LUNDS UNIVERSITET
Ekonomihögskolan

Department of Business Administration
Bachelor's Thesis in Corporate Finance
FEKH89
Fall semester 2020

Wisdom in absence of a crowd

A study of drivers in valuation of small cap stocks, with limited ownership, listed on the Swedish market

Authors:

Allgulander, Oscar	961119 - 2775
Bladh, Olivia	970403 - 7846
Kuntscher, Herman	971023 - 5970

Supervisor: Anamaria Cociorva

Abstract

Title: Wisdom in absence of a crowd

Seminar Date: 01/15/2021

Course: FEKH89

Authors: Oscar Allgulander, Olivia Bladh and Herman Kuntscher

Advisor: Anamaria Cociorva

Key words: Non-frequently traded stocks, liquidity, market imperfections, arbitrage, law of one price, effective market hypothesis, share price, valuation

Purpose: The purpose of this study is to investigate what factors, if any, are most prevalent in the valuation of small cap stocks, with limited ownership as compared to their fundamental values, where crowd based trading and insights do not necessarily prevail.

Methodology: A deductive and comparative study realized in quantitative research based upon a sample selection of smaller companies. Findings are evaluated in a regression analysis to statistically prove any relationship between the independent variables.

Theoretical perspective: The foundation of the study is built on the base of three different, but co functioning theoretical frameworks; The effective market hypothesis, arbitrage price theory, and the law of one price.

Empirical foundation: The sample covers 100 companies all listed on the swedish stock market. Each company has a maximum market cap of 250 million SEK and a maximum number of 500 shareholders. The timespan reaches from November 30, 2019 to October 31, 2020.

Conclusions: When analysing enterprise value over book value, volatility was found to be statistically significant throughout all stratums. The results for enterprise value over sales generated a different result showing information asymmetry as significant in the full sample and asset heavy subset.

Sammanfattning

Titel: Wisdom in absence of a crowd

Seminariedatum: 01/15/2021

Kurs: FEKH89

Författare: Oscar Allgulander, Olivia Bladh samt Herman Kuntscher

Handledare: Anamaria Cociorva

Nyckelord: Low volume stocks, likviditet, marknadsstörningar, arbitrage, law of one price, effective market hypothesis, fundamentalvärde

Syfte: Syftet med denna studie är att undersöka vilka faktorer, om några, som är mest prevalenta vid värdering av svenska småbolagsaktier, med få ägare, i relation till deras fundamentala värden, när handel baserad på folkmassor och deras åsikter inte nödvändigtvis råder.

Metod: En deduktiv samt komparativ studie realised i en kvantitativ studie som baseras på ett urval av små företag på den svenska aktiemarknaden. Svaret från utredningen analyseras även i en regressionsanalys med syfte att påvisa eventuella samband mellan de självständiga variablerna.

Teoretiskt perspektiv: Grunderna till studien baseras på tre olika, men samspelta teorier; *The effective market hypothesis*, *Arbitrage price theory* och *The law of one price.*

Empiriskt underlag: Urvalet består av 100 företag som alla är listade på den svenska aktiemarknaden. Varje enskilt företag har en inkomst på maximalt 250 miljoner SEK samt en maxgräns på 500 aktieägare. Tidsperioden sträcker sig från november 2019 till oktober 2020.

Slutsats: Vid analysen av EV/Book så framgick volatilitet som statistiskt signifikant i alla strata. Resultatet för EV/sales visade informationsasymmetri som signifikant när det kom till det hela urvalet samt den övervärderade delmängden.

Acknowledgements

It has been a great pleasure to engage in this study and the authors are extremely grateful to the economics department at Lund University for providing the tools needed to conduct this study. The authors would also like to share their extreme gratitude for their supervisor Anamaria Cociorva for her great support and guidance.

Oscar Allgulander Olivia Bladh Herman Kuntscher

Terms and definitions

Liquidity: Defined as the volume divided by the number of shares outstanding, the liquidity acts a measure of the volume turnover angle of one specific stock - a measurement that validates how rapid an asset can sell, and be bought without negatively impacting its price.

Volume: Total number of shares traded on a security on the current day. If the security has not traded, then it is the total number of shares from the last day the security was traded.

Market cap: Defined as the current share price multiplied with the number of shares outstanding. In the context of this study the measure is interpreted as the market's perception representing fundamental value, and is thus used as the base line representing fundamental value from which over-or-under pricing diverges.

Number of shareholders: Defined as the number of investors, invested in or, listed as shareholders according to Bloomberg Terminals.

Market makers: A firm or individual who actively quotes two-sided markets in a security, providing bids-and-asking-quotas along with the market size of each.

Small cap shares with limited ownership: In this paper defined as shares with a maximum number of shareholders of 500, and a maximum market cap of 250 MSEK

1. Introduction

In this chapter, the background to the chosen study is presented followed by a more in depth description of the problem discussion. The chapter leads up to the main thesis and the sole purpose of this study followed by the chosen limitations and scope.

1.1 Background

The stock market hosts some of the world's most valuable assets, as it acts as a value maker for some of the world's largest fortunes. What is so special about the values on the stock market is that they are mostly based on other stakeholders' beliefs of what that specific asset should be worth. As with all valuable belongings, money is just a social construct and it is only worth as much as someone else is willing to pay for it. There is no universally agreed exact number that defines the value of any company, but there exists a plethora of methods for pricing its stock. Like guessing jelly beans in a jar, some overestimate future cash flows, some underestimate. One could assume that the average of all those estimates should logically be near the true intrinsic value of the company.

Basic theories on how assets are valued on any market are based on a notion that all investors have simultaneous access to all relevant information and that all investors base and continuously adjust their investment portfolio. In this perfect world, trading is a continuous, harmonious connection where buyers and sellers meet and trade based on small variances in perceived value. However, it is very evident that neither markets nor access to information is perfect, and that the behavior of investors is neither fully predictable nor totally rational at all times. Hence, it is important to understand the prevailing drivers for shares in a reality that is somewhat different from the assumptions of the theory.

1.2 Problem Discussion

Two widely used markers for a company's true value are the measurable *fundamental value*, as well as the acting *spot price.* The fundamental value is in its most simplistic form explained as *the present value of future cash flows*. By this, *fundamental values* do not only consider the company's status today when rating its value, but also perceived returns of important future achievements to come in combination with assumptions of external factors and their relevance to the analyzed company's ability to generate future cash flows. *Spot prices* on the stock market could be defined as the average of the wisdom of the crowd, more precisely what investors and buyers are willing to spend in order to possess that exact stock. Brokers' beliefs in the asset as well as their willingness to purchase or sell creates a normal distribution curve providing a clear population mean to what the market believes the underlying asset is worth. In a perfect market all investors have access to all relevant information and the underlying stock can be traded at any time matching fundamental value and spot price. However, the market is not perfect - comprehensive and exhaustive relevant information is not accessible to all investors at any given point on time and all investors do not assess and adjust their holdings continuously based on new market information (Bryman et.al, 2019. p.194-195).

In the financial world, the volume turnover angle of one specific stock is defined as liquidity, a measurement that validates how rapid an asset can be sold, and be bought, without negatively impacting its price. The more liquid an asset is, the more purchases of that exact asset are being made, which in turn makes it easier to sell the asset for what is deemed by the many to be a fair market value (CFI 1, 2020) When a stock is liquid, the value of the underlying company is based on a multitude of information sources coming from multiple trades a day. In addition, given sufficiently many trades, the difference between the Ask (Sell) and Bid (Buy) courses (the Spread) will be small, making for a more exact spot price. However, stocks are likely to be traded on market imperfections which also will drive the price to converge towards a specific point or spot price. The higher number of trades and sources of information results, according to Bryman et. al, in a more correct validation of the stock price (Bryman et.al, 2019. p.194-195).

Thus, when an asset is illiquid, market imperfections are more noticeable and often problematic. Smaller companies, especially in the early development stage, tend to release fewer news for investors to trade on and if they do, finding relevant information could be cumbersome and therefore costly. In addition, fewer investors are interested in small companies, and market makers have found it difficult to make money on the limited trades, which leads to underrepresentation in analyzing smaller companies. Given all this, it perhaps makes sense why institutions and market makers tend to avoid such securities all together. Even though the lack of information makes sense due to the size of the company, the outcome can be a clear threat towards the so-called "Wisdom of the Crowd" – both in terms of the quality of the Wisdom and the size of the Crowd. With less information, and limited trades per day, the normal distribution curve is constrained by only a few individual beliefs. In opposition to the spot price of liquid shares that are based on generally accepted mean assumptions, illiquid shares can fluctuate greatly based on a few brokers' opinions - with a large risk of sample errors or misconceptions (Bryman et.al, 2019. p.194-195). Simply put, due to the illiquidity, larger orders might shift the spot-price considerably. Because of this, investors holding illiquid or thinly traded assets could end up in a position where they cannot sell their shares for an extended period, at least not without changing the price substantially. This problem extends to prospective buyers, both at the time of buying and then later when they want to sell the shares. What this results in is to no surprise that the sales of thinly traded assets often occur at prices far, or at least somewhat, lower or higher than if there was a liquid market (Longstaff, 2014).

Different countries have different rules and regulations to counteract some of the most common market imperfections such as information asymmetry, volatility, illiquidity and high transaction costs. In Sweden, the rules and regulations are divided by state law and codes of conduct from the different markets. The government challenges the information asymmetry and transaction costs by setting up clear guidelines on how market makers and others must behave in asset trading and market entry. The different markets themselves then challenge volatility and illiquidity by putting up requirements, such as different demands for liquidity providers, for their customers to follow.

1.3 Purpose

The purpose of this study is to investigate what factors, if any, are most prevalent in the valuation of Swedish small cap shares with limited ownership, as compared to their fundamental values where crowd based trading and insights do not necessarily prevail.

1.4 Research question

- What variable, to the highest extent within the set models, correlates with the valuation of small cap shares with limited ownership?

1.5 Limitations and scope

1.5.1 Limitations

The study uses a sample selection of shares from small companies listed on the Swedish markets; First North Stockholm, Spotlight Stock Market, and Nordic SME. By limiting the research to one country's exchanges, the number of potential exogenous factors impacting the data is minimized. Using exchanges from multiple countries would have subjected parts of the sample to a range of outlying and distorting variables, such as, e.g., local legislation relating to shares or audit regulations and practises.

The sample covers 100 companies and the chosen study parameters per traded company are a maximum market cap of 250 million SEK and a maximum number of 500 shareholders per share, in accordance with the definition in this paper as small cap shares with limited ownerships. A sample size of 100 was chosen as the size of the sample needed to satisfy two criteria; be as large as possible in order to produce as robust results as possible, and be small enough that time spent on collection and data management does not outweigh time spent on interpretation and analysis.

For purposes of this study only common shares (also known as. ordinary shares) are analyzed with the assumption that other securities (e.g., preferred stock, bonds, and other derivatives) correlate with the value of the underlying common shares. The authors have hence focused the

analysis solely on common shares, and have excluded all other securities related to the shares in focus.

Furthermore the study has chosen to focus on two different measurements of company fundamental value; *enterprise value over book value* and *enterprise value over sales.* These two multiples were chosen since they not only provide the study with measurements regarding over- or undervaluation, but they also cover two different true value-measurements to further provide the study with a multidimensional point of view.

Lastly, in order to answer the last research question "*What variable, to the highest extent within the set model, correlates with the valuation of said illiquid and less traded shares?*" the study has chosen four different hypotheses variables supporting four different market imperfections. The chosen variables are: *Information asymmetry, Volatility, Illiquidity and Transaction costs.*

1.5.2 Scope

The chosen time window from which data is gathered is to be defined as November 30, 2019 to October 31, 2020 with daily readings over "one rolling year". The sample does take action during the time period for the global pandemic of Covid-19. While having this in mind, the study did not proceed any precautions to cancel out any effects in regards to the global virus spread, since the pandemic is still in movement and shows different outcomes everyday. The impact from Covid is difficult to isolate, and the authors believe that isolating or making clear connections to any proven Covid-impact will be close to impossible in many years to come. The pandemic has shown itself to be not a short crisis-period, but rather contributing to global behavioral changes that will go on for several years. Hence, the study will not make any clear connections between the outcome and the ongoing pandemic.

1.6 Contribution

By providing fact based insights this study aims to provide investors, as well as small companies already publicly traded, or those considering an IPO, with an additional tool for their financial toolbox to maximize investors' profits from trading or reduce the cost of raising capital for listed companies. By gaining deeper insights into these individual forces and their potential co-dependencies, a rational investor should be able to make wiser investment decisions in what, to the uniformed, could be viewed as an irrational market. Especially during times of market crisis, as experienced during the 2020 pandemic.

2. Theoretical framework

In this chapter, the three main theories are presented, which later will be used to interpret the results. This is followed up by relevant previous research and statistical methods.

2.1 Arbitrage price theory

In 1976 an American economist named Stephen Ross developed the arbitrage price theory in order to generate more accurate calculations of asset pricing. The arbitrage price theory (later called APT) is now a commonly used theory to calculate asset pricing using a connection of the assets expected returns, as well as macroeconomic factors affecting the risk of the asset. The *Corporate Finance Institute* argues that the main purpose of the APT is to act as a guide to point out a fair market price for assets that may be incorrectly priced during a temporary period of time. Just like the market shows, APT is built on the idea of market imperfections which results in wrongful valuation of assets, which then leads to mispricing during a brief period of time. With time, the market usually corrects itself bringing the asset price back to, what is generally deemed as, a more fair market value. However, the APT focuses on the short time period in between mispricing and correction in order to turn wrongfully priced assets into arbitrage opportunities. The classic meaning of arbitrage suggests that simultaneous purchases and sale of the same assets at different exchanges creates an opportunity to capture profits at a low risk. APT offers a slightly altered approach to arbitrage than the classical definition, since it argues that the opportunity of arbitrage is, not in fact, risk free. Instead it maintains that the presence of risk may offer a high opportunity of success. The addition of the risk component is intrinsic to the model itself since APT offers calculations to a fair market price, without claiming that the mathematical outcome of the calculation will match reality. If it does match reality APT could assist in finding great arbitrage opportunities that are not easily discernible when observing the market (CFI 2,2020).

2.2 Law of one price

The law of one price is the underlying principle to a unified asset valuation where the asset carries the same value no matter what market or currency it is presented in. By this, the principle holds that if one asset trades at two different prices one, or both, of the values are wrong. The absence of one price creates an opportunity for arbitrage, where brokers can take advantage of the price spread and trade the assets on a parallel market to make a profit based on the "buy low, sell high" principle. However, these profits generated by taking advantage of pricing differentials only benefit a few since they are highly desirable and attract profit seekers. With a multitude of brokers wanting to seize the trading opportunity, forces of supply and demand will eliminate all arbitrage opportunities. The spread will shrink until the lower and the higher price have calibrated to one common price point, and thereby again adhering to the law of one price. The only discernible spread should at that point be attributed to a potential difference in the cost of trading. Hence, in order to take advantage of the profit opportunity presented in arbitrage it is essential to be one of the first brokers making the market trade before the market calibrates to eliminate the window of profit opportunity (Berk, 2014, s. 104).

2.3 The Effective Market Hypothesis continued

In the paper *Efficient Capital Markets: A review of Theory and Empirical Work*, Eugene F. Fama introduced what would be named the efficient capital market hypothesis. An efficient capital market is one where the price of securities fully reflects all available information. However, This perfectly efficient capital market has three prerequisites; 1) transaction costs for trading securities do not exist, 2) all available information is easily accessible for all market participants, 3) there are homogeneous expectations regarding the current information and future prices (Fama, 1970). In spite of rendering Fama a Nobel prize in economics 2013, the theory has its flaws and other theories that challenge its positions exist.

From a purely logical standpoint, ignoring empirical data, the theory has a built in paradox. Assuming efficient markets there would be no incentive to spend billions of dollars on money

managers since they, in line with the theory, can not know more than you do (perfect information symmetry) and, following the same line or reasoning, can not reach higher alphas, since there are no opportunity for arbitrage. Furthermore, the theory has a loophole in that there must exist incentive for private investors to collect information to drive trade, as made famous by the two economists Grossman and Stiglitz (Grossman et.al, 1980).

Robert Shiller, widely known as the father of behavioural economics and Co-laureate with Fama, believes that share price stems from fundamental values tainted by bias and mistakes native to human psychology. Humans panic, subject themselves to various herd like behavior, or allow themselves to be swept away in irrational rallies, but as stated in the following literature; assuming that the majority of investors are naive implies that beating the market if you are smart should be very easy, and yet it is clearly not.

The effective market hypothesis as well as the behavioral theories of Shiller are questioned by Lasse Pedersen in *Efficiently inefficient* in which he uses examples featuring money managers and differing share classes to point out that neither the securities market or the money manager market is efficient. Instead these are efficiently inefficient; for money managers, inefficient enough to achieve enough abnormal return in order to validate their own existence and cover their own costs - but not inefficient enough to enable more competition (Pedersen, 2015). For the securities market, an example is made using dual share classes such as BMW preferred and common shares, where the preferred shares tend to be priced lower than the common shares, adjusted for the benefits from holding a common share, and thus allows for statistical arbitrage. Another example is Royal Dutch Petroleum and Shell Transport, who share a 60/40 ratio, that is to say for every 1$ in group cash flow the Shell Transport shareholder receives 0.4$ divided per outstanding share. An efficient market would price these in line with the suggested ratio, instead the valuations have fluctuated, leaving one share more expensive relative to the other, sequentially opening up for arbitrage profits by going short on the overpriced share and long the underpriced one. This kind of arbitrage is naturally tempered by transaction costs, something that only gets worse the smaller the share. For small shares, the transaction costs increase with trade size, often barring hedge funds and other money managers from being able to take positions in them, less they are able to break up the trade into smaller pieces. (Pedersen, 2015).

These examples are clear violations of the efficient market hypothesis, as the hypothesis would imply that the spread would instantly close (Fama, 1970). Instead the spread in the latter example to this day fluctuates around 1, lending proof to Pedersen that markets in fact are efficiently inefficient (Pedersen, 2015).

2.4 Theory discussion

The main choice of theories has its roots in the nature of the market, and all three theories intertwine well with each other as they all support the same market flow. Given that the research is based on a deductive study, the authors have chosen the theories that provide the tools that will be of greatest use and value to support the main thesis of the study. The chosen theories have clear, direct connections to the purpose of this paper, and therefore provide critical value to the research where included. Furthermore, the theories also provide this study with clear models to show how the market behaves differently when a stock is being traded frequently or not. Since the study has a clear focus on non frequently-traded stocks, the authors have chosen to interpret the given theories from a “small-share” perspective.

The effective market hypothesis supports the belief that an efficient capital market is one where the price of securities fully reflects all available information. To prove the theories of an effective market, economists can call for APT to prove that the same asset might have two different values due to lack of information. This is of great importance for smaller non frequently-traded companies since the lack of trades can be viewed as inadequacy in information. As mentioned previously in this study, single trades rules out a price based on a general mean assumption and is reduced down to a few brokers' opinions. In an unregulated market where distinct and varying opinions of a very limited number of brokers prevals, the opportunity to seize larger arbitrage opportunities among non frequently-traded stocks ought to exist. Even though arbitrage opportunities are rare some argue that it could supposingly be easier to find it among smaller non frequently-traded companies, due to those brokers' opinions, again, supporting the arbitrage opportunities of ATP. Another theory presented, the law of one price, also intertwines perfectly with the other theories. Even if the APT focuses on the short time period in between mispricing

and correction in arbitrage opportunities, the law of one price supports the fact that the asset eventually loses its arbitrage opportunity due to the forces of supply and demand of the market. These theories are conjointly presented in this paper to provide a broader understanding for the importance of correct, timely and relevant information and clear validation. Since arbitrage opportunities could be easier to find among smaller non frequently-traded companies, this also means that the law of one price is not as accurate of an assumption for smaller non frequently-traded companies as it is for the market as a whole. The lack of trades and information asymmetry challenge the concept of market corrections, and mispriced assets could stay mispriced for a longer period of time.

2.5 Previous Research

2.5.1 Liquidity measurements

A prevailing and persistent issue from prior research is the question of how to measure liquidity. Since liquidity as such is not a distinct and observable variable in itself, a wide range of different proxies has been used, but there's no consensus on a best practice. This was discussed in *How should liquidity be measured*, where the authors Aitken and Comerton-Forde discussed a range of different proxies and divided them into two categories; trade-based and order-based. In their argument, trade-based proxies are those that use variables such as turnover-ratio, volume, and number of trades as an indication of liquidity. However, it is worth noting that their arguments do not give any indication of future liquidity. Order-based proxies are to those that consider the bid-ask spread, and thus the transaction costs, associated with trading the security. The conclusion of the paper is that the two categories aren't necessarily correlated, and that the choice of measure could thus have significant effects on the outcome of a research (Aitken et.al, 2002).

Illiquidity has also been analyzed from the perspective of providing a limitation to continuous trading at will. Professor Francis Longstaff has published several papers on this topic. In a paper from 2001, Longstaff analyzed the welfare effects investors suffer when there are liquidity restrictions. He later continued within this area and introduced a framework where illiquid shares are viewed as options. In his theory, Longstaff uses the notion that an investor can't exercise an

option or sell an illiquid asset at will. By using the time-frame where there are no buyers as a variable, he integrates that very variable in a function to determine the discount required for an investor to buy an illiquid asset instead of an identical liquid one. Longstaff suggests that the value of liquidity could represent a large type of risk premium in financial markets. The results of his paper *Valuing thinly traded assets,* were several insights in the effects illiquidity has on the pricing of a security; First, the value of immediacy (the ability to sell immediately) is much higher than the value of future liquidity. Second, Longstaff was able to confirm that the value of illiquid assets is often heavily discounted in the market. Third, the effects of illiquidity and volatility on asset prices are entangled fundamentally, which helps to explain why concerns about market liquidity become much more central during financial crises and periods of market stress where volatility can be expected to be higher driven by new parameters of insecurity. Finally, the results show that the discount effect on prices is smaller when the security pays higher dividends and/or cash payouts to investors and thereby reduces welfare effects of holding the stock. Which gives the investors strong economic incentives to lobby for increased payouts (Longstaff, 2014).

2.5.2 Company valuation measurements

Market inefficiency has become more apparent with the arrival of high frequency traders, or HFT's. In the current electronic market, where the barrier of entry for trading is significantly lower than previously, there are more buyers and more sellers. In a perfectly efficient market all the traders know everything there is to know, bid-ask-spreads are zero and liquidity is almost infinite, all this however is empirically not the case (Pedersen, 2015). Bid-ask-spreads are wide depending on what stock is examined and liquidity problems are frequent. HFT's provide in this manner, and in doing so make the share price concentrate nearer the "true" or fundamental value. Akin to the wisdom of the crowd, HFT's acts as a large crowd making hundreds if not thousands of guesses as to what the fundamental value is. This means that the share price will start to bounce around faster than normal and tighter around the fundamental value (Pedersen, 2015).

2.5.3 Connection between valuation and liquidity

The impact that liquidity has on stock prices and valuations have been widely debated in recent years. However, the discussions have had slightly varying angels or themes. Many papers have chosen to focus on the transaction costs in relation to illiquidity and how that variable affects the volume traded. In this area, the two economics Vayanos & Vila found that transaction costs can increase the value of liquid assets, but it doesn't have as strong an effect on the pricing of illiquid assets (Vayanos & Vila, 1999). Another widely discussed theme has been the role of asymmetric information. Mark Lang, Karl V. Linx & Mark Maffet studied the relationship between transparency, stock liquidity and valuation across countries. In their extensive research they found that transparency was significantly associated with lower transaction costs and higher liquidity. In their conclusion they also stressed how companies could use transparency as a strategy to lower the firm's cost of capital (Lang et.al, 2011).

2.5.4 Analysis practice

In *Introductory Econometrics for Finance,* presented the following steps to create an econometric model; (1) Find previous studies or theories on the subject, (2) Formulate an estimable theoretical model, (3) Collect data, (4) Check if model is statistically adequate - if it isn't reformat the model by going back to stage 2 or 3. If the model is statistically adequate, move on to stage (5) Interpret model, (6) Use model for analysis (Brooks, 2014).

2.5.5 Model selection practice

When it comes to the selection process in regards to what regression model to analyze further there exists a plethora of options. What mode of selection is used can impact the data and favour certain statistical qualities more than others (Anderson et.al, 2010). There are several options to consider in this matter, and Benedict Pötscher mentions three main options in his article *Effects of Model Selection on Inference;* Akaike criterion (AIC), the Bayesian method (BIC), and the Hannan-Quinn method (HQ). According to the article, AIC appears to be the superior method to some extent. This is most prominent when compared to BIC and HQ as the AIC tends to, in some regards, perform better in terms of closeness to the selected order (Pötscher, 1991). AIC

selection also results in models in which the parameters are estimated with relatively little bias. However, these models exhibit asymptotic sampling variances that are too small, and achieve confidence interval coverage that is somewhat below the nominal level (Anderson et.al, 2010).

Parsimony has long been a staple in social sciences, and as a rule of thumb aids in decision making and selection in many fields. The principle is however not without its flaws, and is not generally accepted all the time or in all fields of study (Radford, 1995). While these studies have brought forward the inconsistencies and paradoxes that can be created using it in a logical event tree, more recent studies have lended support to the old adage. By interpreting the complexity of the rule geometrically, they have even gone so far as to conclude that the bayesian selection model and parsimony often coincide, and in fact both select the models closest to the truth (Balasubramanian, 2005).

2.5.6 Illiquidity spirals

Liquidity spirals have been observed for a substantial amount of time. Basically described as an adverse feedback loop that makes share prices fall, liquidity dry up and bid ask spreads widen. The onset usually consists of a shock to the market that in turn makes leveraged investors of a particular share lose money. Once these investors reach their thresholds they naturally start to close their positions in the share. The pressure to sell caused by the leveraged investors closing their positions makes prices drop further, thus non-leveraged investors are motivated to close their positions too. From what started as a small shock to the market has now left a share with few owners, illiquidity and a wide bid ask spread (Brunnermeier et.al, 2005). Although common even with light disturbances to the market ecosystem, widespread liquidity spirals were seen in the financial crisis of 2007-2009, making evident that liquidity spirals are contagious, and can move from market to market (Pedersen, 2009).

3. Methodology

In this chapter, the study provides the reader with a clear summary of the scientific approach used throughout the study. It provides the reader with greater understanding for the future results seeing as it clearly explains the choice of methodology, sample collection, timeline, and regression.

3.1 Choice of methodology

In accordance with the purpose of this paper the study herein aims to challenge the wisdom of the crowd by validating the connection between spot value and the fundamental value of smaller non frequently-traded companies. To do so, the authors have used secondary data from Bloomberg's database in order to retrieve the sample information required for the analysis. Furthermore the authors took inspiration in methodology from previous similar research projects made on similar topics. Hence, this paper will include a comparative study realized in quantitative research based upon the chosen sample selection. The methodology is based on quantitative data and the approach has been selected to best answer the research questions. When analysing the relationship between separate variables, quantitative study approaches are quite naturally the most common methodology since they provide a study of the nature of that which this paper is conducted, with clear mathematical connections. In contrast to qualitative measurements, where human validation and unrealized beliefs would hold the ground value of the study, the authors chose to rely on the actual quantitative and verifiable spot market prices. This approach will therefore permeate this study to reach clear mathematical results in order to minimize the bias from subjective or perceived human assumptions about asset valuation and maximize the outcome of these assumptions. The paper also presents a deductive study using current market values in order to question validated theories presented in the chapter for theoretical framework. In regards to an inductive approach, the study can, by adapting this approach, increase its objectiveness when analyzing the relationship between fundamental and spot value, and thereby adopt a more unbiased conclusion. In order to answer the questions that serve as the base for this paper, regression analyzes will be used on the sample data. The

regression analysis will thereby provide clear quantitative outcomes proving how well the chosen market can provide insights into the questions posed herein. Finally, to act as a base for discussion and to further obtain the objectiveness of the study an empirical method is used to explain the results given by the regression analysis. By doing this the authors can mathematically validate the results, rather than conveying personal preferences or making subjective assumptions.

3.2 Sample selection

The sample covers 100 companies all listed on the Swedish stock market. Each company has a maximum market cap of 250 million SEK and a maximum number of 500 shareholders, and the timespan was set to monthly observations of closing rates from November 2019 through October 2020.

3.2.1 Market definition

The study chose to solely focus on shares listed on Swedish exchanges since shares listed on other international exchanges might face different regulation, or be exposed to other external factors. If so, the quality of the produced empiric data would go down, as the sample would then be more heterogeneous, meaning that (1) there might be other variables more relevant for explaining value or (2) the influence of the existing variables would not be the same across different capital markets. The selection is further aided in this regard as the entire sample consists of stocks listed on less regulated markets; First North Stockholm, Spotlight Stock Market, and Nordic SME. If this were not true, differences in market regulations could appear within the data, that nonetheless would not get captured by the analysis. As such the homogeneity of the chosen markets aid in model quality as well as in improving the comparative analysis.

3.2.2 Sample size

A sample size of 100 was chosen as the size needed to satisfy two criteria; be as large as possible in order to produce as robust results as possible, and be small enough that time spent on

collection and data management does not outweigh time spent on interpretation and analysis. With these criteria and help from the supervisor the size of a 100 was determined.

The authors of this study decided to draw the line at 250 million SEK and 500 shareholders to first (1) provide clear guidelines to how this study defines “small companies” and second (2), they found that these limits kept the companies on the market from variating too much, but still provided the right amount valuable shares to operate on. The market cap should be as small as possible as the aim of the study is to investigate pricing methodology concerning small companies. However, at the same time, the market cap should be as large as possible in order to enable a truly random sample selection. The risks associated with drawing the line at too small a market cap are mainly that the sample size decreases as the maximum market cap decreases as well as the risk of there being possible biases within the sample increasing. Reconciling these two things made it clear that drawing the line at 250 MSEK was to be regarded as a near optimal position.

3.2.3 Timespan

Another crucial limitation to make the research relevant was to limit the study to a specific time frame. The stock market moves constantly every day and it is important to use distinct points in time when comparing a sample of companies and shares in order for them to be subjected to the same macroeconomic variables. The study aims to provide the reader with relevant information about today's market relationships with a contemporary timespan. The study is therefore focusing on the time period spanning from November 30, 2019 – October 31, 2020; the logic behind this being the above mentioned need to be contemporary coupled with the fact that a larger time frame would risk the prevalence of potentially a higher share of other variables impacting the data. The fact that a month is missing means that the variables will be lagged, thus helping in limiting and minimising autocorrelation. This study chose the monthly time frame, as opposed to annual data stretching over several years, since the data would then represent averages from the trading years and thus ignore seasonality and increase the risk of inadvertently ignoring the valuable insights provided by day to day variations of the variables.

3.2.4 Timescope and Covid-19

Although not a focus of the study, the data used did take place during the time of the global Covid-19 pandemic. The virus spread started to make headlines in November 2019, but did not have a large impact on the global stock markets until months later. The market's response has also shifted, from an initial crash followed by everything from bearish stagnation to bullish upturn. Given this, it is very difficult to define when and how the spread of the Coronavirus had an impact on the markets, and even more difficult to isolate any Covid-specific effects.

To avoid the impact of Covid-19, this study could have chosen to study data from 2019 (prior to Covid). But that would have yielded a result which would have been out of date; any findings would show results for the market previous to the pandemic, which would have been less interesting since the world has changed. As mentioned in previous studies the effects of illiquidity and volatility on asset prices are entangled fundamentally, which helps to explain why concerns about market liquidity become much more central during financial crises and periods of market stress where volatility can be expected to be higher driven by new parameters of insecurity. The effect of Covid-19 will be present for years, most likely decades, for example, through the enormous debt that has been accumulated, sector rotations in the stock market, and major every day-changes, such as working remotely.

Hence, the authors believe using the time-period selected yields the most interesting results even though it is not perfect. In this study the impact from Covid-19 is difficult to isolate, due to the belief that isolating Covid-impact will be close to impossible in many years to come, caused by the absence of a proper frame of reference. It is no longer possible to view the future of markets as "normal" with a short-term corona-impact, as it is a new world which includes changes driven by the Covid-19 crisis.

3.2.5 Retrieved data

To retrieve the relevant data relevant to perform the study the authors collected data from databases provided by Bloomberg and retrieved through Bloomberg Terminal. The Bloomberg

terminal is a commonly used, well validated and commonly accepted platform for finance research since they include real time data from relevant markets, as well as their own analytics, in-depth research and tools for market communication (Bloomberg, 2020). The choice of data source was based on facilitated availability and access. Firstly, the economics department at Lund University granted the authors free access to the terminal, which allowed the authors to access the information within reasonable measurements for this study. Secondly, the study needed in-depth information about each company's financial statement, and Bloomberg Terminal held the information needed.

3.2.6 Data management

The data is collected for each stock in the sample as defined by the closing rate on the last day of each month within the selected time frame. Thereby, the data used in the regression analysis have characteristics of cross-sectional as well as time series data and could thus be classified as panel data. Using panel data instead of only time-series or cross-sectional data provides three important advantages; (1) Being able to approach more complex relationships, (2) The option to examine how variables, and their relative co-dependencies, changes dynamically over time, (3) The ability to eliminate the impact of certain forms of omitted variables bias in the regression, by using fixed effects (Brooks, 2014).

3.3. Hypotheses

Since this paper aims to find variables that to the greatest extent explains valuations that differ from the fundamental value of securities, the following hypotheses will be used in the regression analysis:

1. First Hypothesis

   H_0 : transaction costs is not a variable correlated to the ratio between fundamental value and valuation
   H_1 : transaction costs is a variable correlated to the ratio between fundamental value and valuation

2. Second Hypothesis

   H_0 : information asymmetry is not a variable correlated to the ratio between fundamental value and valuation
   H_1 : information asymmetry is a variable correlated to the ratio between fundamental value and valuation

3. Third Hypothesis

   H_0 : volatility in stock price is not a variable correlated to the ratio between fundamental value and valuation
   H_1 : volatility in stock price is a variable correlated to the ratio between fundamental value and valuation

4. Fourth Hypothesis

   H_0 : illiquidity is not a variable correlated to the ratio between fundamental value and valuation
   H_1 : illiquidity is a variable correlated to the ratio between fundamental value and valuation

3.4 Regression analysis

To determine whether there is a reasonable relationship or dissociation between fundamental and pricing values of illiquid and low volume shares' this study will use a multiple linear regression analysis. The choice of method was made based on the limited time scope to execute the study, as well as the willingness to reach a conclusion to the purpose of study. The multiple linear regression analysis supports the study in establishing if any linear relationship between the chosen dependent, and independent variables exist (CFI 3,2020). In the preparations leading up to the topic to be analyzed in this paper, the study did not find any studies having used this method to answer the questions posed herein. Instead, the study has followed the 6-step econometric model path presented in *Introductory Econometrics for Finance.* However, considering the time frame for this study, going back to previous stages isn't possible in the extent the study might require. Therefore, some statistical flaws are to be expected, but the author's ambition is that the regression analysis will nevertheless provide a foundation for the later analysis and discussion (Brooks, 2014).

3.4.1 Dependent variables

In both regressions, valuation is set as the dependent variable. However, there will be a different multiple used as a proxy for valuation in each regression. The first regression will use *enterprise value over book value* and the second will use *enterprise value over sales*. The data was extracted from Bloomberg on a monthly basis where enterprise value is computed for every month, meanwhile book value & sales is based on quarterly and annual reports. The multiples were chosen since they show if a security is over- or undervalued as a ratio, while also capturing different types of values; value stemming from assets or from sales respectively. Enterprise value (in this paper defined as the market capitalization adjusted for debt and cash on the balance sheet) over book value (in this paper defined as assets subtracted by liabilities on the balance sheet) is relevant for asset heavy companies, while knowledge intensive companies could be seen as grossly overvalued when applying this ratio. Enterprise value over sales is a better multiple for analyzing the latter, but also has its limitations in that it is based on past cash flows rather than those expected in the future. Therefore, the use of regression analyses with carefully selected variables that are tested for relevance, should provide more robust results.

3.4.2 Explanatory variables

The explanatory variables chosen for this study can be divided into two categories, proxies and determinants. The proxies are aiming to measure variables that are neither directly quantifiable nor easily defined, e.g., illiquidity and information asymmetry. The determinant variables, on the other hand, are directly quantifiable factors, e.g volatility and transaction costs. As explained in previous research, there is no singular perfect method of measuring illiquidity. The chosen proxy for illiquidity is based on trade based measures presented by Aitken & Comerton-Forde (Aitken et.al, 2002). By using average traded volume over shares outstanding for each observed month, a turnover ratio, that can be analyzed across securities with differing amount shares, is produced.

The statistical measure kurtosis, defined as a measure of probability in regards to more extreme outcomes of a distribution, is used as a proxy for information asymmetry. Graphically this can be thought of as how big the tails of a distribution is. While volatility is a measure of the variance of a stock price , kurtosis measures the probability of extreme outcomes. The application of kurtosis

as a proxy for information asymmetry is based on logic rather than previous research. Specifically, companies that are illusive when it comes to information should logically see more drastic price movements than their well known counterparts. The reason for this being that big companies with a strategic and professional IR function, with regular investor meetings and earnings calls, tracked by a multitude of analysts and with professional as well as private investors, are thoroughly analyzed and generally provide relevant preemptive market guidance can not surprise as much. If the company is headed for a record breaking quarter that would have been pre-announced through press releases of changed guidance, or other events that the rational investor on a regulated market can expect to receive information on; hence allowing for investors to incorporate the news into the price they are willing to trade at. For these companies the Wisdom of the Crowds will prevail, and the share price is the mean in a normal distribution of fact based ”guesstimates”. Given that there is a sufficiently extensive supply of relevant information, some analysts and private investors that are paying attention and “reading the clues provided” will have “guessed” correctly in that the company is heading for a never before seen quarter, thus leading to the mean or share price moving upwards. In companies where information is hard to come by, the news comes more as a surprise, thus leading to a sharper or faster correction in share price. As such, when looking at the daily share prices, kurtosis should reasonably be higher for this sample than for larger more frequently traded companies as they instead see their mean drift slowly. In this paper, kurtosis has been calculated for each month using daily price movements.

Volatility has been computed as the standard deviation of day to day logarithmic price changes, to capture the magnitude of relative price changes. With the time parameter of a trading year the volatility subsists of the annualized standard deviation of relative price changes for each shares' daily closing price, sequentially expressed as a percentage of the day prior to the day of computation (Bloomberg, 2020).

Transaction costs related to the trading of shares is a function of supply and demand. If one has a substantial amount of illiquid shares there is a lowered chance of someone showing up and buying said shares at the offered price as perceived by the seller to be a representation of fair market value. This problem is one of the reasons market makers exist, to lower the transaction

costs by offering better spreads, something that is essential for hedge funds as well as wealth managers. The transaction costs variable was calculated for each month in the form of the average bid-ask spread over the closing price of the last trading day of each month. This generates a quota which could be used to study securities with different prices. This was described by Aitken & Comerton-Forde as an order-based approach to measure liquidity. Thus, it could be seen as this study uses two different proxies for liquidity (Aitken et.al, 2002).

3.4.3 Dummy Variables

To be able to run various regression tests and distinguish differences across industries and valuations this study used two dummy variables; industry and if a security is over or under-valued according to enterprise value over book value. The industry dummy was used to compare companies in asset heavy industries to service companies where, generally, assets are not as common. This will allow the regression to generate a more diversified result for the upcoming analysis, since the dependent variable favors asset heavy balance sheets.

Given that EV/B as a ratio has a quantifiable border where companies become overvalued, the over-under-valuation dummy was created. This was done by dividing the sample into one group with an EV/book > 1, thus considered overvalued, and one group with an EV/book < 1, thus considered undervalued. The dummy variable was only used based on the dependent variable EV/book since the multiple provides clear guidelines on the current valuation of a security. The dummy will further bring a more nuanced result for the later analysis. When considering EV/sales as a measurement if a security is under or overvalued, no such guidelines exist that could be compatible across industries.

By using two dummy variables in the regression analysis it allows for testing the impact of the explanatory variables in different subsets of the data. Impacts that usually could get cancelled out and become part of the statistical noise can instead potentially be brought forward and highlighted.

3.4.4 Regression line

To estimate the regression line *ordinary least squares (OLS)* is used. This method is commonly used in econometrics and estimates a straight regression line where the variance from the observations are minimized. Since ordinary least squares estimates a straight regression line, there's limitations to the validity of the outcome when correlation between variables is not linear. To test the assumption that OLS is a good fit for the model, a test for general misspecification of functional form will be executed. By using quadratic terms of the fitted values in the regressions, the output will be another regression model where the significance values and statistical diagnostics will either reject or support OLS as a good fit (Brooks, 2014)

3.4.5 Residual Analysis

When creating a linear regression model, three assumptions are made: (1) The collective average value of all errors is zero. This assumption relates to the regression line, or more precisely its intercept, but is rarely violated when running the model without interfering with the intercept. If the assumption is violated the model's coefficient will be biased, making model fit measurements irrelevant since the mean of the dependent variable won't be equal to the mean of the fitted values. (2) The assumption of homoscedasticity. This means that when a regression model is created, potential errors are assumed to have a constant variance. If the variance is not constant throughout the model, the model is classified as heteroscedastic. When using OLS in a Heteroscedastic model, the estimation will still generate unbiased coefficients, but the coefficient standard errors will no longer hold. As the heteroskedasticity test is not built-in with panels and is not always working together with fixed effects, this study assumes, as it is common in finance research, that the data is heteroskedastic and uses robust standard errors for all models. Which will result in a more conservative p-value, even in the presence of autocorrelation (Brooks, 2014) (3) The last assumption is considering OLS in that the error terms are not correlated with each other. If that happens, the errors are autocorrelated. As with heteroskedasticity, relevant tests cannot be performed accurately for unbalanced panel data, but once again, by using robust standard errors, the p-values will be more reliable in spite of autocorrelation.

3.4.6 Multicollinearity

A prerequisite for successfully using OLS is that the explanatory variables are not correlated with one another, a relationship described by Brooks as being orthogonal. If this relationship requirement is met, adding or removing variables from the model won't change the coefficients of those left to remain in the model. However, this is rarely the case, and a small correlation between explanatory variables are to be expected. When variables become more closely correlated though, a problem called multicollinearity emerges. Generally, when a model suffers from multicollinearity R^2 will be high, but the individual coefficients will have a high standard error, thus leaving the model looking good, but where the individual variables are not statistically significant. The three problems multicollinearity brings is (1) Difficulties when observing the individual contribution of each variable. (2) A sensitivity model where adding or removing explanatory variables greatly changes coefficients and significance of other variables. (3) Confidence intervals will be very wide, and the significance test might give inappropriate conclusions. Measuring multicollinearity in a precise way is difficult, but a correlation matrix of the variables could be used as a substitute. A correlation matrix allows the observer to simply study the presence and easily detected forms of multicollinearity and will therefore be the test used in this paper (Brooks, 2014).

3.4.7 Model selection

To test how well the model fits the data, adjusted R^2 will be used. Adjusted R^2 is based on R^2 which is defined as the square of the correlation between the observations of the dependent variable and the fitted value from the regression model. The adjusted R^2 differs from its unadjusted peer in the sense that losses in degrees of freedom, associated with adding another variable, is taken into account, and subsequently falls. A value close to 1 means the model fits the data excellently, meanwhile a value close to zero suggests the opposite. Adjusted R^2 is widely used because it is not complicated to calculate, effortlessly interpreted, and gives a good overall understanding of how well the model fits the data (Brooks, 2014).

In addition to the adjusted R^2, the Akaike criterion is used. The criterion is a measure of the number of explanatory variables used in order to predict the dependent combined with a measure

of how well the model fits. More variables used means a higher and worse score, while fewer variables used means a lower and better score. As such, given that models are suited for comparison and in conjunction with chapter 2.5.5, AIC is a chosen method for selection.

Finally, the comparability and parsimony of each model relative to its peers will be evaluated. Seeing as adding different effects to the regression such as cross or period fixation, comparability between the models will be sacrificed if the selection used is varied in form and sample size. As such, given that differences in adjusted R^2 and Akaike are relatively small, the primary method of selection will be comparability in tandem with parsimony. Parsimony, all though not as apparent in the models, can be used in the sense that simpler models will in general be prefered over more complex ones, in accordance with chapter 2.5.5 (Balasubramanian, 2005).

3.4.8 Regression formula

Considering what has been discussed previously, the following formulas were used for the upcoming results:

$$EV/BOOK_{it} = C + \beta_1 * Transactioncosts_{it} + \beta_2 * Informationasymmetry_{it} + \beta_3 * Volatility_{it} + \beta_4 * Liquidity_{it} + u_{it}$$

$$EV/SALES_{it} = C + \beta_1 * Transactioncosts_{it} + \beta_2 * Informationasymmetry_{it} + \beta_3 * Volatility_{it} + \beta_4 * Liquidity_{it} + u_{it}$$

Where *EV/BOOK* and *EV/SALES* are the dependent variables, *C* the intercept and *u* the error term.

For each dependent variable, five strata of regression models were produced separately. Each stratum consisted of; (1) regression without dummy variables, (2) regression with service companies, (3) regression with asset heavy companies, (4) regression with undervalued companies, and (5) regression with overvalued companies. As a result of the usage of panel data in the regression, each stratum consists of one model for each panel option; (1) non-fixed, (2) period-fixed, (3) cross-fixed, and (4) both-fixed. Using the above described adjusted data sets resulted in eight regression models for each stratum, generating a total of forty models.

3.5 Method discussion

3.5.1 Chosen parameters and variables

There is, as always, a multitude of parameters, variables, and methodologies that could potentially be altered or adjusted in as far as the methodology used in this study. Any such change could inevitably impact the outcome of the study. One such parameter is the chosen number of variables.

For the purpose of this analysis the authors have opted to base it on only four explanatory and two dependent variables. The reasoning behind this, potentially limited number of variables, is logical in nature; frequently traded large-cap stocks are more complex than small less frequently traded small-cap stocks. However, the first group of companies and their stocks are generally covered and continuously analyzed by a very large number of professional and private investors. Any glitch in pricing is eliminated almost instantly by arbitrage or other corrective behaviors oscillating the stock price back to a small spread. Measuring, or quantifying, if some – or even quite a few investors act irrationally or get their math wrong – is almost impossible given the almost instant elimination of any such potential effects.

For less frequently traded small-cap stocks however, the number of parameters affecting prices are often distinguishable and hence measurable – and at best predictable. There are generally no analysts covering them (with the exception of analyses paid for by the company or newsletters provided by professional shareholder associations such as Aktiespararna) thus not swaying investors to use their valuation. The number of shareholders are fewer and trades occur more seldom. Even if a hypothetical investor gets a hunch that a company will release a stellar financial report and thus buys a lot of stock, it might not even pay off.

Taking all this into account it makes sense that small and less frequently traded stocks require fewer variables to explain why they are priced the way they are.

3.6 Reliability and validity

3.6.1 Reliability

Reliability is a term commonly used to determine how well one specific study can be replicated, and still result in the same outcome. The higher the reliability, the more independent the study is from its authors or executors, which consequently means it could be replicated by studies following the same instructions and procedures. Reliability is often considered of higher importance in quantitative research, since executors want to determine the stability of a measure (Bryman et.al, 2019. p.46). In order to increase the reliability of the study, the authors presented in-depth information to how the study was conducted. The study shows a clear regression formula which was conducted through the public econometric program EViews, giving potential replicators easy access to the same formula and econometric program. Furthermore, all data was collected from the Bloomberg terminals, a commonly used, well validated and commonly accepted platform for finance research. Because of this the authors believe that the data provided on the terminal is true to reality, and that data found on other well respected platforms would show the same results.

3.6.2 Validity

Measurement validity is commonly defined as the measurement of how well a study succeeded in measuring the phenomenon which it was meant to capture. In other words, if spotlights the absence of measurement errors (Bryman et.al, 2019, p. 46-47). To further increase the study's validity, several alternative models were applied and compared in order to distinguish the most applicable one for this study. Furthermore, the sample was divided based on industry and valuation, in order to capture potential differences in how the variables explain the two valuation multiples. In continuation, the study has taken into account the impact of outliers and compared their effects on the various regression models' output. in the end describe the final models. Finally, the study was conducted in the previously mentioned econometric program EViews - a program that is well used and highly trusted to increase the chance of econometric calculator errors.

4. Results

In this chapter, results of the study are presented. It starts with descriptive statistics of the data and continues with the assumptions, and statistical effects present in the regressions. The chapter leads up to the presentation of the regression models used for the later analysis.

4.1 Descriptive statistics

The sample as a whole showed drastically varying values depending on the variable and the company. For example, the maximum value of the enterprise value over sales ratio was around 4304 while the minimum was roughly 0.1. Depending on the variable, some deviations turned out to be extraordinarily high. As such, more of the following analysis would be spent explaining the more extreme company variables, rather than the cluster of company variables centered around the average. This, coupled with high spreads in almost every variable, suggested that some data alteration could be beneficial to the study - in order to provide a more normally distributed data curve.

EVB	EV Book value	Transaction costs	Information asymmetry	Volatility	Liquidity
MEAN	6.146679	0.084290	0.062874	92.29205	0.056316
MEDIAN	2.809191	0.042077	-0.496275	84.57577	0.024575
MAXIMUM	695.5988	2.972367	12.87878	329.4530	2.191727
MINIMUM	0.209193	0.001798	-2.194350	21.60257	0.001178
STD. DEV.	27.16482	0.234156	1.854694	41.91793	0.141676
SKEWNESS	22.87865	7.776401	2.702396	1.540568	9.712802
KURTOSIS	576.8309	72.72885	13.66143	8.402176	120.6507
SUM	4437.902	60.85759	45.39493	66634.86	40.66034
SUM SQ. DEV.	532045.6	39.53178	2480.162	1266878.	14.47188
OBS.	722	722	722	722	722

Table 1: Descriptive statistics of EV/Book value pre-adjustments of outliers

EVS	EV sales	Transaction costs	Information asymmetry	Volatility	Liquidity
MEAN	34.36533	0.116650	0.072828	92.92768	0.059409
MEDIAN	1.954487	0.043515	-0.477098	85.35401	0.026568
MAXIMUM	4304.030	3.977087	12.87878	329.4530	2.191727
MINIMUM	0.100195	0.001798	-2.194350	21.60257	0.000466
STD. DEV.	228.1926	0.375649	1.809927	42.07456	0.142840
SKEWNESS	14.58570	6.890360	2.458051	1.507323	9.361137
KURTOSIS	241.4379	55.65710	11.82044	8.157786	114.5532
SUM	25189.79	85.50422	53.38284	68115.99	43.54665
SUM SQ. DEV.	38116619	103.2943	2397.912	1295837.	14.93522
OBS.	733	733	733	733	733

Table 2: Descriptive statistics of EV/Sales pre-adjustments of outliers (difference in observations used in the two regressions is a consequence of data points missing from one dependent but not the other, a discrepancy caused by the information availability)

The data was examined in order to determine what gets omitted and what stayed in the sample. The alteration considered would have to satisfy a compromise between two opposing factors; (1) keeping as many observations as possible and (2) as many extreme outliers removed as possible. An optimum tangent between both of the factors would provide the most observations per unit of quality. Furthermore, there was a desire to modulate the data uniformly, more specifically, to omit the same relative piece from each variable. If possible, the removal of relatively identical pieces would provide statistical power to the results at the cost of data variability. Remove too much however, and predictability is forced upon the data, statistically insignificant variables can suddenly become significant. Upon examination it was found that a removal method that satisfies all the above was to remove all data points that lied further than two standard deviations away from the average in both directions. This did not only help in providing statistical power, but also kept enough data points to maintain the variables inherent uncertainties to an acceptable extent. The above resulted in the following:

EVB ADJ	EV Book value ADJ	Transaction costs ADJ	Information asymmetry ADJ	Volatility ADJ	Liquidity ADJ
MEAN	3.629546	0.047593	-0.168048	89.49299	0.042476
MEDIAN	2.606847	0.040227	-0.523488	83.69660	0.023533
MAXIMUM	19.33851	0.193023	4.878729	207.9210	0.482043
MINIMUM	0.209193	0.001798	-2.194350	21.60257	0.001178
STD. DEV.	3.365037	0.035121	1.284333	37.80855	0.055314
SKEWNESS	2.150289	1.451519	1.314506	0.690881	3.406642
KURTOSIS	8.705321	5.704629	4.772419	3.066998	18.20903
SUM	2301.132	30.17374	-106.5425	56738.55	26.92952
SUM SQ. DEV.	7167.758	0.780807	1044.141	904865.1	1.936777
OBS.	634	634	634	634	634

Table 3: Descriptive statistics of EV/Book value post-adjustments of outliers

EVS ADJ	EV sales ADJ	Transaction costs ADJ	Information asymmetry ADJ	Volatility ADJ	Liquidity ADJ
MEAN	3.377870	0.048992	-0.156695	89.09318	0.046624
MEDIAN	1.553667	0.041667	-0.490608	82.50911	0.026877
MAXIMUM	23.98918	0.193023	4.878729	207.9210	0.482043
MINIMUM	0.100195	0.001808	-2.194350	21.60257	0.001178
STD. DEV.	4.205955	0.035465	1.297853	38.72461	0.057975
SKEWNESS	2.195311	1.468659	1.349233	0.748584	3.043008
KURTOSIS	8.740949	5.753539	4.907577	3.066282	15.07009
SUM	2023.344	29.34612	-93.86026	53366.82	27.92768
SUM SQ. DEV.	10578.65	0.752153	1007.285	896757.9	2.009939
OBS.	599	599	599	599	599

Table 4: Descriptive statistics of EV/Sales post-adjustments of outliers

The new, adjusted, data sets showed improvements when it came to deviation and started to resemble a more normally distributed set. Seeing as the trimming of the data resulted in improved clarity without sacrificing too much inherent randomness, the sets were chosen to act as the basis for the regression analysis.

4.2 Model selection and regression

The models varied drastically in their quality and produced correlations to a varying extent. In consideration of the described method found in chapter 3.4.7, the selection of models was governed by a combination of two tangible factors, and one more intangible one. More precisely, the selection resulted in "the adjusted R^2", in combination with "the Akaike criterion" as tangible

factors, and “the comparability of a model relative its peers” as the more intangible one. When taking all these factors into account it was found that the cross-fixed models produced the best results. The cross-fixed models did not only have the best scores according to both adjusted R^2 and Akaike, but more importantly had the best comparability and were the most parsimonious. Exclusively using the cross-fixed models cancels out the addition of a plethora of dummies found in the both fixed models, that in turn, were found to not contribute enough to make up for their deduction in comparability and parsimony.

Summarily the cross-fixed models will serve as the regression models used in the analysis and answering of the research questions, justified by their statistical quality, degree of comparability and their level of parsimony All other models, as well as the descriptive statistics for each subset, can be found in the appendix 3.

4.3 Regression diagnostics

To cope with the assumed heteroscedasticity and autocorrelation within the models, all regressions were run by using Whites cross section. Whites cross section utilizes robust standard errors when running regressions and limits the impact of both heteroscedasticity and autocorrelation. In addition to fixed effects, the lagged values used in the regression account for reverse causality, which is known as a cause for endogeneity.

As provided in the appendix 2, the heatmaps shows the correlations between the different variables used in the regressions. As expected, there are some correlations between the variables. However, most of them are not significant enough to affect the models and should not be considered a problem. Furthermore, the heatmap distinguishes a correlation between the to variables transaction costs and liquidity in the regressions. However, dropping one of the correlated variables didn't change significantly the results, and therefore, based on theory and previous literature, this study chose to keep both.

As this study assumed the relationships are realistically non-linear, the test of adding quadratic terms were run on all regressions. The output can be seen in appendix 1, where the models did not significantly improve by this method and the assumption that OLS is a good fit were accepted for all regressions.

Furthermore, the lack of significance for the explanatory variables makes the regression models as a whole, unusable to predict and/or explain movements in the two valuation ratios chosen as dependent variables. Considering this, going back to an earlier stage in the process of building the regressions with the knowledge obtained and attempting to create more precise models would be the optimal choice. However, given the time constraints, this study is better served by stating the anomalies above as a consideration when interpreting the results.

4.4 Enterprise value over Book as dependent variable

Pertaining to the first dependent variable; EV over Book value, the following results were found:

EV/BOOK	Full Sample	Undervalued	Overvalued	Service	Asset Heavy
TRANSACTION COSTS ADJ	-0.989214	-0.233680	0.998933	-7.022557	2.843128
INFORMATION ASYMMETRY ADJ	0.045743	0.001571	0.088994	0.078130	0.020453
VOLATILITY ADJ	0.040314***	-0.002867***	0.034449***	0.034778**	0.041031***
LIQUIDITY ADJ	-0.612597	0.499598**	-1.742875	-0.648840	-0.874675
AKAIKE CRITERION	4.300382	-1.677280	4.416079	4.199331	4.375852
ADJ. R SQUARED	0.656847	0.690176	0.625809	0.668524	0.613065
OBS.	634	117	517	257	377

Table 5: All models include cross section fixed effects, feature one month lagged explanatory variables, are adjusted for extreme values and have robust standard errors.

*** = *significance on a 1% level of confidence*

** = *significance on a 5% level of confidence*

* = *significance on a 10% level of confidence*

4.4.1 Results for EV/Book

Table 5 shows volatility as a statistically significant correlation at the 1% confidence level for all subsets but one; the service company subset, where it showed a 5% confidence level. Furthermore, liquidity was found to be significant at the 5% level for the under-valued subset, while none of the other variables showed signs of significance. Volatility appeared highly prevalent, and featured similar coefficients for all subsets except for the under-valued subset where it instead took on a negative value. Compared to the Volatility coefficients, the significant coefficient for liquidity in the undervalued subset was superior. The R^2 was found to be in a range of roughly 0.61 to 0.69 and averaging at roughly 0.65 for all the subsets of this stratum.

4.5 Enterprise value over sales trailing as dependent variable

Pertaining to the second dependent variable; Enterprise value over sales trailing, the following results were found. As evident by the descriptive statistics the average ratio was found to be roughly 3.4 after the removal of extreme outliers. In conjunction with cross fixed effects the following regression model was produced:

EV/SALES	Full Sample	Undervalued	Overvalued	Service	Asset Heavy
TRANSACTION COSTS ADJ	-1.761780	-1.869348	-1.159606	-5.965507	-0.261257
INFORMATION ASYMMETRY ADJ	0.092684***	0.090262	0.052094	0.061424	0.106127**
VOLATILITY ADJ	0.007636	0.023767	0.010987	-0.015310	0.009804
LIQUIDITY ADJ	-1.118423	1.631705	-2.751981*	-2.550904	-0.323337
AKAIKE CRITERION	3.535788	1.888868	3.642431	3.419535	3.592673
ADJ. R SQUARED	0.897284	0.875661	0.900820	0.931464	0.866841
OBS.	599	103	474	203	396

Table 6: All models include cross section fixed effects, feature one month lagged explanatory variables, are adjusted for extreme values and have robust standard errors.

**** = significance on a 1% level of confidence*

*** = significance on a 5% level of confidence*

** = significance on a 10% level of confidence*

4.5.1 Results for EV/Sales

Table 6 makes it clear that different variables are affecting the valuation in ratio to sales in comparison to book. The regressions show information asymmetry as significant on a 1% confidence in the full sample and on a 5% on the asset heavy subset, with no significance in the other subsets. Regarding liquidity as a variable, significance is found on a 10% interval in the overvalued subset which, even if interesting, does not bring any value for the forthcoming analysis. Also notably different here is the relatively remarkable values of R^2, roughly covering a spread from 0.87 to 0.93 and averaging roughly 0.89.

4.6 Hypotheses outcomes

1. First Hypothesis

   For the transaction cost variable, the null hypothesis was not disproven for any subset of data.

2. Second Hypothesis

   Information asymmetry was proven to be statistically significantly correlated at the 1% level in the full sample, and 5% in the asset heavy subset in the EV Sales regression. Thus disproving the null hypothesis in those subsets. However, the null hypothesis was not disproven for any other subset.

3. Third Hypothesis

   Volatility was found to be statistically significant in the range 1% and 5% for all subsets in the EV Book regression, which disproves the null hypothesis in those subsets. In the EV Sales regression, none of the subsets could disprove the null hypothesis.

4. Fourth Hypothesis

   Illiquidity was found to be significant at a 5% interval within the undervalued subset of the EV Book regression, thus disproving the null hypothesis for that subset. However, the null hypothesis could not be disproven for any other regression.

5. Analysis

In this chapter, the results of the study are analyzed using the given theoretical framework in order to answer the research questions posed. Given the results in the previous chapter a combinatory effort of deduction and induction is used in order to form conclusions.

Given the results there are a number of observations that can be made. The study therefore focuses on the two main topics; chosen theories, and previous research - in relation to the studys' results.

5.1 Analysis pertaining to assets

When analysing the results pertaining to EV/B the two variables producing statistically significant coefficients were volatility and liquidity, where the former was significant throughout the whole sample and the latter variable only in the undervalued subset. This points to the fact that, according to the models, neither transaction costs nor information asymmetry are highly prevalent in shares priced differently than their fundamental values in the form of the EV/B ratio.

5.1.1 Analysis pertaining to theoretical perspective

Combining the results with the theory of the law of one price should entail that companies with the same volatility should be priced the same. If volatility is the primary tool of measurement used by investors to gauge valuations, pertaining to assets, the law of one price entails that one unit of volatility should equal the same unit of valuation, no matter the company. In this sample however this is not the case, with valuations varying widely in spite of similar volatility. As a consequence of this the findings in this study directly stand in opposition to the law of one price theorem.

Correlation of course does not necessarily entail causation, rather it could be the other way around that overvaluation causes volatility, or something else entirely. The possibility for arbitrage, in accordance with *arbitrage pricing theory,* still exists however seeing as the prevalence of statistically significant correlation can be used to exploit mispricing. The

opportunity is limited however, as timing windows can be too long to justify holding severely illiquid assets that are hard to unravel if the need arises.

As to the causes of the shares' more prominent overvaluation, rather than undervaluation, one can only speculate. When taking the efficient market hypothesis into account it becomes troubling, as the theorem suggests that all information that could possibly be extracted about the companies have already been extracted as well as priced into the share price. If that were true, the study failed to capture that information, which in itself is a paradox seeing as the information should be highly prominent and near impossible to miss. Therefore, the consistent overpricing within the sample instead serves as contradictory evidence against the efficient market hypothesis.

Transaction costs not showing any correlation in any stratum or subset is a finding at odds with most of the literature. As it is not evident from the statistical analysis, speculation is the only option as to why that is, but the result in itself is remarkable. One of the possible reasons for this result is a combination of factors. Firstly, institutions and companies dealing in securities often have limits when it comes to the spreads between sellers and potential buyers. As pointed out by Pedersen these limitations are in place to minimize transaction costs, and as such this could mean that these types of investors never deal with the securities in the sample. The investors in that case might instead be only private investors, however, it can be argued that the vast majority of private investors do not break away from the herd. In line with the thoughts of Robert Schiller, most people remain within the herd and trade securities within the “normal bubble”. What possibly could make an investor leave the bubble is a personal connection to the security. Per definition this should lead to a more irrational pricing of the security, and as the pricing becomes less and less connected to the fundamental value, its volatility starts to fluctuate.

5.1.2 Analysis pertaining to previous research

When putting this in the context of previous research, the fact that there is a strong correlation between transaction costs and liquidity, provided in appendix 2, is contradictory to Aitken & Comerton-Fordes paper. In the paper they suggest that both transaction costs and volume act as

proxies for liquidity, and the choice could have a significant impact on the results. Something the correlation between the two variables suggests should not happen. The conclusion that could be drawn from this correlation is that liquidity is heavily dependent on transaction costs.

With crises in mind, it is possible that Covid-19 generated countless liquidity spirals, perhaps even ones that pushed shares into the parameters of the sample for this study. If this were to be true, it would entail that those shares would appear to be undervalued as compared to their fundamentals. Given what is said in chapter 2.5.6, that these types of differences in pricing and fundamental value have been historically linked primarily to illiquidity and wide bid ask spreads, these should reasonably show significant correlation in the undervalued subset. As is made evident in chapter 4.3.1, illiquidity was significantly correlated in the undervalued subset, however transaction costs were not. This finding lends credence to Pedersens theory, as discussed in chapter 2.5.6, that times of crisis bring about bouts of illiquidity. The same can however not be said for transaction costs in the form bid ask spreads.

5.2 Analysis pertaining to sales

When analysing the results pertaining to EV/Sales the two variables producing statistically significant correlations were information asymmetry and liquidity.

5.2.1 Analysis pertaining to theoretical perspective

When comparing the results to the arbitrage price theory and law of one price they are at odds. Whereas these theories see mispricings as temporary oddities that the market sequentially corrects, either by arbitrageurs or private investors in general. The “mispricing windows” do not only appear to be open for longer than should be theoretically possible, they are also seemingly driven by other factors than what should be expected. For example, arbitrage price theory suggests that expected returns tempered by macroeconomic risk should be what closes the window. The results of the study however point to information asymmetry being highly prevalent in mispriced shares, especially asset heavy ones. Something that is not accounted for in either theory.

In the service company and undervalued subset, no statistically significant correlations were found. Although surprising, for service companies, it might actually be consistent with the efficient market hypothesis as well as Pedersens' take on it. Companies with this classification within the sample tended to be consultant orientated companies or pharmacological companies. Viewing these companies through the eyes of a randomly selected private investor, a common question is what the primary thing one considers before making a purchase decision actually is. Seeing as assets are not an essential part of the business and past performance is not indicative of the future, one might guess that projected future performance is the main driver. Naturally this guess is predicated on the fact that the randomly selected private investor can make this distinction, something that should be true if one believes in efficient markets. Projected future income was left unaccounted for in the models, and as such only appears in the form of an exogenous factor.

What is prevalent however since it goes against the theories, is that information asymmetry was significant. Based on the results it seems that informational gaps between insiders and outsiders are linked with companies being overvalued pertaining to their sales. The potential causes for this relationship can only be speculated, but it would make sense if apparent information asymmetry made people stay away from the security. Primarily because it could be risky to hold a security that there exists little information about. As an outsider looking in, insider trades will not make sense if the information causing them never leaves that inner circle, and as such the shifts in value can not be predicted. Looking through the lense of the chosen theories of the study, the market if anything appears to be efficiently inefficient, with some people having, and being incentivised, to find more information than others. This in turn possibly causes the securities of this category to be priced according to what the few insiders know, all whilst the greater market steers clear in blissful ignorance.

6. Concluding remarks

In this chapter, the answers to the thesis are concluded in a concrete and precise way. Successively, potential alterations to different aspects of the study are discussed as to what impact they might have had on the results. Lastly, the analysis is put into context of the previous research, ending with suggestions for further research.

6.1 Conclusion

To answer the research question of this study; *"What variable, to the highest extent within the set models, correlates with the valuation of small cap shares with limited ownership?"*, results varied depending on what stratum and subset was analysed. Beginning with enterprise value over book value; volatility was found to be statistically significant in all subsets at the 1% level in each but one where it was at the 5% level. Liquidity was also found to be significant at the 5% level for the undervalued subset. For enterprise value over sales the results were more varied. Information asymmetry was found to be significant in the full sample and the asset heavy subset at the 5 % level. All the above is summarized in the tables below, displaying what variables were significant in what subset.

EV/BOOK					
	Full sample	**Undervalued**	**Overvalued**	**Service**	**Asset heavy**
Volatility	1%	1%	1%	5%	1%
Liquidity	Not significant	5%	Not significant	Not significant	Not significant

Table 7: Variables that were statistically significantly correlated are featured. All significance levels under 10% are displayed for what variable they stem from and in what subset they appear

EV/SALES					
	Full sample	**Undervalued**	**Overvalued**	**Service**	**Asset heavy**
Information asymmetry	1%	Not significant	Not significant	Not significant	5%

Table 8: Variables that were statistically significantly correlated are featured. All significance levels under 10% are displayed for what variable they stem from and in what subset they appear

6.2 Discussion

There is much to discuss when it comes to different aspects of this essay. Seeing as the chosen topic and subsequent question pertaining to it was not a simple path that required no decision making. Indeed it was a forked road where different decisions had to be made, thus opening up for potential errors and compromises of the validity and integrity of the results.

6.2.1 Choice of proxies for fundamental value and explanatory variables

Perhaps one of the largest deciding factors in the outcome of the essay was the chosen estimates for fundamental value. Predicated on the fact that they needed to be as close to the "real number" as possible. As such, the question of which estimates that capture the inherent value in the best way, was raised. Finally, two variables governed which estimates were chosen; accuracy and feasibility. The first of which goes without saying, the second of which was necessary due to limitations in time and resources.

Besides the multiples, the method of DCF; discounted cash flow valuation was considered - a method that is very commonly used to value stocks. The main benefits with the model in this case is that the computed valuations tend to be more accurate for companies with fewer comparables, as well as often being better for companies that at the time of valuation are running a loss. The potentially biggest downside however is that the model, like any type of forecast, requires assumptions in order to work.

In the case of valuing a single company, perhaps one that already has analyst coverage and plentiful information online, the assumptions will be more robust. For example, forecasting the depreciation expenses ten years into the future will be easier as one can compute an average based on the wisdom of the crowds and a trend line fitted to historical data. Given the sample set, the forecast would in almost every case be based solely on the historical data, thus lowering the quality of the forecast. The problem is further worsened by the fact that an almost moderate sized proportion of the companies were only recently formed, thus in turn leading to low amounts of historical data points to work with.

In the case of this essay, the above mentioned downside was deemed to outweigh the above mentioned upsides, especially once taking time and resources into account. Given more time or resources the option could have been viable, but with the aforementioned constraints coupled with the lack of robustness in the valuations as a consequence of assumptions with low validity, it was not chosen for this essay.

Another huge factor deciding the outcome of this paper is the explanatory variables used in the regression. As with the estimates for valuation, many of the explanatory variables were simple proxies attempting to measure factors that are not easily defined. Thus, the question if a different proxy would have generated a different result is raised. To reiterate, the choice was based on accuracy and feasibility, whereas previous research provided a wide range of choices. Most of them however, where complicated and thus not applicable on the sample size used in the regression. Considering the lack of significance, going back and reconstructing the model with other variables, for example projected future income, might have resulted in more significant correlations. Something that could have been beneficial for the latter analysis and conclusion.

6.2.2 The potential of changed parameters leading to differing conclusions

There is no doubt that changing the parameters of a study can impact the results. It is naturally wrong however to postulate that the hypothetical changes one could make would have equal impact either way. As such it is worth discussing if a change can lead to better and closer to the truth-results.

The dummy variables were defined as asset heavy companies versus service companies, and overvalued companies versus their inverse. Both divisions were made as they directly tie into the chosen dependent variables, but they are not the only divisions possible. Besides these, a division was considered that would separate the outlier companies from the herd; in other words, companies with more extreme values in terms of all the variables as compared to companies with rather similar values. This division was considered as there was a substantial number of outliers that were sequentially removed for the adjusted samples. In the non-adjusted samples they remain, but without separation it is impossible to tell what impact they had on the overall results. Comparing the groups might have shown that as companies approach extreme

over-or-undervaluation, or are extreme in some other regard, they are also judged by new variables as compared to their more reasonable counterparts. This comparison is something that is highly suggested as a possible future angle of research, as it might highlight a swing in relevant factors determining pricing as companies become extreme in some regard.

A plethora of company specific factors were unaccounted for in the model, one of those being leverage. More precisely the dilutionary effect it can have on fundamental value. A heavily leveraged company will have a higher balance sheet total, higher enterprise value as well as a hefty tax shield thus increasing its book value. It is however important to note that the creditors are sequentially entitled to a sizable part of the incoming cash flows, leaving less to be divided among shareholders, something not reflected by comparing the enterprise value to book value or sales. Also omitted is the positive effect of leverage in that it can enable longer reach and open up investment opportunities for companies. If the company sees steady cash flows and has no problems meeting its obligations, leverage allows them to grow at much faster speeds. This positive effect is something that reasonably might impact investors pricing of the shares but also could be argued to be a part of the fundamental value.

6.2.3 Wisdom of the crowd in the context of this study

The study focused solely on the conditions and factors pertaining to shares that could be classified as unpopular or small. But it is interesting to speculate the granularity of said conditions and factors and to what extent they change as companies are traded more and become bigger. The wisdom of the crowd, as a rule of thumb, suggests that many guesses generate better averages than fewer. If applied to the stock market however it is unclear as to where the line should be drawn where “few guesses” becomes “many”. If one were to attempt to capture the hypothetical “entire sample” of companies that have many guesses, it is unclear how big the smallest company of that sample would be.

Seeing as this study found statistically significant correlations pertaining to small shares, a follow-up question can be formulated. If the most prevalent factors found in low volume and illiquid shares are volatility and information asymmetry, and one presumes the opposite side of

the market to be defined by a myriad of factors and potentially the wisdom of the crowd, is there a "medium stratum" of companies that are defined differently as compared to both other groups? If one tried to analyze this "medium stratum", what variables showed high correlation could be either different from both other samples, or perhaps a combination of both. If a "medium stratum" is found, it is also not clear as to where it fits in the different theories. Through the lens of Pedersen; are different strata of the market more or less efficiently inefficient than others? Given the above this topic is suggested as a potential future focus of research.

Although it should be improbable that the majority of the shares within the sample are priced according to speculation bubbles, there at least exists a chance that some of them are. If this is the case, the effects of such kinds of pricing impacts the model, and should do so in a negative fashion. As pointed out by Pedersen in Efficiently inefficient; "*the process of a stock price overvaluation can be significantly amplified when investors start to speculate in the future forecasts of other investors, rather than focusing on the company's fundamentals.*". In the context of the wisdom of the crowds, this can be seen as an evolution of the concept which has dilutionary properties. If part of the crowd starts to modulate their guesses to incorporate their individual ideas of what the other part is guessing, the average guess of the crowd could actually become worse. Like that of a self incorporating feedback loop, eventually the wisdom of the crowd could be nothing more than an average guess as to what the average should be, rather than what it actually is. Once applied to the context of the sample in the study, this dilutionary effect could sequentially be described as part of the error margin when it comes to the pricing side of the variables.

7. Sources:

Aitken M., & Comerton-Forde C.,(2002). *How should liquidity be measured?*, Pacific-Basin Finance Journal

Anderson, D. R. , Burnham, K. P. & G. C. White (1998). *Comparison of Akaike information criterion and consistent Akaike information criterion for model selection and statistical inference from capture-recapture studies*, Journal of Applied Statistics [Online] Available:https://doi.org/10.1080/02664769823250 [Retrieved January 6, 2021]

Balasubramanian, Vijay (2005). *Bayesian Inference and the Geometry of the Space of Probability Distributions* [Online] Available:https://www.sas.upenn.edu/~vbalasub/public-html/Inference_files/MDLChapter.pdf?fbclid=IwAR3ea8SxV9_oEUI6wSaVYH383CKTPhpv7XOtDEvnGdCZOkQxU88jm1vzaTc [Retrieved January 6, 2021]

Berk, J. B. & DeMarzo, P. M. (2014). *Corporate Finance*. New York: Pearson Education

Bloomberg L.P. (2020) Stock price graph for targeted sample. 19/21/11 to 20/01/12. Bloomberg terminal.

Brooks, Chris., (2014). *Introductory Econometrics for Finance*, Cambridge University press, Third edition.

Brunnermeier, M., and Pedersen, L.H. (2005) *Predatory trading*, Journal of finance 60, p. 1825-1863.

Bryman, A., & Bell, E., (2019). *Business research methods*. Malmö: Liber AB

Corporate finance institute 2 (2020). *Arbitrage pricing theory* [Online] Available:https://corporatefinanceinstitute.com/resources/knowledge/finance/arbitrage-pricing-theory-apt/ [Retrieved November 15, 2020]

Corporate finance institute 1 (2020). *Liquidity* [Online]
Available:https://corporatefinanceinstitute.com/resources/knowledge/finance/liquidity/
[Retrieved November 15, 2020]

Corporate finance institute 3 (2020). *Multiple-linear-regression* [Online]
Available:https://corporatefinanceinstitute.com/resources/knowledge/other/multiple-linear-regression/
[Retrieved November 15, 2020]

Fama, E. and Burton G., (1970).Efficient capital markets: *A review of theory and empirical work*. The journal of Finance 25.2, s. 383-417.

Grossman, S., & Stiglitz, J., (1980). *On the Impossibility of Informationally Efficient Markets.* The American Economic Review, *70*(3),

Lang M., Linz KV., and Maffett M., (2011). *Transparency, Liquidity and Valuation: International Evidence on When Transparency Matters Most*, Journal of Accounting Research Vol. 50

Longstaff, Francis A., (2014). *Valuing thinly traded assets*, National Bureau of Economic Research

Longstaff, Francis A., (2001). *Optimal Portfolio Choice and the Valuation of Illiquid Securities,* Review of Financial Studies 14, s. 407-431.

Pedersen, L. H., (2015). *Efficiently Inefficient: How Smart Money Invests and Market Prices Are Determined*

Pedersen, L.H. (2009), *When everyone runs to the exit*, International journal of central banking 5, p. 177-199

Pötscher, Benedikt (1991). *Effects of Model Selection on Inference,* Cambridge University Press [Online]
Available:https://econpapers.repec.org/article/cupetheor/v_3a7_3ay_3a1991_3ai_3a02_3ap_3a163-185_5f00.htm?fbclid=IwAR3NAAIWxDDBIffc4MuvfliKY_4MF2JdylzoacmrxAxifLqZV7J_pxShr3U [Retrieved January 6, 2021]

Radford, Neil (1996). *Bayesian Learning for Neural Networks.* Springer-Verlag New York, p. 103-104

8. Appendix:

1. OLS assumption test

EV/BOOK	Full Sample	Undervalued	Overvalued	Service	Asset Heavy
TRANSACTION COSTS ADJ	-0.989214	-0.233680	0.998933	-7.022557	2.843128
INFORMATION ASYMMETRY ADJ	0.045743	0.001571	0.088994	0.078130	0.020453
VOLATILITY ADJ	0.040314***	-0.002867***	0.034449***	0.034778**	0.041031***
LIQUIDITY ADJ	-0.612597	0.499598**	-1.742875	-0.648840	-0.874675
OBS.	634	257	517	257	377
AKAIKE CRITERION	4.300382	-1.677280	4.416079	4.199331	4.375852
ADJ. R SQUARED	0.656847	0.690176	0.625809	0.668524	0.613065

EV/BOOK	Full Sample	Undervalued	Overvalued	Service	Asset Heavy
TRANSACTION COSTS ADJ ^2	-5.649908	-0.016025	6.515257	-40.54930	13.13684
INFORMATION ASYMMETRY ADJ ^2	-0.026238	0.001566	-0.008952	0.030135	-0.053904
VOLATILITY ADJ ^2	0.000190***	-0,0000121	0.000147***	0.000204***	0.000187
LIQUIDITY ADJ ^2	-2.250862	3.067405***	-3.780247	-4.545474	0.605988
OBS.	634	117	517	257	377
AKAIKE CRITERION	4.281116	-1.629369	4.416535	4.185186	4.349007
ADJ. R SQUARED	0.663395	0.674970	0.625638	0.673180	0.623314

EV/SALES	Full Sample	Undervalued	Overvalued	Service	Asset Heavy
TRANSACTION COSTS ADJ	-1.761780	-1.869348	-1.159606	-5.965507	-0.261257
INFORMATION ASYMMETRY ADJ	0.092684***	0.090262	0.052094	0.061424	0.106127**
VOLATILITY ADJ	0.007636	0.023767	0.010987	-0.015310	0.009804
LIQUIDITY ADJ	-1.118423	1.631705	-2.751981*	-2.550904	-0.323337
OBS.	599	103	474	203	396
AKAIKE CRITERION	3.535788	1.888868	3.642431	3.419535	3.592673
ADJ. R SQUARED	0.897284	0.875661	0.900820	0.931464	0.866841

EV/SALES	Full Sample	Undervalued	Overvalued	Service	Asset Heavy
TRANSACTION COSTS ADJ ^2	-1.761780	-1.869348	-1.159606	-5.965507	-0.261257
INFORMATION ASYMMETRY ADJ ^2	0.092684***	0.090262	0.052094	0.061424	0.106127*
VOLATILITY ADJ ^2	0.007636	0.023767	0.010987	-0.015310	0.009804
LIQUIDITY ADJ ^2	-1.118423	1.631705	-2.751981*	-2.550904	-0.323337
OBS.	599	103	474	203	396
AKAIKE CRITERION	3.535788	1.888868	3.642431	3.419535	3.592673
ADJ. R SQUARED	0.897284	0.875661	0.900820	0.931464	0.866841

Shown as the first regression for each dependent variable is the OLS regression presented in the results and the second is Quadratic.

2. Correlation Matrix:

EVB Full sample	TRANSACTION COSTS	INFORMATION ASSYMETRY	VOLATILITY	LIQUIDITY	ENTERPRISE VALUE OVER BOOK
TRANSACTION COSTS	1,0000	0,0136	-0,0071	1,0622	-0,1895
INFORMATION ASYMMETRY	0,0136	1,0000	-0,0000	0,0091	0,0028
VOLATILITY	-0,0071	-0,0000	1,0000	-0,0039	-0,0010
LIQUIDITY	1,0622	0,0091	-0,0039	1,0000	0,0290
ENTERPRISE VALUE OVER BOOK	-0,1895	0,0028	-0,0010	0,0290	1,0000

EVS Full sample	TRANSACTION COSTS	INFORMATION ASSYMETRY	VOLATILITY	LIQUIDITY	ENTERPRISE VALUE OVER SALES
TRANSACTION COSTS	1,0000	0,0165	-0,1184	0,9625	-0,2984
INFORMATION ASYMMETRY	0,0165	1,0000	-0,0000	0,0222	0,0044
VOLATILITY	-0,0112	-0,0000	1,0000	-0,0067	-0,0015
LIQUIDITY	0,9625	0,0222	-0,0067	1,0000	0,1001
ENTERPRISE VALUE OVER SALES	-0,2984	0,0044	-0,0015	0,1001	1,0000

3. All Regressions

								No dummy variables									
Dependent Variables	EVB ADJ	EVS ADJ	EVB	EVS	EVB ADJ PERIOD FIXED	EVS ADJ PERIOD FIXED	EVB PERIOD FIXED	EVB PERIOD FIXED	EVB ADJ CROSS FIXED	EVS ADJ CROSS FIXED	EVB CROSS FIXED	EVS CROSS FIXED	EVB ADJ BOTH FIXED	EVS ADJ BOTH FIXED	EVB BOTH FIXED	EVS BOTH FIXED
Explanatory Variables and Model quality																	
BID ASK OVER CLOSING PRICE	-2.932639	3.007359	0.326699	-10.27613	-2.528211	4.065353	0.679732	-11.42395	-0.989214	-1.761780	-0.005283	-1.003869	0.320672	0.690113	-0.442214	-2.105164
KURTOSIS	0.112283	0.202474	-0.293333	6.804361	0.131764	0.186835	-0.386703	7.121337	0.045743	**0.092684****	-0.420339	0.522042	0.023118	0.070208	-0.781719	0.686812
ADJUSTED VOLATILITY	**0.020572*****	**0.037096*****	**0.069139*****	**0.512567****	**0.020947*****	**0.037291*****	**0.066440****	**0.512724****	**0.040314*****	**0.00726***	0.017463	0.087751	**0.045415*****	**0.010209****	-0.004832	0.075135
ADJ. VOLUME OVER SHARES OUTS.	**-5.730396****	**-12.86605*****	-5.101809	**-137.1992****	**-5.838437*****	**-13.19955*****	-3.942106	**-143.9084****	-0.613597	-1.118423	3.636325	1.911262	-0.485812	-1.341886	5.194073	12.15206
N.O OBSERVATIONS	634	599	722	733	634	599	722	733	634	599	722	733	634	599	722	733
R SQUARED	0.085478	0.109497	0.009622	0.010728	0.103072	0.114187	0.022823	0.020712	0.694794	0.908105	0.117029	0.941426	0.703408	0.913268	0.172458	0.942385
ADJUSTED R SQUARED	0.078671	0.103500	0.004097	0.005293	0.078813	0.089834	-0.000172	-0.001172	0.656847	0.897284	0.060615	0.934739	0.659249	0.900830	0.060382	0.934613
PROB (F-STATISTIC)	0	0	0.138971	0.090725	0	0	0.463103	0.510039	0	0	0.001181	0	0	0	0.002267	0
DURBIN-WATSON STAT.	0.225266	0.131406	1.763847	0.078157	0.230775	0.129961	1.754305	0.081433	0.615681	0.880541	2.067514	1.111356	0.611063	0.871354	2.068256	1.120069
AKAIKE CRITERION	5.189586	5.609935	9.444518	13.69975	5.208024	5.644721	9.465158	13.72235	4.500382	3.535788	9.477270	11.06680	4.309608	3.518033	9.492038	11.08304
								Service companies									
BID ASK OVER CLOSING PRICE	9.530674	15.12863	**3.098418****	-28.19067	9.996844	12.89101	3.554311	-29.43020	-7.022557	-5.965007	-0.971297	0.213136	-5.202045	-3.622235	0.115682	6.111964
KURTOSIS	0.191705	0.188380	-0.218678	12.38547	0.226279	0.193878	-0.316216	10.92259	-0.078130	0.061424	-0.189006	2.044986	0.054095	0.052703	-0.317381	1.737522
ADJUSTED VOLATILITY	0.071176**	**0.089480*****	0.001559	**1.834756***	**0.017656****	**0.090991*****	0.001890	0.930446	**0.034778****	-0.015310	0.017914	0.066552	**0.040360****	-0.018918	0.027863	0.194844
ADJ. VOLUME OVER SHARES OUTS.	-5.044351	**-14.76118*****	0.569384	-205.2641	-5.850864	**-16.43116*****	0.118755	-208.1317	-0.648840	-2.550904	1.532735	2.389777	-1.506784	-2.981326	1.021978	18.23694
N.O OBSERVATIONS	257	203	295	281	257	203	295	281	257	203	295	281	257	203	295	281
R SQUARED	0.075686	0.465739	0.018024	0.018253	0.109772	0.473169	0.032908	0.047147	0.707369	0.939607	0.647254	0.940897	0.718497	0.943963	0.684294	0.943035
ADJUSTED R SQUARED	0.061015	0.454946	0.004480	0.004023	0.093424	0.427850	-0.022752	-0.010602	0.668324	0.931464	0.625182	0.933001	0.663248	0.931810	0.627239	0.932127
PROB (F-STATISTIC)	0.000520	0	0.218543	0.270911	0.021582	0	0.809770	0.666583	0	0	0	0	0	0	0	0
DURBIN-WATSON STAT.	0.239147	0.215926	0.216159	0.088448	0.238673	0.228185	0.233729	0.096257	0.655889	1.234836	0.617691	1.123007	0.643917	1.225402	0.679978	1.140479
AKAIKE CRITERION	5.147134	5.402502	6.879903	14.60195	5.202845	5.500724	6.945986	14.65749	4.199331	3.419535	5.984527	11.99830	4.253947	3.462906	6.023114	12.04686
								Asset heavy companies									
BID ASK OVER CLOSING PRICE	**-8.506635***	-7.337337	-3.135193	**32.02047***	-8.248947	-6.121910	-3.805260	**32.90228***	2.843128	-0.261257	0.959158	-7.040514	3.750007	1.786663	-1.949526	-0.096115
KURTOSIS	0.013779	0.176611	-0.260829	2.753812	0.024860	0.146237	-0.394179	2.569270	0.020453	**0.186127***	-1.019018	-0.489836	0.005224	0.081366	-1.223411	-0.426579
ADJUSTED VOLATILITY	**0.033840*****	**0.019911*****	**0.118258*****	**0.287609*****	**0.033870*****	**3.310055*****	**0.109273****	**0.287008*****	**0.041831*****	**0.009804*****	0.014971	0.086456	**0.043288*****	**0.013029*****	-0.027469	0.067982
ADJ. VOLUME OVER SHARES OUTS.	-5.375119	-5.686824	-2.156381	-161.9340	-4.829810	-5.339287	4.689836	**-164.7035*****	-0.874675	-0.323537	10.71953	-1.506135	0.067569	-0.118259	18.88726	1.383211
N.O OBSERVATIONS	377	396	427	452	377	396	427	452	377	396	427	452	377	396	427	452
R SQUARED	0.132292	0.033457	0.017167	0.047180	0.140288	0.044984	0.038337	0.056066	0.675315	0.881000	0.141461	0.941261	0.667212	0.888996	0.167750	0.942507
ADJUSTED R SQUARED	0.113865	0.023569	0.007851	0.038654	0.102078	0.034666	0.008008	0.021346	0.613065	0.866841	0.084272	0.934751	0.610192	0.873418	0.041787	0.934189
PROB (F-STATISTIC)	0	0.009732	0.118693	0.000235	0.000003	0.337905	0.432292	0.061489	0	0	0.042409	0	0	0	0.065481	0
DURBIN-WATSON STAT.	0.275240	0.116049	1.821645	0.126157	0.287996	0.113585	1.805388	0.126664	0.599909	0.763392	2.076173	1.144777	0.603427	0.753857	2.075714	1.140279
AKAIKE CRITERION	5.108319	5.495554	9.948054	11.31161	5.152405	5.543962	9.974485	11.35534	4.375825	3.592073	9.992200	8.706701	4.410207	3.583716	10.01731	8.738360
								Undervalued companies									
BID ASK OVER CLOSING PRICE	0.489839	**-7.438118***	-0.050597	-70.78874	0.606062	**-7.723136***	-0.059151	-101.4737	-0.233680	-1.869548	-0.104632	17.12957	0.113903	-1.942330	-0.001468	-49.01672
KURTOSIS	-0.001421	0.100059	-0.001644	-4.990003	-0.003693	0.102720	-0.003038	-5.155219	0.001571	**0.098262****	0.001583	-2.410130	-0.004014	0.078012	-0.003412	-1.95394
ADJUSTED VOLATILITY	-0.000422	**0.027193*****	-0.000265	0.086770	-0.000485	**0.026701*****	-0.000276	0.083113	**-0.002867*****	**0.023767*****	**-0.003191*****	**0.078228*****	**-0.002713****	**0.032591*****	**-0.003044*****	**0.029311***
ADJ. VOLUME OVER SHARES OUTS.	0.046898	-6.156466	0.037636	**-337.7020****	-0.068826	**-7.521608***	-0.083142	**-337.3296****	0.499598	1.631705	0.430545	-45.48619	0.459969	1.201868	0.410647	-25.53561
N.O OBSERVATIONS	117	103	121	117	117	103	121	117	117	103	121	117	117	103	121	117
R SQUARED	0.010853	0.281460	0.004351	0.072664	0.034812	0.338404	0.023816	0.103257	0.751606	0.901269	0.752302	0.735341	0.773395	0.919073	0.774957	0.763001
ADJUSTED R SQUARED	-0.024474	0.252131	-0.029962	0.039337	-0.017298	0.215316	-0.126367	-0.040222	0.690176	0.875661	0.693569	0.673400	0.675479	0.880369	0.682292	0.666733
PROB (F-STATISTIC)	0.872628	0.000001	0.972507	0.074904	0.994888	0.001384	0.999922	0.768109	0	0	0	0	0	0	0	0
DURBIN-WATSON STAT.	0.357129	0.131233	0.351830	0.452190	0.398556	0.305338	0.396651	0.417663	1.339487	0.394736	1.368980	1.314080	1.320314	0.577957	1.391421	1.337311
AKAIKE CRITERION	-0.620238	3.545507	-0.618135	10.72618	-0.441704	3.695951	-0.639532	10.90747	-1.677280	1.888068	-1.495270	5.789702	-1.563956	1.922839	-1.592840	9.884446
								Overvalued companies									
BID ASK OVER CLOSING PRICE	-1.874727	8.900739	-0.092516	-20.61660	-2.186331	9.415343	0.476839	-22.69181	0.998933	-1.159606	-0.632052	1.352405	2.336419	1.573053	-0.212349	-3.906756
KURTOSIS	0.178142	**0.252521***	-0.344705	6.646671	**0.286421***	0.257695	-0.428279	6.980384	0.088994	0.052094	-0.735612	0.920654	0.053409	0.036995	-0.926286	1.040249
ADJUSTED VOLATILITY	**0.070223*****	**0.018450*****	**0.075811*****	0.047188	**0.029558*****	**0.038520*****	**0.072097****	0.453916	**0.034449*****	**0.010907*****	-0.147664	0.024077	**0.036700*****	**0.012587****	-0.103385	0.043239
ADJ. VOLUME OVER SHARES OUTS.	**-6.975186****	**-12.17053*****	-6.342938	-106.4370	**-6.053763*****	**-12.87014*****	-5.072289	-112.0135	-1.742875	**-2.751901***	4.289549	2.007814	-1.449129	**-3.116971***	5.971061	13.48483
N.O OBSERVATIONS	517	474	601	568	517	474	601	568	517	474	601	568	517	474	601	568
R SQUARED	0.094529	0.119235	0.009569	0.007305	0.109591	0.123630	0.027416	0.023853	0.672945	0.912982	0.210471	0.941844	0.684940	0.918369	0.231831	0.942972
ADJUSTED R SQUARED	0.087455	0.111733	0.002922	0.000253	0.081097	0.092947	0.000770	-0.004491	0.623869	0.906820	0.187877	0.933919	0.629679	0.904190	0.211943	0.933604
PROB (F-STATISTIC)	0	0	0.219389	0.388019	0.000001	0	0.425505	0.638021	0	0	0.000006	0	0	0	0.000010	0
DURBIN-WATSON STAT.	0.261768	0.128891	0.773739	0.079014	0.279391	0.132105	0.792014	0.081561	1.640954	0.902431	0.970247	1.148143	0.643708	0.897443	0.999836	1.157513
AKAIKE CRITERION	-5.198450	5.729256	9.622639	13.90498	-5.228678	5.778887	9.644189	13.93042	4.410679	3.642431	9.612242	11.29304	4.425136	3.629157	9.624749	11.31370