuous position	-0.041	0.156
Positive return	0.157	0.256
Negative return	-2.792	<0.001
Book-to-market	-0.003	<0.001

Market age	0.000	<0.001
Firm size	0.973	<0.001
Share price	0.437	<0.001
Leverage	0.002	0.071
Beta	0.096	<0.001
Number of analysts	0.050	<0.001
<i>Model information</i>		
Number of Observations	57551	
R ²	0.620	
R ² Adj.	0.618	
R ² Within	0.608	
R ² Within Adj.	0.608	
AIC	258135.8	
BIC	260483.4	
RMSE	2.27	

Table 14 reports results when the log of total trading volume is used as the dependent variable. In this specification, continuous position is statistically significant at the one percent level, with the coefficient being negative, as hypothesised. This indicates that trading volume decreases as the one goes down the list of alphabetically sorted shares.

Table 14 Regression of logged trading volume on continuous position

Variable	Statistic	p-value
Continuous position	-0.126	<0.001
Positive return	-0.120	0.112
Negative return	-1.043	<0.001
Book-to-market	-0.005	<0.001
Market age	0.003	<0.001
Firm size	1.081	<0.001
Share price	-0.718	<0.001
Leverage	0.004	<0.001
Beta	0.053	<0.001
Number of analysts	0.085	<0.001
<i>Model information</i>		
Number of Observations	57293	
R ²	0.452	

R ² Adj.	0.450
R ² Within	0.433
R ² Within Adj.	0.433
AIC	244161.7
BIC	246508.2
RMSE	2.03

Table 15 shows the results from tests where turnover is regressed on position dummies. The coefficients and corresponding p-values are displayed in the table. As a reminder, position dummy 5 is assigned one where the firm's name places in the top 5% of shares, or assigned zero if it is not. Similarly, position dummy 25 is one if the firm's name falls in the range of 5% to 25% of firms when alphabetically sorted. If the firm name falls outside this range, the value will be zero. In the same vein, position dummy 50 is set to one if the firm's name lies within 25 and 50 percent of firms sorted alphabetically. Similar logic applies for position dummy 75.

The results from Table 15 are not consistent with the hypothesis that turnover reduces as one traverses down an alphabetically sorted list of firms. This is despite the statistical significance of the coefficients. The reason for this is that two of the coefficients are negative, namely coefficients for position dummy 5 and position dummy 50. This means that turnover is 14% lower for firms in the top 5% relative to other firms. Likewise, firms in the 25 to 50 percentile of firms have turnover that is 6.4% less than other firms. This runs contrary to the hypothesis that turnover uniformly decreases for alphabetically listed firms.

Table 15 Regression of turnover on position dummies

Variable	Statistic	p-value
position5	-0.134	<0.001
position25	0.179	<0.001
position50	-0.064	0.004
position75	0.226	<0.001
<i>Model information</i>		
Number of Observations	65979	
R ²	0.042	
R ² Adj.	0.038	

R ² Within	0.001	
R ² Within Adj.	0.001	
AIC	356554.4	
BIC	359210.8	
RMSE	3.59	

As stated earlier, the overwhelming trend from the results of different models is that continuous position does not have a strong statistically significant relationship with trading activity on the JSE. Conversely, most of the control variables were consistently statistically significant. Based on the persistence of the statistical significance of these variables, one can confidently conclude that negative return, book-to-market, market age, share price, firm size, leverage, beta and number of analysts have a significant impact on the trading activity of the JSE.

One reason for this result is that the JSE has a much smaller universe of shares to select from in comparison to other more developed markets. As a result, investors on the JSE do not face the same search costs as do investors in other markets. Figure 1 shows that at any given moment, the number of shares listed on the JSE does not exceed 300.

The JSE being a smaller market also means large firms dominate the market. This gives them additional visibility over other firms that naturally increases their liquidity. This explains the significance of firm size, market age and share price, as well as their correlation. This makes sense as large firms are likely to be older, have higher share prices and hence higher market capitalisations.

The key takeaway from these results is that in a developing market like the JSE, traditional markers for visibility take precedence as investors are less constrained given the lower number of firms. It would be interesting to see this study being replicated on similar markets.

4.4. Illiquidity

4.4.1. Illiquidity and continuous position

This section reports results of regressions with the Amihud illiquidity measure as the dependent variable. Firstly, liquidity is regressed on continuous position only, the results of which are reported in Table 16 below. The expected relationship is positive, since the Amihud illiquidity measure reflects the level of illiquidity of a stock (that is lower values are better). Therefore,

the hypothesised relationship between illiquidity and continuous position is positive. This means that the illiquidity measure increases as continuous position increases.

However, as evidenced by the coefficient on continuous position being negative, the hypothesised relationship is not realised in the regression output. Moreover, the coefficient is not statistically significant at the five percent level.

The R-squared value of the model is 1.1%, which indicates a poor model fit. The R-squared value is 0.6%, which further shows that the model explains very little of the variation in liquidity.

Table 16 Regression of liquidity on continuous position

Variable	Statistic	p-value
continuous	-0.016	0.067
<i>Model information</i>		
Number of Observations	65967	
R ²	0.011	
R ² Adj.	0.006	
R ² Within	0.000	
R ² Within Adj.	0.000	
AIC	167078.2	
BIC	169707.3	
RMSE	0.85	

4.4.2. Illiquidity and firm characteristics

In Table 17 below, the log of the Amihud (2002) illiquidity measure is regressed on continuous position and control variables from Chordia *et al.*, (2007). As was the case with regressions using turnover, continuous position is not statistically significant even in the presence of control variables. On the other hand, all the control variables have statistically significant coefficients, including positive return, which was not statistically significant with turnover.

The r-squared value of the model improves from 1.1% (0.6% adjusted) to 24.7% (24.3% adjusted), which shows that including the variables from Chordia *et al.*, (2007) greatly improves the model. The variables from Chordia *et al.*, (2007) can explain a significant amount of the variation of the liquidity measure.

The AIC of the multivariate model in Table 17 is lower than the AIC of the single variable model in Table 16. This means that adding the additional variables was beneficial to the model.

Table 17 Regression of liquidity on continuous position and control variables

Variable	Statistic	p-value
Continuous position	0.013	0.176
Positive return	0.350	(<0.001)
Negative return	-0.754	(<0.001)
Book-to-market	0.003	(<0.001)
Market age	0.000	0.013
Share price	-0.064	<0.001
Firm size	-0.149	<0.001
Leverage	0.000	0.082
Beta	-0.029	<0.001
Number of analysts	-0.024	<0.001
<i>Model information</i>		
Number of Observation	57549	
R ²	0.247	
R ² Adj.	0.243	
R ² Within	0.239	
R ² Within Adj.	0.238	
AIC	126581.3	
BIC	128929.0	
RMSE	0.72	

4.4.3. Alternative regressions

Just as before, adding multiple variables that are correlated with each other raises issues of multicollinearity. To combat this, alternative regressions are performed with some of the highly correlated control variables are dropped.

Table 18 excludes market age and firm size from the model, both of which are highly correlated with share price, as was demonstrated earlier. The results do not change from the model in Table 17 as continuous position is still statistically insignificant while all the control variables are statistically significant.

The r-squared value drops to 20.6%, which shows that the reduced model does not improve on the model from Table 17. Similarly, the AIC of the model increases, which shows that the model is not better than the model in Table 17.

Table 18 Regression of illiquidity on continuous position and control variables (excluding market age and firm size)

Variable	Statistic	p-value
Continuous position	-0.002	0.790
Positive return	0.409	<0.001
Negative return	-1.055	<0.001
Book-to-market	0.002	<0.001
Share price	-0.142	<0.001
Leverage	0.000	0.021
Beta	-0.030	<0.001
Number of analysts	-0.005	<0.001
<i>Model information</i>		
Number of Observations	57761	
R ²	0.206	
R ² Adj.	0.202	
R ² Within	0.197	
R ² Within Adj.	0.197	
AIC	129912.1	
BIC	132241.8	
RMSE	0.74	

The model in Table 19 below substitutes market age for share price, whilst still excluding firm size. This model produces a statistically significant coefficient on continuous position. The coefficient has the hypothesised sign. This shows that illiquidity reduces when with continuous position.

However, this result cannot be generalised given that this relationship is not found in the other model. Furthermore, the r-squared value of 13.9%, which is well below the r-squared values of previous models, shows that this model is not the best in explaining variation in illiquidity. The much higher AIC further demonstrates that this is a relatively poor model. The significance of continuous position here could simply be due to omitted variable bias. This can be seen by the small coefficients of market age and leverage, which shows that the excluded variables are much stronger in the relationship with illiquidity.

Table 19 Regression of liquidity on continuous position and control variables (excluding share price and firm size)

Variable	Statistic	p-value
Continuous position	0.059	<0.001
Positive return	0.485	<0.001
Negative return	-1.415	<0.001
Book-to-market	0.003	<0.001
Market age	0.000	0.345
Leverage	0.000	0.029
Beta	-0.017	0.019
Number of analysts	-0.036	<0.001
Number of Observations	57764	
R ²	0.139	
R ² Adj.	0.135	
R ² Within	0.130	
R ² Within Adj.	0.130	
AIC	134592.8	
BIC	136923.5	
RMSE	0.77	

Table 20 reports results from regression of illiquidity on continuous position and control variables, excluding share price and market age. In this case, continuous position is statistically significant at the one percent level. The coefficient has the hypothesised sign, which shows that illiquidity increases as continuous position increases. Therefore, according to this result, traversing an alphabetically sorted list of shares will result in decreasing liquidity. This finding suggests the firms listed at the top of the alphabetically sorted lists have more liquidity, when controlling for positive return, negative return, book-to-market, firm size, leverage, market beta and number of analysts.

This model measures up well with the main model reported in Table 17, although it does not improve on it. The difference between the two is slight. The r-squared value of the model in Table 20 is 23.8%, compared to 24.7% for the model from Table 17. The AIC of the model in Table 20 is 127251.7, which is only slightly higher than the AIC of the model from 16 (126579.8). Both measures show that the models are similar.

Table 20 Regression of illiquidity on continuous position and control variables (excluding share price and market age)

Variable	Statistic	p-value
Continuous position	0.034	<0.001
Positive return	0.355	<0.001
Negative return	-0.776	<0.001
Book-to-market	0.003	<0.001
Firm size	-0.191	<0.001
Leverage	0.000	0.698
Beta	-0.024	<0.001
Number of analysts	0.023	<0.001
<i>Model information</i>		
Number of Observations	57550	
R ²	0.238	
R ² Adj.	0.234	
R ² Within	0.230	
R ² Within Adj.	0.230	
AIC	127253.7	
BIC	129583.4	
RMSE	0.73	

4.5. Discussion of results

An interesting result from this study is the significance of continuous position in some of the liquidity models. This seems to suggest that alphabetic listing matters more for liquidity than it does for trading activity. Although trading activity and liquidity are related, they are not the same. Loosely defined trading activity is how much a share trades, while liquidity is how many shares can be traded without severely impacting the price of the share.

Continuous position being significant in some of the liquidity models suggests that when controlling for some of the Chordia et al., (2007) variables, liquidity reduces by alphabetic ordering. The fact that this result is not persistent even when contained to the liquidity models means that there is a slight alphabetic bias that exists on the JSE, but it is not as strong as in western markets. Moreover, it could also mean that the other variables have a greater impact and once that has been factored in, alphabetic listing then comes in. Practically, this would

mean that investors would first use share price, firm size, market age, leverage, number of analysts, market beta and negative return before sorting firms alphabetically, almost as a last resort tie breaker.

The coefficients on positive return are all positive in the liquidity models, while the coefficients on negative return are all negative in the liquidity models. Recall that the Amihud illiquidity measures the level of illiquidity of a share. This means that a positive relationship with the illiquidity measure indicates that liquidity reduces with positive returns and increases with negative returns. This suggests that positive returns discourage selling or encourages increasing the asking price to gain more, which reduces liquidity. The reverse of this insight would be the case for negative returns.

Nonetheless, the lack of a persistently strong showing from continuous position ultimately leads to the conclusion that continuous position is not a significant factor in determining trading activity on the JSE. Therefore, the answer to the research question stated in Chapter 1.4 is that alphabet bias on the JSE has not been substantiated.

The findings in this study are largely consistent with the behavioural finance and microstructure literature. The absence of evidence for alphabetic bias aligns with the position of Amihud (2002) discussed in earlier in the literature review notes that illiquidity in less-developed markets may constrain attention-driven trading strategies. The results of Jacobs and Hillert (2015) may only be generalisable to larger, more complex markets with several thousands of companies listed.

The firm-specific determinants of trading activity are statistically significant, which is in line with the visibility hypothesis stated by Barber and Odean (2008). Characteristics that are easily observable may increase investor attention and increase trading activity. Firms with higher share prices and more analyst coverage are proxies for firm visibility. Smaller firms, despite being less covered, may attract attention due to their perceived speculative potential.

The decomposition of returns into positive and negative components further highlights the asymmetric reaction of investors to share price changes. The finding that negative returns are associated with increased turnover is consistent with the disposition effect (Shefrin and Statman, 1985), where investors are more likely to act after losses than after gains. This asymmetry is also reflected in the relatively weaker and less consistent effect of positive returns on trading activity.

Overall, the results confirm the relevance of firm visibility in trading activity. The significance of firm characteristics indicates that investor behaviour on the JSE may be guided more by firm visibility and relative valuation measures than by search heuristics such as alphabetical position.

4.6. Conclusions on hypotheses

Therefore, results from the regressions do not fully support the stated hypotheses of the study. The results of the tests on alphabetic bias does not provide enough evidence to reject the null hypothesis that there is no statistically significant relationship between alphabet bias and trading activity.

In contrast, the strong evidence found for the relationship between trading activity and firm-level characteristics leads to the rejection of the null hypothesis. Therefore, the results of the hypothesis testing are that firm-specific characteristics such as share price, market capitalisation, and market age have a significant association with trading activity.

The firm-level characteristics consistently have statistically significant coefficients across multiple model specifications. The results suggest that visibility is an important factor driving attention to a particular share, which may increase trading activity for that share. These results underline the importance of firm visibility with regards to trading behaviour on the JSE.

In conclusion, this study finds no statistically significant evidence of an alphabetical bias on the JSE. Therefore, it cannot be confirmed that alphabetical sorting affects trading activity in the South African context.

The results suggest that trading activity is sensitive to both positive and negative returns, but the direction and strength of the relationship differ. In particular, the coefficient on negative returns is positive and statistically significant, indicating that downward price movements tend to elicit increased trading. This may reflect rebalancing activity, stop-loss triggers, or heightened investor attention in response to negative performance. Positive returns also show a positive relationship with trading, although the magnitude is comparatively smaller and less consistent across model specifications.

5. Conclusion

Listing items by their names in an alphabetic order has been a feature throughout history. The effect of this may be that more attention is given to people or things that have names that start

with letters closer to the start of the alphabet. This is due to humans psychologically favouring items at the start of the list. It has been shown, for example, that election candidates who appear at the top of a ballot are more likely to receive votes. In academia, authors with names that start with the first letters of the alphabet attain more success than their counterparts with names beginning with letters further down the alphabet. This has been termed as alphabet bias.

The traditional assumption in economic theory is that market participants evaluate decisions utilising a rational expectations framework. In financial markets, investors are assumed to make decisions using publicly available information and the expected mathematical payoff from the investment. The EMH falls into this stratum. Over the years, numerous empirical evidence was produced that contradicted the EMH. The explanation for these apparent contradictions is that the rational expectations model is not a fully accurate representation of how humans make decisions. The field of behavioural finance has shown that human psychological predispositions have an impact on how market participants evaluate investment choices.

In finance, publicly traded shares are typically listed in alphabetic order. Since it has been shown that humans exhibit alphabet bias and that human biases affect decision making in capital markets, one would expect shares at the top of the list to receive considerably more attention from investors compared to shares at the bottom of the list. The hypothesis is that search costs are high and therefore investors will likely settle on a share before they get to the bottom of the list. The probable impact of this is that trading activity will be affected by a firm's position in a list.

Jacobs and Hillert (2015) address this by testing for alphabet bias in trading activity. The study shows that shares trading in United States of America (US) and other international markets that are at the top of an alphabetically sorted list trade 5 to 15% more frequently than shares at the bottom of the list. Furthermore, shares with names that start with an early alphabet letter are more liquid than shares with names starting with letters further down the alphabet.

Therefore, the results of Jacobs and Hillert (2015) shows that firm position has an impact on trading activity, even after controlling for variables that are known to influence trading activity. The results hold true even in international markets. In Japan, which has numerical ticker names, shares at the top of the list have more trading activity than shares at the bottom of the list.

The aim of this study was to determine if alphabet bias exists and whether it affects trading

activity on the JSE. The methodology used in this study closely follows Jacobs and Hillert (2015) except for the omission of variables for which data is unavailable.

The key independent variable of interest for this study was continuous position, which is a measure of a firm's position in the list of companies sorted alphabetically. Continuous position is the company's rank in an alphabetically sorted list divided by the total number of shares in a month. This value is therefore between 0 and 1. As per Jacobs and Hillert (2015), the continuous position variable was constructed from a list of firm names in each month.

As in Jacobs and Hillert (2015), two measures of trading activity are employed, namely share turnover and liquidity. Share turnover is the log of the ratio of volume of shares traded for company i at time t to the number of shares outstanding at time. Raw values of turnover were also used in robustness checks. Liquidity is proxied by the Amihud (2002) measure of liquidity, which is the log of the annual average ratio of the absolute share return to the trading volume (in rands). Trading activity and liquidity are then regressed on a vector of firm characteristics, which includes continuous position (proxying for alphabetical share listing). Control variables are taken from Chordia *et al.*, (2007), which are: continuous position, positive return, negative return, book-to-market, market age, firm size, share price, leverage, beta and number of analysts.

The results of univariate regression of turnover and liquidity on continuous position variable did not produce statistically significant coefficients. This remained the case for the most part when control variables were included. Two of the liquidity models had a statistically significant coefficient on continuous position. However, both models excluded key variables, namely share price and firm size. Based on this, it could not be reliably concluded that continuous position had an impact on trading activity or liquidity, although it appeared to be more impactful on liquidity than trading activity, two related but ultimately different measures. The results were persistent even in robustness checks and alternative models.

On the other hand, the control variables were consistently statistically significant across the models. Consequently, the study concludes that negative return, book-to-market, market age, share price, firm size, leverage, beta and number of analysts have a significant impact on the trading activity and liquidity of the JSE.

One possible explanation is the relatively small number of listed shares on the JSE, which may reduce investor search costs and thereby diminish the impact of alphabetical ordering. This

reduces search costs for investors and therefore alphabetic listing is less of a factor. In addition to this, large firms are known to dominate liquidity and trading activity on the JSE.

In conclusion, this study does not find strong evidence for an alphabetic bias on the JSE. While this study does not find evidence of alphabetic bias, the significant impact of firm size, share price, and analyst coverage on trading activity may suggest the presence of other investor preferences or attention-driven biases.

This conclusion is in line with the psychological research showing that investors do have biases that manifest themselves on the stock market. For the JSE, this would be biases towards large firms with high share prices and several analysts covering the firm.

The study recommends further research of the same on similar markets to validate the conclusions made here. Future research could explore the impact of market structure, investor type, or digital trading interfaces on behavioural patterns in emerging markets.

References

- Admati, A. R., & Pfleiderer, P. (1988). A theory of Intraday patterns: Volume and price variability. *Review of Financial Studies*, 1(1), 3-40.
- Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, 5(1), 31-56.
- Amihud, Y., & Mendelson, H. (1987). Trading mechanisms and stock returns: An empirical investigation. *The Journal of Finance*, 42(3), 533.
- Atiase, R. K., & Bamber, L. S. (1994). Trading volume reactions to annual accounting earnings announcements. *Journal of Accounting and Economics*, 17(3), 309-329.
- Bamber, L. S. (1986). The information content of annual earnings releases: A trading volume approach. *Journal of Accounting Research*, 24(1), 40.
- Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies*, 21(2), 785-818.
- Brandt, M. W., Brav, A., Graham, J. R., & Kumar, A. (2009). Idiosyncratic volatility puzzle: Time trend or speculative episodes? *The Review of Financial Studies*, 23(2), 863-899.
- Brennan, M. J., & Hughes, P. J. (1991). Stock prices and the supply of information. *The Journal of Finance*, 46(5), 1665.
- Brennan, M. J., & Subrahmanyam, A. (1995). Investment analysis and price formation in securities markets. *Journal of Financial Economics*, 38(3), 361-381.
- Brogaard, J., Engelberg, J., & Parsons, C. A. (2014). Networks and productivity: Causal evidence from editor rotations. *Journal of Financial Economics*, 111(1), 251-270.
- Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *Journal of Finance*, 52(1), 57-82.
- Carney, D. R., & Banaji, M. R. (2012, June 27). *First is best*. PLOS ONE.
- Chau, M., & Vayanos, D. (2008). Strong-form efficiency with monopolistic insiders. *Review of Financial Studies*, 21(5), 2275-2306.
- Chordia, T., Huh, S., & Subrahmanyam, A. (2007). The cross-section of expected trading activity. *The Review of Financial Studies*, 20(3), 709-740.

- Diether, K. B., Malloy, C. J., & Scherbina, A. (2002). Differences of opinion and the cross section of stock returns. *The Journal of Finance*, 57(5), 2113-2141.
- Diether, K. B., Malloy, C. J., & Scherbina, A. (2002). Differences of opinion and the cross section of stock returns. *The Journal of Finance*, 57(5), 2113-2141.
- Dimson, E. (1979). Risk measurement when shares are subject to infrequent trading. *Journal of Financial Economics*, 7(2), 197-226.
- Einav, L., & Yarov, L. (2006). What's in a surname? The effects of surname initials on academic success. *Journal of Economic Perspectives*, 20(1), 175-188.
- Gallant, A. R., Rossi, P. E., & Tauchen, G. (1992). Stock prices and volume. *Review of Financial Studies*, 5(2), 199-242.
- Goetzmann, W. N., & Kumar, A. (2008). Equity portfolio diversification. *Review of Finance*, 12(3), 433-463.
- Grinblatt, M., & Keloharju, M. (2001). How distance, language, and culture influence stockholdings and trades. *The Journal of Finance*, 56(3), 1053-1073.
- Grossman, S. J., & Stiglitz, J. E. (1980). On the Impossibility of Informationally Efficient Markets. *The American Economic Review*, 70(3), 393-408.
- Grullon, G., Kanatas, G., & Weston, J. P. (2004). Advertising, breadth of ownership, and liquidity. *The Review of Financial Studies*, 17(2), 439-461.
- Harris, M., & Raviv, A. (1993). Differences of opinion make a horse race. *Review of Financial Studies*, 6(3), 473-506.
- Hillert, A., Jacobs, H., & Müller, S. (2014). Media makes momentum. *Review of Financial Studies*, 27(12), 3467-3501.
- Ho, D. E., & Imai, K. (2008). Estimating causal effects of ballot order from a randomized natural experiment. *Public Opinion Quarterly*, 72(2), 216-240.
- Itzkowitz, J., Itzkowitz, J., & Rothbort, S. (2015). ABCs of trading: Behavioral biases affect stock turnover and value *. *Review of Finance*, 20(2), 663-692.
- Jacobs, H., & Hillert, A. (2015). Alphabetic bias, investor recognition, and trading behavior *. *Review of Finance*, 20(2), 693-723.

- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263.
- Karpoff, J. M. (1986). A theory of trading volume. *The Journal of Finance*, 41(5), 1069.
- Kim, O., & Verrecchia, R. E. (1991). Trading volume and price reactions to public announcements. *Journal of Accounting Research*, 29(2), 302.
- Lee, C. M., & Swaminathan, B. (2000). Price momentum and trading volume. *The Journal of Finance*, 55(5), 2017-2069.
- Lo, A., & Wang, J. (2000). Trading volume: Definitions, data analysis, and implications of portfolio theory. *The Review of Financial Studies*, 13(2), 257-300.
- Meer, J., & Rosen, H. S. (2011). The ABCs of charitable solicitation. *Journal of Public Economics*, 95(5-6), 363-371.
- Morse, D. (1980). Asymmetrical information in securities markets and trading volume. *The Journal of Financial and Quantitative Analysis*, 15(5), 1129.
- Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3), 703-708.
- Richardson, M. L. (2008). Alphabetic bias in the selection of reviewers for the American Journal of Roentgenology. *American Journal of Roentgenology*, 191(6), W213-W216.
- Ritter, J. R. (1991). The long-run performance of initial public offerings. *The Journal of Finance*, 46(1), 3.
- Rosenberg, B., Reid, K., & Lanstein, R. (1985). Persuasive evidence of market inefficiency. *The Journal of Portfolio Management*, 11(3), 9-16.
- Shiller, R. J. (2003). From Efficient Markets Theory to Behavioral Finance. *Journal of Economic Perspective*, 17(1), 83-104.
- Simon, H. A. (1957). *Models of man: Social and rational; Mathematical essays on rational human behavior in society setting*. John Wiley & Sons, Inc.
- van Praag, C. M., & van Praag, B. (2008). The benefits of being economics professor a (rather than Z). *Economica*, 75(300), 782-796.

Young, N., & Auret, C. (2018). Liquidity and the convergence to market efficiency. *Investment Analysts Journal*, 47(3), 209-228.