Table of Contents

Tables

Figures

1 Introduction

This chapter introduce the reader to the subjects of the thesis: the liquidity premium and exchange traded funds as well as to the problem and purpose. In the end of the chapter, two research questions are stated which are investigated further in the thesis.

1.1 Background

The liquidity premium is a central concept in equity pricing theory. Essentially, the liquidity premium is a return premium demanded by investors for holding illiquid assets. Equities with low liquidity are characterized by higher trading costs due to their wider bid-ask spreads (Amihud & Mendelson, 1986a). As a result of the higher trading cost, investors require a higher return from these investments. Historically, equities with less market liquidity have generated a better return than equities with more liquidity over time (Amihud & Mendelson, 1986b). Because the financial market has evolved since the first discovery of the premium it is of interest to investigate how the premium has developed in recent years.

A noticeable development on the market is the surge of capital into passive investing. The financial market has experienced a shift in capital allocation, from active management to passive management (Investment Company Institute [ICI], 2017). One kind of a passive investment that have gained ground are index funds. The goal of an index fund is to mimic a specified benchmark. The assets in index funds have grown from $11 million in 1975 to $4 trillion in late 2015, representing 34% of the equity funds' market (Bogle, 2016). The rise in popularity of passive investments may be credited to the debate whether active fund managers can generate a better return than a passive index over time, and hence justify a management fee (Frino, Gallagher & Oetomo, 2005).

There are two different types of index funds: mutual index funds and Exchange Traded Funds (ETFs). This study focus on ETFs because of their increasing popularity and because of how they are structured. However, a section with a discussion and comparison of selected features between ETFs and mutual index funds is included. ETFs have gained much attention due to their low fees and their ability to be traded intra-daily on a stock exchange. The net assets of the United States ETF market have grown from $7 billion in 1997 to more than $2,500 billion in 2016 (ICI, 2017). An important distinction between mutual index funds and ETFs is that the price of an ETF is determined by the market participants. To keep the ETF price in line with the Net Asset Value (NAV) of the underlying components, the issuer of an ETF authorizes an external partner to create and redeem ETF shares in exchange for these components. The external partners are called Authorized Participants (APs) and are typically large institutions. When the price of an ETF is above the NAV of the underlying equities the AP buys the securities and create new ETF shares pushing the price of the securities up and the price of the ETF down. If an ETF is priced at a discount the AP buys the ETF and redeem it with the issuer which pushes the price up (ICI, 2017).

When large amounts of capital are being transferred between different securities it could impact their valuations. Coval and Stafford (2007) conducted a study on asset fire sales and purchases and found that underlying securities in a fund that had large out(in)flows experienced a negative (positive) price pressure. In the case of ETFs, this indicates that the valuations of the underlying securities are affected due to the massive capital inflows to these funds. A consequence of this could be that equities which are overrepresented in ETFs have their valuations positively skewed compared to equities underrepresented in ETFs.

ETF funds' performances can be evaluated by assessing their tracking error. The tracking error is a measurement of how well a fund has been able to track its benchmark. The presence of illiquid stocks in the benchmark can make it harder to achieve a good tracking error. This is due to their wider bid-ask spread which increases the cost for the fund (Buetow & Henderson, 2012). Hence, an underrepresentation of illiquid stocks is expected in ETFs. The existence of ETF issuers' ability to "sample" their holdings strengthens the argument. Sampling means that they have the right to choose a representative basket from an index while excluding certain components. It is common to exclude illiquid stocks that would be costly to include (Buetow & Henderson, 2012; Maurer & Williams, 2015). ETF issuers have to assess what impair the tracking error most: 1) Including the illiquid stock, incurring additional costs or 2) Excluding the illiquid stock which degenerates the mimicry of the benchmark. Since liquidity and size tend to be positively correlated, illiquid assets usually only constitute a small portion of a benchmark. Excluding them could in fact ameliorate the tracking error of the fund. Considering the massive capital inflow to ETFs it is possible that it has impacted the liquidity premium.

1.2 Problem

Equities with low liquidity have historically generated excess return in order to compensate for the liquidity risk. If illiquid stocks would not yield excess return they would underperform relative to their risk and would be unattractive to investors. If a liquidity discrimination exists on the ETF market, a consequence could be that the most liquid stocks perform better than they otherwise would and the most illiquid stocks perform worse than they otherwise would. A change in the return distribution between more and less liquid stocks would mean that the magnitude of the liquidity premium has been altered as well. We argue that due to how ETFs are structured, the capital that is allocated to them drives up the prices of the underlying assets, which we believe are mainly constituted by exceedingly liquid stocks. This in turn could have mitigated the liquidity premium that have existed on the stock market which would mean that investors are no longer compensated for the additional risk of illiquidity.

1.3 Purpose

This study has two main purposes. The first purpose is to investigate how the liquidity premium has developed in the United States between 1997 and 2016. This is interesting for two reasons: 1) If a premium still exists, investors that are willing to take on the extra risk of investing in illiquid securities can receive a higher yield, and 2) By doing the study over two decades we are able to spot possible trends in the premium, which in turn can be used for assessing how it might continue to develop. The second purpose is to explore whether possible changes in the liquidity premium can be linked to the capital inflow to the United States ETF market during this period. This has to the best of our knowledge not been done before. Our hypothesis is that the growth of ETFs has mitigated the liquidity premium because ETF issuers include liquid stocks to a larger extent than illiquid stocks.

Our key research questions are as follows:

1. How has the equity liquidity premium developed in the United States between 1997 and 2016?
2. Has the capital inflow to the ETF market had an impact on the equity liquidity premium?

1.4 Outline

The thesis is organised as follows. In Chapter 2 the literature and financial theory associated with the purpose is explored. Chapter 3 includes the philosophies underlying the method, an explanation on how the quantitative study is conducted and how the data is collected. In Chapter 4 the empirical results produced from the study are presented. A deeper analysis of the results is done in Chapter 5. In Chapter 6 conclusions are made based on the results and the following analysis and in Chapter 7 there is a discussion of the results as well as suggestions for further research within the subject.

2 Frame of Reference

The aim of this chapter is to provide the reader with information concerning past research within the subjects of the study. Moreover, the chapter introduce the reader to the financial theories used in the method.

2.1 Literature Review

2.1.1 Market Liquidity

Market liquidity is part of what is called the market microstructure and is an important subject for traders, scholars and regulators. Bodie and Merton (2000) define liquidity as "[...] the relative ease, cost, and speed with which an asset can be converted into cash". Huberman and Halka (2001) define a liquid market as "[...] if one can trade a large quantity shortly after the desire to trade arises at a price near the prices of the trades before and after the desired trade".

2.1.2 Liquidity Premium

An investor who is considering purchasing an illiquid stock needs to, not only consider the firm specific risk, but also the risk of illiquidity. This means that the investor should require a premium for taking on the risk of illiquidity, the risk of having to sell at an unfavourable price due to a lack of buyers. The required compensation is called the liquidity risk premium (Amihud & Mendelson, 1986b).

There are two types of liquidity risks, systematic liquidity risk and characteristic liquidity risk (Bradrania & Peat, 2014). Markets can experience fluctuations in the level of liquidity and investors that purchase stocks which returns are more sensitive to changes in liquidity should require higher expected returns (Pástor & Stambaugh, 2003). These stocks are more susceptible to liquidity market shocks which implies stock prices are more volatile in times of dire liquidity. This higher expected return is explained by the systematic liquidity premium. On the other hand, the characteristic liquidity premium is related to the cost of trading for a specific stock. The characteristic premium is firm specific and does in contrast to the systematic liquidity premium not necessarily have to be affected by market-wide liquidity fluctuations (Ben-Rephael, Kadan & Wohl, 2015).

The origins of the research concerning the equity liquidity premium was made in the 1980s by Stoll and Whaley (1983) and Amihud and Mendelson (1986b). The latter found that market returns have an increasing relationship with the bid-ask spread and that average returns net of trading costs increase with the spread. The increasing returns could not be explained by the firm size effect, which suggests that smaller firms generate excess returns compared to larger firms. They argued that the firm size effect might as well be a result of the liquidity premium which in turn is a response from an efficient market to the existing spread (Amihud & Mendelson, 1986a). In contrast to Amihud and Mendelson, Eleswarapu and Reinganum (1993) found a significant size effect even after controlling for the spread.

The existence of a liquidity premium gained further support from Eleswarapu and Reinganum (1993), however, they found a strong seasonality effect. They presented results for a positive liquidity premium only in the month of January and that the liquidity premium was negative for all other months. The seasonality effect has later received critique by Brennan and Subrahmanyam (1996) and Eleswarapu (1997). The former found no evidence for a seasonality effect while conducting a similar study, whereas the latter found a liquidity premium for the remaining 11 months of the year as well. The reason for the possible seasonality effect is unknown but it could be part of a broader puzzle.

Amihud (2002) solidified the previous research that illiquidity has a positive effect on stock returns. He found that expected stock returns are related to the sensitivity to expected and unexpected market illiquidity. This sensitivity is stronger for more illiquid stocks, which also explain the phenomenon "flight to liquidity", where in times of crisis investors tend to view liquid stocks as more attractive (Amihud, 2002; Pástor & Stambaugh, 2003). The finding suggests that the phenomenon occurs during shorter periods but that the longer trend is the reversed in that illiquid stocks generate excess returns.

Unlike previous studies, Ben-Rephael et al. (2015) found that the liquidity premium had almost vanished. They studied how the liquidity premium had developed in the United States between 1964 and 2011 on NYSE, NASDAQ and AMEX. They found that the premium could be observed in the earlier periods but only on NASDAQ's smallest stocks in the latest test period between 2000 and 2011. Index funds and ETFs are highlighted as a possible source for the vanishing premium with the argument that investors to a large extent have switched from owning illiquid stocks directly to owning them through index structured products which tends to hold stocks for a longer period and consequently incur lower transaction costs. As transaction costs in the form of spreads have been found to be a determinant of the liquidity premium the argument is valid. If it is true, it would mean that it is mainly the characteristic risk premium which has decreased because of ETFs.

Due to the conflicting results between older research which found significant evidence of a liquidity premium and more recent research which have only found weak evidence it is important to conduct further studies exploring why this distinct change has occurred. It could be an effect of more efficient markets, the popularity of indexed products such as ETFs or a reason not yet proposed.

2.1.3 Liquidity Measures

Liquidity is a subtle concept which cannot be seen directly, for this reason researchers use proxies to estimate it. With the help of proxies, liquidity can be roughly measured in a plenitude of different ways. In order for this thesis to use the best measurement based on our aim this section is dedicated to encapsulating some of the methods used to measure liquidity. We can decide between looking at the "tightness" of the market microstructure (the bid-ask spread), the "depth" (the liquidity ratio), or a combination of both. There are strengths and weaknesses for the different options which makes it important to use the most suitable one. The liquidity associated with the tightness of a stock is called the characteristic liquidity and is individual across stocks. On the other hand, the systematic liquidity is associated with the depth of the market (Ben-Rephael et al., 2015).

The bid-ask spread can be seen as the most intuitive measurement of liquidity as it informs a trader what the immediate cost of executing an order is. The cost of immediate execution is the spread between the highest bid offer and lowest ask offer on the market. If a trader would instantly make a trade in a stock at the given bid and ask prices he or she would incur the spread as a loss. Amihud and Mendelson (1986a) used the bid-ask spread as a variable to measure the liquidity premium and found that the average returns are an increasing function of the spread. They found that a stock with a 1.5% spread had a monthly excess return of 0.45% compared to a stock with a 0.5% spread. Brennan and Subrahmanyam (1996) on the other hand found that the spread had a significant negative impact on the returns of stocks. Grossman and Miller (1988) criticized using the bid-ask spread as a measure of liquidity since the cost of the spread only is realised if a trader buy and sell simultaneously. While it is true that the full cost of the spread only is realised if the trade happens simultaneously, the spread can still be used as a proxy to measure the level of liquidity for a stock.

Even though the bid-ask spread has been found to have explanatory power on the liquidity premium, it does exclude the depth of the microstructure. Chalmers and Kadlec (1998) found that using what they called "amortized spreads" had stronger evidence for a correlation between liquidity and return than the bid-ask spread alone. The amortized spread includes the dollar turnover of a share in the model which make it account for both the tightness and the depth of a stock. Even though their amortized spread variable showed stronger evidence for being priced than unamortized spreads, they highlight that the study had a limited sample period between 1983 and 1992.

To understand the source of the spread it is important to know what creates it. The spread has been linked to other variables such as trading volume, stock price and number of shareholders (Stoll, 1989). These variables can be linked to what is called the depth of the bid-ask order book. Kempf and Korn (1999) explain market depth as the relationship between order flow and price changes, and Huberman and Halka (2001) use the number of shares at the bid and ask price (quantity depth) and the dollar value of the shares at the bid and ask price (dollar depth) as measures of depth. These measures can be linked to what is called the systematic liquidity of the market. When there is a negative liquidity shock on the market, stocks that are affected to a larger degree are said to be more susceptible to systematic liquidity risk.

The liquidity ratio is a measurement of the depth and is the most frequently used liquidity measure (Bernstein, 1987). It measures the ratio of the dollar volume of trading by the percentage change in price. Grossman and Miller (1988) criticize the liquidity ratio by arguing that it can only be used to learn about past correlation between the trading volume and price changes. Since this thesis is investigating the historical liquidity premium, the past correlation between price changes due to trading volume is still of interest.

Amihud (2002) suggested an illiquidity measure called “ILLIQ”. It is related to the liquidity ratio as it is a measurement of a stock’s price response to trading volume. Amihud’s result imply that trading volume, all else equal, has a larger impact on the price of more illiquid stocks than on more liquid stocks. He does however argue that the bid-ask spread is a better measurement for assessing liquidity, but that he still used ILLIQ for two reasons: 1) He did not have the microstructure data necessary to obtain the bid-ask spread and 2) ILLIQ is a good measurement to use for time series analysis.

According to Kluger and Stephan (1997), most of the commonly used liquidity measures can be used to conclude that a liquidity premium exists, but that a composite measure, consisting of several measures can explain it in a better way. Chen and Sherif (2016) also highlights the benefits of making a composite measure, arguing that single measures are not reliable. The arguments indicate that liquidity is a multidimensional phenomenon. In Table 1 below, the mentioned liquidity proxies are presented with their formulas and input variables.

Table 1: Liquidity Proxies Used in Different Studies

Liquidity Proxy	*Formula*	*Variables*
Bid-Ask Spread	$(P_a\text{-}P_b)/P_a$	P_a = Ask Price
Amihud & Mendelson (1986a)		P_b = Bid Price
Amortized Spread	$((\|P\text{-}M\|)/P)*(S/S_O)$	P = Transaction Price
Chalmers & Kadlec (1998)		M = Midpoint of Spread
		S = Number of Shares Traded
		S_O = Shares Outstanding
Quantity Depth	S_a+S_b	S_a = Number of Shares Ask Side
Huberman & Halka (2001)		S_b = Number of Shares Bid Side
Dollar Depth	$S_a*P_a+S_b*P_b$	S_a = Number of Shares Ask Side
Huberman & Halka (2001)		P_a = Ask Price
		S_b = Number of Shares Bid Side
		P_b = Bid Price
Liquidity Ratio	$(P_c*S)/(\|R\|)$	P_c = Closing Price
Cooper, Groth & Avera (1985)		S = Number of Shares Traded
		R = Absolute Return of Stock
ILLIQ	$\|R\|/VOL$	R = Absolute Return of Stock
Amihud (2002)		VOL = Dollar Trading Volume

2.1.4 Indexation

With an increasing demand for passive investments and low-cost funds, the role of indices has grown. Indices are no longer only an instrument to measure stock market performance, but also an important investment tool in asset allocation. Furthermore, they play a central role for many funds and derivatives that rely on index replication (International Organization of Securities Commissions [IOSCO], 2003). As the interest for passive investments has increased, more capital has been allocated into funds that follow a particular index. Several studies have been performed to investigate how an inclusion in an index or a rebalance of an index affects a stock's price (Shleifer, 1986; Harris & Guel, 1986; Beneish & Whaley, 1996; Kaul, Mehrotra & Morck, 2000; Chen, Noronha & Singal, 2004; Baran, & King, 2014). In rebalancing situations, passive index funds need to replicate the new weights, which means sell-offs in excluded stocks and purchases in included stocks. These actions lead to massive reallocations of the funds' capital and it is common that an index fund purchase 3% of a recently included firm's outstanding shares (Shleifer, 1986).

When calculating the weights of stocks in a market capitalisation index, index creators usually also consider how liquid the stocks are. The most well-known indices such as the S&P 500 use what is called a "Free-Float adjusted market capitalisation" measure to ascertain what weight a stock should have in an index (FTSE Russell, 2015; MSCI, 2017; Standard & Poor's [S&P], 2018). This means that shares which are held by certain institutions, insiders and other parties that are not likely to sell their shares are excluded or have a lower weight when calculating the market capitalisation. In effect this means that all else equal more liquid stocks receive higher weights in indices. The adjustment made for free-float can greatly impact the weights for individual stocks in an index. Schmidt and Fahlenbrach (2017) made an example of this with the CNH Global stock. In 2010 it would have the 412th position in the Russell 1000 index if only accounting for market capitalisation. However, float-adjusted it only held the 973th position in the index. This means that in regard to their own methodology most indices are biased against illiquidity.

The inclusion of a stock in the S&P 500 index, without any other news, has generated abnormal returns for the shareholders which have persisted for 10 to 20 trading days (Shleifer, 1986). This is related to index funds' investments as Shleifer (1986) presented about rebalancing situations. The abnormal returns are consistent with the price pressure hypothesis, where increased purchase demand tends to affect stock prices (Harris & Guel, 1986). These results demonstrate that trading by institutions and index funds impact the market prices. This was also supported by Harris and Guel (1986); Beneish and Whaley (1996) and Kaul et al. (2000), who found evidence close to identical to Shleifer (1986). However, Chen et al. (2004) argued that the excess return as an effect of an index inclusion is a result from increased investor awareness. Investor awareness increases for added stocks while there is a small drop of awareness for removed stocks. Increased awareness also leads to enhanced monitoring by investors and analysts which decrease the asymmetric information in the bid-ask spread (Chen et al., 2004).

The results above show that an index inclusion of a stock leads to temporary abnormal returns. There are different opinions regarding the cause of the excess return, but nevertheless capital inflow to included stocks play a major role. Due to the extensive research concerning the causes of stock price movements it is of interest to investigate whether the phenomenon of index inclusion is applicable to the recent trend with ETFs and if there are similar results in the underlying holdings of ETFs.

2.1.5 Exchange Traded Funds

ETFs can be considered a relatively young type of fund as they had their inception in the early 1990s. ETFs share many of the same characteristics as index mutual funds, such that they usually mimic an index at a low management fee, however, they differ in terms of structure. These differences become apparent when investigating the market microstructure of the funds. While mutual funds only trade at market end at the given market NAV, ETFs can trade throughout the day and may deviate from the NAV of the underlying securities. These deviations enable arbitrage opportunities for traders. In order to keep the ETF price in line with the NAV, APs are able to create and redeem ETF shares when its price deviates from that of the underlying securities. This procedure pushes the price of the ETF share in line with the underlying securities' NAV (ICI, 2017).

There are several factors that have contributed to the growth of ETFs. Some of these factors are related to money management features and others are ETF specific characteristics. However, one of the fundamental arguments are the high management fees that actively managed funds charge their investors. Indexed products, such as ETFs have a clear advantage in terms of cost structure compared to actively managed mutual funds (ICI, 2017).

The option of intra-day trading is also a favourable feature for investors. Institutional investors find the option an attractive way to access liquidity and a variety of asset classes, while the arbitrage opportunity prevents the ETF price to deviate greatly from the NAV in contrast to closed-end funds (Madura & Ngo, 2008a). Another favourable feature is the ability to offer investors a broad market exposure which has been proven to be an efficient approach to hedge against market and sector corrections or for speculative purposes. Furthermore, the index-tracking focus of ETFs decrease the portfolio turnover compared to actively managed funds, which makes them tax efficient. ETFs use the redemption process to distribute securities that was purchased at a lower price than the current market price and thus reduce their unrealised gains. This process does not incur any capital gains taxes to the investors until they sell their ETF shares (ICI, 2017).

Due to the perks of investing in ETFs it has become a popular form of investment and has grown significantly amongst institutional and individual investors. This trend is clear in terms of capital allocation where actively managed United States equity mutual funds have experienced capital outflows every year since 2005 in favour for passive United States equity ETFs. The increasing demand for ETFs has led to the creation of funds with different alternative benchmarks such as specific markets, asset classes, volatility and smart beta to name a few. The emergence of different types of ETFs has enabled small-scale investors to get exposure to markets which have been closed to them before. In 2008 the United States Securities and Exchange Commission approved the launch of a fully actively managed ETF (ICI, 2017).

2.1.6 ETF Spillovers

As with any new popular investment form it is interesting to see how ETFs' affect the financial market both directly and indirectly. Directly it impacts where investors allocate their capital, a consecution of this is that other investment forms might instead lose capital. Indirectly it is interesting to see what consequences it may have on the volatility, liquidity and valuation for the underlying securities. Krause, Ehsani and Lien (2014) found indications that suggests an ETF generate volatility in its' largest component stocks and Hegde and McDermott (2004) as well as Madura and Ngo (2008b) found that a stock's inclusion in an ETF can increase its liquidity. Madura and Ngo also found that the largest component stocks of an ETF have an elevation in their valuations. Even though there is limited research in the subject, these studies suggest that there are real spillover effects from an ETF to its underlying components.

2.1.7 ETF Sampling

The tracking error is often used as a measure of the performance of index mutual funds and ETFs. It measures how well a fund has been able to track its benchmark. It is reasonable to assume that ETF issuers would try to minimise this error as much as possible in order to attract capital. Less liquid securities in the benchmark have a negative impact on the tracking error due to increases in the expense ratio of the fund (Buetow & Henderson, 2012). Keim (1999) found that a complete replication strategy can induce more costs than a simple sample replication of an index due to trading in the benchmark's most illiquid stocks. Frino and Gallagher (2001) acknowledges that the best technique to use, whether it is full replication or partial replication depends on the liquidity of the underlying index. Hence, an inclusion of the most illiquid securities is costly for the fund and as a result many ETF issuers are sampling their benchmark instead of using full replication.

The combined factors of the surge to ETF and their ability to sample their holdings to improve their tracking error could indicate that there are consequences for traded stocks, more specifically for the most liquid stocks and the least liquid stocks on the market. The second research question of this study is to investigate if this is the case by researching if the magnitude of the liquidity premium has been altered as a consequence of the popularity in ETFs. Madura and Ngo (2008b) found that there is a positive valuation effect for component stocks in ETFs, but they did not discuss the possibility of sampling and hence not how this practice could have impacted the relative valuation between more and less liquid stocks on the broader market, the liquidity premium. They did find evidence that less liquid stocks which were included in ETFs tended to have larger positive valuation effects than more liquid stocks. This means that on the one hand, illiquid stocks are expected to be excluded to a large extent in ETFs. But on the other hand, the illiquid stocks which are included tends to have an enhanced valuation effect. The net effect of these two phenomena on the liquidity premium is ambiguous.

2.2 Financial Theories

2.2.1 Efficient Market Hypothesis

The Efficient Market Hypothesis (EMH) was developed by Fama (1965) and states that stock prices reflect all available information about a company. With all information available to investors, each investor value a stock based on the same information. Stocks are then said to be priced at their fair value, which eliminates the opportunity for investors to outperform the market without taking on a higher risk (Fama, 1970).

The EMH states that a stock market can have three different efficiency forms: weak, semi-strong or strong form. The weak form asserts that the current stock price reflects all past publicly available information. Based on this information, investors cannot outperform the market. More specifically, stock prices cannot be predicted since they follow a random walk pattern (Fama, 1965). The semi-strong form asserts that in addition to all past public information, stock prices also reflect all current publicly available information. The strong form of efficiency states that stock prices reflect all public and private information. This form implies that investors cannot earn excess return even by trading on insider information (Degutis & Novickyte, 2014).

2.2.2 Capital Asset Pricing Model

The Capital Asset Pricing Model (CAPM) is an early framework of the relationship between risk and return. The model was introduced in the 1960s by Sharpe (1964), Lintner (1965a & 1965b) and Mossin (1966). The CAPM is a further development of Markowitz's (1952) portfolio theory which states that firm-specific risk can be eliminated by diversification, however, systematic risk can only be reduced but not eliminated. The CAPM is based on four assumptions (Sharpe, 1964):

1. All investors are risk averse and evaluate their investment opportunities based on the expected rate of return and risk - expressed as standard deviation.
2. Capital markets are perfect; all assets are infinitely divisible, there are no transactions cost, short sales are restricted, all information is public and available to everyone and investors can borrow and lend at the risk-free rate.
3. Investors have the same investment opportunities.
4. Investors estimate the same values for expected return, standard deviation and correlation for individual assets.

Since all investors are risk averse, each investor hold the portfolio that maximises the risk-adjusted return, the Sharpe ratio. Investors hold risky assets in the same relative proportions, which result in that each investor hold the market portfolio. In a market equilibrium state, the market portfolio is also the portfolio that has the highest Sharpe ratio. By adding or deducting risky assets from the portfolio, investors are not able to increase the Sharpe ratio of the portfolio. The risk premium of each asset must satisfy:

$$E(R_i) - Rf = \beta_1(E(R_m) - Rf)$$

for i = 1,2,…,N (Asset)

Where $E(R_i)$ represent the expected return of asset i and $E(R_m)$ is the expected return of the market portfolio, β_1 represent the sensitivity of the asset's return with respect to the market portfolio. The Rf variable represents the risk-free rate of return, while the $(E(R_m) - Rf)$ function represents investors' demand for excess return for holding risky assets. The expected rate of return for an asset is derived by:

$$E(R_i) = Rf + \beta_1(E(R_m) - Rf)$$

for i = 1,2,…,N (Asset)

2.2.3 Fama-French Three-Factor Model

The Fama-French Three-Factor Model, created by Fama and French (1993) is an extension of the CAPM. In addition to the volatility of a stock's return to the market return, the model takes into account the size and book-to-market value of a company to predict the return of a stock. Fama and French found that the model had more explanatory power for explaining stock returns than the CAPM which only includes the market risk as a factor. The model in expected return form is:

$$E(R_i) = Rf + \beta_1(E(R_m) - Rf) + \beta_2(R_{SMB}) + \beta_3(R_{HML})$$

for i = 1,2,…,N (Asset)

The extensions from the CAPM model are the explanatory variables R_{SMB} and R_{HML}. R_{SMB} is included in order to capture the excess return small stocks have over big stocks and R_{HML} is included for the excess return of stocks with a high book-to-market value over stocks with a low book-to-market value. Fama and French divide stocks into a total of six baskets. The first step is to divide stocks in two baskets depending on stocks' market capitalisation. Which basket a stock is part of depends on its market capitalisation compared to the median of the market, a stock is hence part of either the basket with big (B) or small (S) stocks. The second step is to divide both baskets into three sub-baskets depending on the book-to-market value of the constituent stocks, the lowest 30% (L), the medium 40% (M) and the highest 30% (H). By performing the previous steps, six baskets are created: SL, SM, SH, BL, BM and BH. The formulas for calculating R_{SMB} and R_{HML} are:

$$R_{SMB} = \frac{R_{SL} + R_{SM} + R_{SH}}{3} - \frac{R_{BL} + R_{BM} + R_{BH}}{3}$$

$$R_{HML} = \frac{R_{SH} + R_{BH}}{2} - \frac{R_{SL} + R_{BL}}{2}$$

2.2.4 Fama-MacBeth Approach

The Fama-MacBeth regression approach was developed by Fama and MacBeth (1973) and was originally developed to test the efficiency of the CAPM model. The model has been extended to estimate different kinds of risk factors that can determine assets' prices and are frequently used as a tool to empirically test asset pricing models. Furthermore, the model works well in cross-sectional regressions with panel data (Fama & MacBeth, 1973).

The approach involves performing two steps. The first step is to regress each asset's individual return against a proposed risk factor in a time-series procedure to determine that asset's level of exposure to the risk factors, retrieving the asset's factor exposure. The coefficients (β's) in the equation below are the factor exposure of portfolio *p* to risk factor *j*:

$$\begin{aligned} &R_{1,t} = \alpha_1 + \beta_{1,F1}F_{1,t} + \ldots + \beta_{1,Fm}F_{m,t} + \varepsilon_{1,t} \\ &\vdots \\ &R_{P,t} = \alpha_P + \beta_{P,F1}F_{1,t} + \ldots + \beta_{P,Fm}F_{m,t} + \varepsilon_{P,t} \end{aligned}$$

for p = 1,2,...,P (Portfolio), t = 1,2,...,T (Time) and j = 1,2…,m (Risk factor)

$R_{p,t}$ is the return of portfolio *p* at time *t*. $F_{j,t}$ is the value of risk factor *j* at time *t* and $\beta_{p,Fj}$ is the factor exposure of portfolio *p* to risk factor *j*.

Obtaining the estimated factor exposures "$\hat{\beta}_{p,Fj}$" is the first step. The second step involves performing cross-sectional regressions on the returns of portfolios against the retrieved factor exposures for each time period. $R_{p,t}$ are the dependent variables and $\hat{\beta}_{p,Fj}$ are the explanatory variables. The purpose of the second step is to determine how each factor exposure affects returns, which result in a risk premium for each factor. The second equation is:

$$\begin{aligned} &R_{p,1} = \gamma_{0,1} + \gamma_{1,1}\hat{\beta}_{p,F1} + \ldots + \gamma_{m,1}\hat{\beta}_{p,Fm} + \varepsilon_{p,1} \\ &\vdots \\ &R_{p,T} = \gamma_{0,T} + \gamma_{1,T}\hat{\beta}_{p,F1} + \ldots + \gamma_{m,T}\hat{\beta}_{p,Fm} + \varepsilon_{p,T} \end{aligned}$$

for p = 1,2,...,P (Portfolio), t = 1,2,...,T (Time) and j = 1,2,...,m (Risk factor)

$R_{p,t}$ is the return of portfolio p at time t. $\hat{\beta}_{p,Fj}$ is the estimated factor exposure of portfolio p to risk factor j and $\gamma_{j,t}$ is the regression coefficient of factor exposure j at time t. The estimated market risk premium for a risk factor is γ_j and is calculated by averaging an estimated risk factor over the number of periods.

3 Method

In this chapter, the philosophies underlying the method are presented, the econometric equations used are introduced as well as a description of how data was acquired. Moreover, the portfolio creation process and the regression methods are presented.

3.1 Methodology

In order to investigate the research questions in an objective manner, this thesis takes on a positivistic approach. In a study based of positivism the researchers are independent and results are based upon facts. This approach is the most suitable for the thesis since it allows to test the hypothesis on the basis of a theoretical framework through the medium of statistical analysis. The potential downside of a positivistic approach is that the result may support a hypothesis while disregarding the forces and processes behind the result (Esterby-Smith, Thorpe & Jackson, 2015).

The purpose of the study is to investigate how the liquidity premium has developed between 1997 and 2016 and if the capital inflow to the ETF market has impacted the premium. The first question is developed from earlier research which implies that the reasoning behind the first question takes on a deductive approach (Saunders, Lewis & Thornhill, 2009). The second question is not anchored in any earlier work, however, several studies have investigated how a security's inclusion in an index affects its price (Shleifer, 1986; Harris & Guel, 1986; Beneish & Whaley, 1996; Kaul et al., 2000; Chen et al., 2004; Baran, & King, 2014). These articles created the basis for the second question which implies that it has a deductive approach as well (Saunders et al., 2009).

Throughout the study, estimates are based on numeric data rather than qualitative features which makes it a quantitative study. When testing the liquidity premium, the approach was a longitudinal panel study. A longitudinal panel is a research design which has multiple dimensions, such as time and subjects (Menard, 2002). Examples of subjects are different firms, individuals or countries. When testing how the growth of ETFs has affected the liquidity premium, a time-series regression was made since there was only a time dimension.

Longitudinal panel studies consists of a combination of cross-sectional and time series data which means that heteroscedasticity related to cross-sectional regressions and autocorrelation related to time series regressions needs to be addressed (Gujarati & Porter, 2009). The major distinction between cross-sectional and longitudinal studies is that longitudinal studies consists of data collected over at least two periods. Since longitudinal studies analyse data over time it is common to have data sets with missing data. As a result, several issues arise related to how one deal with the missing data. There is always a risk that interference by a researcher on the data lead to biased estimates or inaccurate estimates in the descriptive statistics (Menard, 2002). The benefit of a longitudinal study is that a researcher can observe changes and developments between units over time.

3.2 Models

To recap, two research questions were stated at the beginning of the study: 1) How the liquidity premium has developed between 1997 and 2016 and 2) How possible alterations in the liquidity premium can be linked to the steep capital inflow to the ETF market. We hypothesize that the characteristic risk premium has decreased between 1997 and 2016 but have no hypothesis on how the systematic risk premium has developed. Moreover, we hypothesize that the capital inflow to the ETF market has been a factor which has influenced the decrease in the characteristic risk premium. In Chapter 3.2.1 and Chapter 3.2.2 below, two different models are presented. In Chapter 3.2.1, the model to test the development of the liquidity premium is presented and is hence the model which is used to test the first research question. The model that is used to test if ETFs have affected the liquidity premium and hence the second research question is presented in Chapter 3.2.2.

3.2.1 Liquidity Premium Model

In order to make the foundation of the model to test the liquidity premium, the three risk factors used in the Fama-French Three-Factor Model were included. The risk factors that have been found to impact the returns of stocks are: the market risk, the market capitalisation of a company and the book-to-market value of a company. The reason for including these variables is to risk-adjust the returns of securities. The factors are commonly used by researchers when investigating the liquidity premium such as by Bradrania and Peat (2014), Liu (2006) and Pástor and Stambaugh (2003). In this study, portfolios are used instead of stocks since there can be large fluctuations in the risk factors between individual stocks. Creating portfolios is done in order to avoid errors in the estimations (Amihud & Mendelson, 1986b). The variables for the market capitalisations and the book-to-market values are added as portfolio specific variables instead of the market-wide excess return variables which are used in the Fama-French Three-Factor Model. This means that small-minus-big (SMB_t) is replaced by "MV_{pt}" and high-minus-low (HML_t) is replaced by "$BTMV_{pt}$". MV_{pt} and $BTMV_{pt}$ are specific to each portfolio and can hence be used in cross-sectional regressions. These changes were necessary since global factors are omitted in Fama-MacBeth cross-sectional regressions which are performed to acquire the liquidity premium in this thesis. Consider the modified Fama-French Three-Factor Model in a panel data setting with portfolios as assets in equation 1 below.

$$RPRF_{pt} = \alpha_{pt} + \beta_1(RMRF_t) + \beta_2(MV_{pt}) + \beta_3(BTMV_{pt}) + \varepsilon_{pt} \quad (1)$$

for t = 1,2,...,T (Month) and p = 1,2,...,P (Portfolio)

Where:

$RPRF_{pt}$ was the excess return of portfolio *p* in month *t*.

α_{pt} was the alpha of portfolio *p* in month *t*.

$RMRF_t$ was the excess return of the market in month *t*.

MV_{pt} was the market capitalisation of portfolio *p* in month *t*.

$BTMV_{pt}$ was the book-to-market of portfolio *p* in month *t*.

ε_{pt} was the error term of portfolio *p* in month *t*.

To be able to test for the liquidity premium two additional explanatory variables were included, one proxy for the characteristic liquidity risk and one proxy for the systematic liquidity risk. The proxy used for the characteristic liquidity risk was the bid-ask spread introduced by Amihud and Mendelson (1986a), and the proxy used for the systematic liquidity risk was the ILLIQ measure introduced by Amihud (2002).

To construct the proxy for the characteristic liquidity risk the daily bid-ask spreads of each stock were used to calculate the average daily bid-ask spread of a portfolio *p* in month *t*. The rationale for using this measure is because it indicates what the immediate cost of execution is. The proxy was named SPR_{pt} since it is a measure of the spread. The SPR variable for individual stocks and for portfolios in month *t* was calculated as:

$$SPR_{it} = \left(\frac{1}{D_t}\right) * \sum_{d=1}^{D} \left(\frac{PH_{id} - PL_{id}}{PH_{id}}\right) \qquad (2)$$

$$SPR_{pt} = \left(\frac{1}{I_{pt}}\right) * \sum_{i=1}^{I} SPR_{it} \qquad (3)$$

for i = 1,2,...,I (Stock), t = 1,2,...,T (Month), d = 1,2,...,D (Day) and p = 1,2,...,P (Portfolio)

Where:

SPR_{it} was the average daily bid-ask spread of stock *i* in month *t*.

PH_{id} was the intraday high price of stock *i* on day *d*.

PL_{id} was the intraday low price of stock *i* on day *d*.

D_t was the number of trading days in month *t*.

SPR_{pt} was the average daily bid-ask spread of the constituent stocks in portfolio *p* in month *t*.

I_{pt} was the number of constituent stocks in portfolio *p* in month *t*.

The systematic liquidity risk proxy was constructed by using the ILLIQ measure created by Amihud (2002). It measures the absolute price change to the dollar trading volume. The higher the variable, the more the stock price moves given the turnover by value of the stock. Stocks with a high ILLIQ ratio are more prone to have large price fluctuations given the same turnover as stocks with a small ILLIQ ratio. In dire stock market times, stocks with a high ILLIQ ratio should experience more severe fluctuations in their prices. The variable was named DPT_{pt} in order to highlight that it is a test of the depth. The average DPT of a stock and of a portfolio in month *t* was calculated as:

$$DPT_{it} = \left(\frac{1}{D_t}\right) * \sum_{d=1}^{D} \left(\frac{|R_{id}|}{VOL_{id}}\right) \tag{4}$$

$$DPT_{pt} = \left(\frac{1}{I_{pt}}\right) * \sum_{i=1}^{I} DPT_{it} \tag{5}$$

for i = 1,2,...,I (Stock), t = 1,2,...,T (Month), d = 1,2,...,D (Day) and p = 1,2,...,P (Portfolio)

Where:

DPT_{it} was the average daily ratio of absolute return to dollar trading volume of stock *i* in month *t*.

$|R_{id}|$ was the absolute return of stock *i* on day *d*.

VOL_{id} was the dollar trading volume of stock *i* on day *d*.

D_t was the number of trading days in month *t*.

DPT_{pt} was the average daily ratio of absolute return to dollar trading volume of the constituent stocks in portfolio *p* in month *t*.

I_{pt} was the number of constituent stocks in portfolio *p* in month *t*.

By adding the liquidity explanatory variables to equation (1), the new equation becomes:

$$RPRF_{pt} = \alpha_{pt} + \beta_1(RMRF_t) + \beta_2(MV_{pt}) + \beta_3(BTMV_{pt}) + \beta_4(SPR_{pt}) + \beta_5(DPT_{pt}) + \varepsilon_{pt} \tag{6}$$

for t = 1,2,...,T (Month) and p = 1,2,...,P (Portfolio)

Where:

$RPRF_{pt}$ was the excess return of portfolio *p* in month *t*.

α_{pt} was the alpha of portfolio *p* in month *t*.

$RMRF_t$ was the excess return of the market in month *t*.

MV_{pt} was the market capitalisation of portfolio p in month *t*.

$BTMV_{pt}$ was the book-to-market of portfolio p in month *t*.

SPR_{pt} was the average daily bid-ask spread of the constituent stocks in portfolio p in month *t*.

DPT_{pt} was the average daily ratio of absolute return to dollar trading volume of the constituent stocks in portfolio p in month *t*.

ε_{pt} was the error term of portfolio p in month *t*.

The coefficient for the SPR_{pt} variable ($\hat{\beta}_4$) represents the estimated excess monthly return of a stock for each percentage spread in the bid-ask prices. Similarly, the coefficient for the DPT_{pt} variable ($\hat{\beta}_5$) represents the estimated excess return of a stock given the ratio between the absolute return and trading volume. The estimated coefficients $\hat{\beta}_4$ and $\hat{\beta}_5$ are hence the approximated characteristic and systematic liquidity premiums.

3.2.2 ETF Model

To be able to test the effect of the capital inflow to the ETF market against the liquidity premium the ratio of the total ETF value in equities to the total value of the equity market in the United States was calculated for each year. This ratio made it possible to see how the liquidity premium was affected when the ETF share of the total equity market changed. The formula for the ETF variable "ETF_y" was calculated as:

$$ETF_y = \frac{Total\ ETF\ value\ of\ equity_y}{Total\ market\ value\ of\ equity_y} \tag{7}$$

for y = 1,2,...,Y (Year)

Where:

ETF_y was the ratio of United States equity which was allocated in ETFs in year y.

$Total\ ETF\ value\ of\ equity_y$ was the total value ETF's had allocated in United States securities in year y.

$Total\ market\ value\ of\ equity_y$ was the total market capitalisation of the United States equity market in year y.

After calculating the ETF variable for the twenty years between 1997 and 2016, it was used as an explanatory variable when regressing against the estimated liquidity premium coefficients $\hat{\beta}_{4y}$ and $\hat{\beta}_{5y}$. The ETF variable was not tested against the liquidity variables SPR and DPT directly since it is only hypothesized in this thesis that ETFs affects the magnitude of the liquidity premium and not the spreads or the depths of stocks.

$$\hat{\beta}_{4y} = \alpha_{SPR} + \beta_6(ETF_y) + \varepsilon_y \quad (8)$$

$$\hat{\beta}_{5y} = \alpha_{DPT} + \beta_7(ETF_y) + \varepsilon_y \quad (9)$$

for y = 1,2,...,Y (Year)

Where:

$\hat{\beta}_{4y}$ was the previously estimated coefficient for the SPR_{pt} variable in year y.

$\hat{\beta}_{5y}$ was the previously estimated coefficient for the DPT_{pt} variable in year y.

α_{SPR} is the estimate of the mean value of the characteristic liquidity premium if ETF_y = 0.

α_{DPT} is the estimate of the mean value of the systematic liquidity premium if ETF_y = 0.

ETF_y was the ratio of United States equity which was allocated in ETF funds in year y.

ε_y was the error term in year y.

3.3 Data Collection

Equations (6), (8) and (9) were used in regressions while equations (2), (3), (4), (5) and (7) were used in order to acquire the variables needed for those equations. The data needed in the equations was collected for the years between 1997 and 2016. There were two reasons for choosing this time-period. Firstly, there is limited recent research about the liquidity premium, providing an opportunity to contribute to the research field. Secondly, these years made it possible to incorporate capital inflow to the ETF market due to their inception in the early 1990s.

The tests in this study are performed on stocks included in the market index NASDAQ Composite. The stocks in the index have varying size and characteristics which enables a diversified sample. The reason for conducting the study on a market index with corporations from mainly the United States was because it is the world’s largest ETF market and because reliable data from the country is readily accessible.

All stock data was downloaded from Thomson Reuters Datastream. Because the data was originally collected by someone else and compiled into a database it was secondary data. The decision to use Datastream was based on its broad available time series data. Data was obtainable for each component stock of the NASDAQ Composite index for the relevant time period from the database. Furthermore, the data output was presented in a pliable way which made it possible to work with the data efficiently. The data retrieved from Datastream was the market values (MV), market-to-book values (MTBV), prices of stocks adjusted for capital actions (P), daily high prices of stocks (PH), daily low prices of stocks (PL), turnover by value (VA), turnover by volume (VO) and volume weighted average price (WVAP).

The one-month risk-free rates and the excess market returns were obtained from Kenneth French's website (French, 2018) and was hence secondary data. The risk-free rates obtained were the one-month United States Treasury bill rate.

The acquired data mentioned above was used to calculate the variables $RPRF_{pt}$, $RMRF_{t}$, MV_{pt}, $BTMV_{pt}$, SPR_{pt} and DPT_{pt} used in equation (6). Further details on how the variables were calculated is stated in Appendix A.

There were three steps involved when removing stocks from the sample:

1. In the data output retrieved from Datastream, stocks which were listed in a later period or that had already been delisted were included. Because of the inclusion of stocks which were not listed in the corresponding time period, those entries had to manually be removed from the output.
2. Stocks with incomplete data observations were manually removed. This was done in order to assure a balanced panel.
3. Stocks which were considered severe outliers were removed due to the risk of having possible negative interference with results reflecting the nature of the stock market. Furthermore, the presence of severe outliers increases the risk of heteroscedasticity (Gujarati & Porter, 2009). Stocks with an observation above a 25% SPR value and stocks with an observation above a 500% DPT value were removed. To put these values in a relation, the 95% percentile of the SPR variable was 7.9% and the 95% percentile of the DPT variable was 10.2%. The five stocks that were removed in this process can be found in appendix B.

In Table 2 below, the descriptive statistics for the 210 stocks used as the sample is presented. All the included stocks can be found in appendix C.

Table 2: Descriptive Statistics Stocks

Variable	#Obs.	Mean	Std. Dev.	Min.	Max.	Kurtosis	Skewness
RIRF	50400	0.3%	14.0%	-146.2%	164.9%	7.51	-0.34
MV	50400	9653	36566	3	752016	110.68	9.10
BTMV	50400	0.48	0.44	-2.56	20.00	224.53	8.96
SPR	50400	3.8%	2.1%	0.7%	24.6%	5.47	1.82
DPT	50400	2.4%	11.6%	0.0%	486.5%	352.96	15.11

Note that RIRF has a minimum observed return in excess of -100%, this was possible because the calculations of the returns were done using logarithmic returns. Also note the high kurtosis in some of the variables. The high kurtosis is present due to deviant observations still being present in the data even after removing the most severe outliers.

The data for the capital allocated to ETFs was retrieved from ICI (2017). ICI had data for all the years in the relevant time period 1997-2016. The data for the total United States market equity value was retrieved from The World Bank (n.d.) and it covered the entire time period. Both of these datasets were secondary data in this study.

3.4 Portfolio Construction

Ten basis portfolios were constructed from the stock sample using a composite measure of both the characteristic and systematic liquidity variables. For each year, stocks were ranked based on their SPR multiplied by their DPT in January and the portfolios were then formed based on this measure. When testing the liquidity premium, portfolio formation is often done by using stocks' liquidity ratios, for example by Amihud and Mendelson (1986a) and Eleswarapu (1997). The stocks with the highest value constituted portfolio 1 and the stocks with the lowest value constituted portfolio 10 for each year. The portfolios were constructed equally-weighted with 21 companies in each and were rebalanced annually. The creation of portfolios was done in order to decrease the high idiosyncratic volatility of individual stocks.

Table 3: Descriptive Statistics Portfolios

Variable	#Obs.	Mean	Std. Dev.	Min.	Max.	Kurtosis	Skewness
RPRF	2400	0.3%	7.2%	-34.2%	30.2%	1.73	-0.45
MV	2400	9206	20605	73	124088	11.49	3.39
BTMV	2400	0.48	0.19	0.17	2.67	18.55	2.87
SPR	2400	3.8%	1.5%	1.2%	12.2%	1.49	1.09
DPT	2400	2.4%	7.0%	0.0%	85.3%	38.39	5.58

The kurtosis and skewness of the data has decreased at the expense of less observations. Even though more observations generally provide better results, the loss of observations does not critically impact the method since there are still 2400 observations for each variable.

3.5 Regression Method

The Fama-MacBeth regression approach is a common method to estimate the risk premium of different risk factors. It is also common to employ this method when investigating the liquidity premium. Authors which use the Fama-MacBeth approach when researching the premium includes Amihud and Mendelson (1986), Eleswarapu (1997), and Ben-Rephael et al. (2015) to name a few. Another option to test the relationship between variables and returns is to do a time-series regression with Fama-French return factors. When testing the liquidity premium in this thesis, the Fama-MacBeth approach was applied on equation (6) to test the liquidity risk factors SPR and DPT because it is a common approach when investigating risk factors.

To be able to draw conclusions of trends in the liquidity premium as well as for the existence of a premium for the whole period and sub-periods, 1-year regressions, 5-year regressions and a full 20-year regression was conducted. The 1-year regressions consisted of twelve periods, one for each month. With ten portfolios and twelve periods the 1-year regressions had 120 observations. These regressions were made in order to investigate the trend of the liquidity premium as well as a way to obtain the coefficients used to test how ETFs has affected the premium. It made it possible to investigate how the liquidity coefficients changed over time. Each of the 5-year regressions consisted of 600 observation. In total four 5-year regressions were made: 1997-2001, 2002-2006, 2007-2011 and 2012-2016. The full 20-year test period regression had 2400 observations and was made in order to see if any conclusions could be drawn by looking at the entire period.

When the 1-year coefficients for the liquidity variables were obtained from the Fama-MacBeth regression, the impact of ETFs could be tested. The coefficient results were used as dependent variables and the ETF variable as the explanatory variables. The tests were done using time-series regression because the data only had one dimension - time. The time-series regressions were made on equations (8) and (9).

3.6 Quality of the Method

Results from empirical studies are only useful if they are obtained by using a correct and unbiased process. When conducting research there is always a possibility of errors in the data or in the calculations. The data from Datastream and the other sources have been trusted to be correct due to their authoritativeness. The risk of being exposed to inaccuracies in the data has been mitigated by being critical and by verifying with additional sources. There is also the element of human error present when handling and compiling data after it has been retrieved. By sticking to declared standards and being attentive in the procedure of creating the finalised spreadsheets the risk of human error has been minimized. Additionally, it is assumed that the models are correctly specified and that they include all the relevant variables. If any relevant variables would be omitted, the models would suffer from specification bias which could lead to autocorrelation and heteroscedasticity (Gujarati & Porter, 2009).

When working with data, biases may occur. Two common types of biases are "sample selection bias" and "survivorship bias". Sample selection bias means that the sample which is tested does not represent the entire population (Heckman, 1979). A problem that can arise with this bias is how well the results corresponds to the entire population and the subjects which were excluded. The sample selection bias can hence impair the external validity of a test. Survivorship bias means excluding units which have not made it past some kind of barrier, this means that "failures" are excluded (Linnainmaa, 2013). This is a common bias in financial research since companies which have been delisted or has gone into bankruptcy are often excluded. Imagine that a researcher wants to test the average stock market return over a long time period, by excluding delisted stocks the researcher would typically receive overly optimistic results. In this thesis, both sample selection bias and survivorship bias may be present. Because of the removal of stocks with incomplete data, a certain type of stocks are excluded, the ones which lack data, typically stocks with extremely low liquidity. However, because the sample still includes stocks with a wide range of liquidity values, results can still be received on the relationship between returns and liquidity. Survivorship bias may also be present because only stocks which have data in all time periods are included. Because the cross-sectional regressions are performed on monthly returns, the effects of a possible survivorship bias are not as significant.

4 Empirical Results

In this chapter, the results from the diagnostics regressions for heteroscedasticity and autocorrelation as well as the results from the statistical measurements of the models (6), (8) and (9) are presented.

All of the statistical measures performed in this chapter are calculated in the software Stata. The output is assumed to have been correctly calculated inside the program. This also include the Boston College Archive package program "xtfmb" which is used to execute the Fama-MacBeth regression (Hoechle, n.d.). For each model, results are only considered to be statistically significant if they have a p-value below 0.05.

4.1 Robustness Tests

Because panel data is a combination of cross-sectional and time-series data, both autocorrelation and heteroscedasticity can exist in the data. If heteroscedasticity or autocorrelation is present it can impair the tests of significance. To ensure regressions without these drawbacks, diagnostics regressions were made to enquire whether they were present. To test for heteroscedasticity in the data, Breusch-Pagan tests (Breusch & Pagan, 1979) were carried out. The results of the tests are presented in Table 4 below.

Table 4: Breusch-Pagan Tests

H_0: Constant Variance	
Variables: Fitted Values of $RPRF_{pt}$	
Chi-square(1)	9.41
Prob > Chi-square	0.00
H_0: Constant Variance	
Variables: $RMRF_t$ MV_{pt} $BTMV_{pt}$ SPR_{pt} DPT_{pt}	
Chi-square(5)	783.98
Prob > Chi-square	0.00

The null hypothesis of homogeneity in the variance was rejected, both in the fitted values of the dependent variables as well as for the independent variables. It can be concluded that the data suffers from heteroscedasticity. Without proper actions, the standard errors would be biased when conducting the regressions. A consequence of biased errors is the risk of making wrong conclusions due to inaccuracies in the levels of significance.

To test for autocorrelation in the data, a Wooldridge test (Wooldridge, 2002) was performed. The result from the test is presented in Table 5.

Table 5: Wooldridge Test

H_0: No First-order Autocorrelation	
F(1,9)	6.26
Prob > F	0.03

The null hypothesis of no autocorrelation was rejected, meaning that autocorrelation is present. If unaccounted for, autocorrelation would bias the standard errors, which would impact the results.

To address the problems in the data, lags were included in the Fama-MacBeth regressions. The lag length was set to T-2, where T is the total number of months in the regression. T-2 is the maximum number of lags possible when executing the xtfmb package in Stata. By including the lags, Newey-West standard error estimates were obtained (Hoechle, n.d.). The Newey-West procedure is done in order to create a covariance matrix which accounts for heteroscedasticity and autocorrelation in the data which can make the estimates more accurate (Newey & West, 1987).

4.2 Liquidity Premium Results

To recap, the formula on which the liquidity premium regressions are based upon is:

$$RPRF_{pt} = \alpha_{pt} + \beta_1(RMRF_t) + \beta_2(MV_{pt}) + \beta_3(BTMV_{pt}) + \beta_4(SPR_{pt}) + \beta_5(DPT_{pt}) + \varepsilon_{pt} \quad (6)$$

Since the focus of this study is the liquidity premium, the interesting empirical results are the sign, the magnitude and the significance of $\beta_4(SPR_{pt})$ and $\beta_5(DPT_{pt})$ which are proxies for the characteristic risk premium and the systematic risk premium. The full regression outputs are presented in appendices D-F. Only the liquidity premium coefficients are presented when we report the empirical results below in order to avoid deviating from the purpose of the thesis.

4.2.1 20-Year Results

The results for the entire 20-year period cross-sectional Fama-MacBeth regression is presented in Table 6. The table includes the coefficients of the SPR and DPT variables, along with the corresponding Newey-West standard errors and the resulting t-statistic.

Table 6: 20-Year Regression

Years	Variable	Coefficient	Newey-West Std. Err.	t-Statistic
1997-2016	SPR_{pt}	-0.49%**	0.20%	-2.48
1997-2016	DPT_{pt}	0.15%***	0.05%	3.03

**** Significant at 1%*

*** Significant at 5%*

** Significant at 10%*

The regression result for the SPR variable show that there is a significant negative relationship between the spread and the returns of stocks. Given these results, a stock with a 1% higher spread than another stock would yield on average 0.49% lower monthly return. The DPT on the other hand, has a significant positive relationship with return, indicating that there is a positive systematic liquidity premium present. The result implies that a stock with a 1% higher DPT ratio than another stock return on average 0.15% better on a monthly basis.

4.2.2 5-Year Results

The output for the cross-sectional Fama-MacBeth regression for the four sub-periods are presented below in Table 7. The two tables present the results for the periods 1997-2001, 2002-2006, 2007-2011 and 2012-2016 for SPR_{pt} and DPT_{pt}.

Table 7: 5-Year Regressions

Years	Variable	Coefficient	Newey-West Std. Err.	t-Statistic
1997-2001	SPR_{pt}	-0.07%	0.23%	-0.30
2002-2006	SPR_{pt}	0.33%	0.39%	0.09
2007-2011	SPR_{pt}	-0.55%	0.54%	-1.01
2011-2016	SPR_{pt}	-1.44%***	0.16%	-9.14
1997-2001	DPT_{pt}	0.05%**	0.02%	2.11
2002-2006	DPT_{pt}	-0.19%	0.21%	-0.91
2007-2011	DPT_{pt}	0.48%	0.39%	1.23
2011-2016	DPT_{pt}	0.27%***	0.07%	3.88

**** Significant at 1%*

*** Significant at 5%*

** Significant at 10%*

The beta coefficient for SPR_{pt} is only positive for the period 2002-2006, while the beta coefficient for DPT_{pt} is only negative for same period. SPR_{pt} is significant between 2011 and 2016 on the 1% level. DPT_{pt} on the other hand, is significant at the 5% level between 1997 and 2001 and significant at the 1% level between 2011 and 2016. Although not all results for the sub-periods are significant at the 5% level, the results show that there has been a negative characteristic liquidity premium and a positive systematic liquidity premium in the majority of the sub-periods.

4.2.3 1-Year Results

The results for the cross-sectional Fama-MacBeth regressions for each year are presented in Table 8 and Table 9. The sign, magnitude and level of significance of the coefficients vary between the years. Ignoring the significance level, a majority of the coefficients for SPR_{pt} was negative and a majority of the coefficients for DPT_{pt} was positive.

Table 8: 1-Year Regressions SPRpt

Year	*Variable*	*Coefficient*	*Newey-West Std. Err.*	*t-Statistic*
1997	SPR_{pt}	-0.02%	0.76%	-0.03
1998	SPR_{pt}	-0.74%	1.15%	-0.64
1999	SPR_{pt}	1.71%**	0.56%	3.05
2000	SPR_{pt}	-1.21%	0.99%	-1.22
2001	SPR_{pt}	-0.07%	0.49%	-0.15
2002	SPR_{pt}	-1.57%**	0.61%	-2.56
2003	SPR_{pt}	0.67%	0.38%	1.77
2004	SPR_{pt}	-0.83%	0.50%	-1.66
2005	SPR_{pt}	1.37%*	0.63%	2.18
2006	SPR_{pt}	0.64%	0.41%	1.55
2007	SPR_{pt}	-3.57%***	0.39%	-9.22
2008	SPR_{pt}	-0.64%	0.55%	-1.17
2009	SPR_{pt}	1.12%**	0.38%	2.96
2010	SPR_{pt}	0.89%	0.60%	1.49
2011	SPR_{pt}	-0.54%	0.66%	-0.81
2012	SPR_{pt}	-2.02%**	0.74%	-2.73
2013	SPR_{pt}	-0.86%	0.62%	-1.38
2014	SPR_{pt}	-0.55%	0.61%	-0.90
2015	SPR_{pt}	-2.63%***	0.74%	-3.53
2016	SPR_{pt}	-1.12%***	0.26%	-4.30

**** Significant at 1%*

*** Significant at 5%*

** Significant at 10%*

Table 9: 1-Year Regressions DPTpt

Year	Variable	Coefficient	Newey-West Std. Err.	t-Statistic
1997	DPT_{pt}	-0.03%	0.06%	-0.52
1998	DPT_{pt}	0.08%	0.11%	0.69
1999	DPT_{pt}	0.04%	0.10%	0.44
2000	DPT_{pt}	-0.03%	0.02%	-1.47
2001	DPT_{pt}	0.17%***	0.06%	3.11
2002	DPT_{pt}	0.25%*	0.13%	2.00
2003	DPT_{pt}	0.09%	0.07%	1.21
2004	DPT_{pt}	0.47%	0.30%	1.56
2005	DPT_{pt}	-1.37%***	0.27%	-4.97
2006	DPT_{pt}	-0.39%	0.42%	-0.91
2007	DPT_{pt}	2.87%***	0.37%	7.86
2008	DPT_{pt}	-0.16%	0.30%	-0.53
2009	DPT_{pt}	-0.04%	0.21%	-0.18
2010	DPT_{pt}	-0.39%**	0.13%	-2.83
2011	DPT_{pt}	0.11%	0.15%	0.73
2012	DPT_{pt}	0.37%*	0.20%	1.87
2013	DPT_{pt}	0.56%**	0.23%	2.44
2014	DPT_{pt}	-0.08%	0.18%	-0.43
2015	DPT_{pt}	0.54%*	0.26%	2.10
2016	DPT_{pt}	-0.04%	0.16%	-0.28

**** Significant at 1%*

*** Significant at 5%*

** Significant at 10%*

4.3 ETF Results

The formulas which the ETF regressions are based upon are:

$$\hat{\beta}_{4y} = \alpha_{SPR} + \beta_6(ETF_y) + \varepsilon_y \tag{8}$$

And

$$\hat{\beta}_{5y} = \alpha_{DPT} + \beta_7(ETF_y) + \varepsilon_y \tag{9}$$

In order to test the impact of the capital inflow to the ETF market on the liquidity premium, the two time-series regressions were performed with the estimated liquidity premiums as dependent variables and ETFs' market share of the total United States equity market as the explanatory variable. In Table 10 below, the data used in the regressions is presented. To clarify that the value is a correlation coefficient, the notation beta (β) is used instead of gamma (γ) even though a gamma notation is used in the original Fama-MacBeth model. The estimated beta coefficients ($\hat{\beta}$'s) for SPR and DPT were retrieved from the previous 1-year regressions. Not all of the estimated betas were significant on the 5% level so no final statistical conclusions can be made from the regressions. However, the output of the two time-series regressions can be used to spot trends and gain insight in how the characteristic and systematic liquidity premium have been affected by the changes in ETFs' relative market share.

Table 10: Inputs Used in the Time-Series Regressions to Estimate ETFs Impact on the Liquidity Premium

	Variables		
Year	$\hat{\beta}_{4y}$ (SPR_{pt})	$\hat{\beta}_{5y}$ (DPT_{pt})	ETF_y
1997	-0,02%	-0,03%	0,03%
1998	-0,74%	0,08%	0,07%
1999	1.71%**	0,04%	0,12%
2000	-1,21%	-0,03%	0,17%
2001	-0,07%	0.17%***	0,21%
2002	-1.57%**	0.25%*	0,38%
2003	0,67%	0,09%	0,54%
2004	-0,83%	0,47%	0,82%
2005	1.37%*	-1.37%***	0,96%
2006	0,64%	-0,39%	0,84%
2007	-3.57%***	2.87%***	0,90%
2008	-0,64%	-0,16%	0,85%
2009	1.12%**	-0,04%	0,69%
2010	0,89%	-0.39%**	1,13%
2011	-0,54%	0,11%	1,32%
2012	-2.02%**	0.37%*	1,24%
2013	-0,86%	0.56%**	1,99%
2014	-0,55%	-0,08%	2,90%
2015	-2.63%***	0.54%*	3,09%
2016	-1.12%***	-0,04%	2,98%

***** Significant at 1%*
*** Significant at 5%*
** Significant at 10%*

The results from the two regressions are presented below in Table 11.

Table 11: Regression Output based on the input from Table 10

Years	Dependent Variable	Explanatory Variable	Coefficient	Std. Err.	t-Statistic	P-value
1997-2016	$\hat{\beta}_{4y}$ (SPR_{pt})	ETF_y	-0,39%	0,32%	-1,21	0,24
1997-2016	$\hat{\beta}_{5y}$ (DPT_{pt})	ETF_y	0,04%	0,19%	0,23	0,82

**** Significant at 1%*

*** Significant at 5%*

** Significant at 10%*

The first result represents the relation between the characteristic liquidity premium and the change in the ETF market share while the second result represent the relation between the systematic liquidity premium and the change in the ETF market share. None of the obtained results are significant on any significance level. However, the impact on the characteristic liquidity premium appears to be negatively impacted by the ETF growth, the premium decrease by 0.39% for each 1% increase in ETFs share of the market. Adversely, the systematic liquidity premium appears to be slightly positively impacted by the ETF growth with a premium increase of 0.04% for each 1% increase in ETFs share of the market.

A possible drawback in the method is the risk of excluding variables that could have had an impact on the level of the premium, such as other indexing products, average holding period of investors or a proxy which reflects the overall total market efficiency. A Ramsey RESET test can be performed in order to assess if a variable has been omitted in a regression and is hence a test for specification errors (Ramsey, 1969). The test was done on the ETF regressions performed on equation (8) and equation (9). The results are presented in Table 12 below and show no significance that these equations lack a variable that could help explain the changes in the liquidity premium. Even though the tests show no statistical significance that the equations lack explanatory variables, it is possible that there are variables that would improve them.

Table 12: Ramsey RESET Test

Model	$\hat{\beta}_{4y} = \alpha_{SPR} + \beta_6(ETF_y) + \varepsilon_y$
H_0: Model Has No Omitted Variables	
F(3,15)	0.8
Prob > F	0.51
Model	$\hat{\beta}_{5y} = \alpha_{DPT} + \beta_7(ETF_y) + \varepsilon_y$
H_0: Model Has No Omitted Variables	
F(3,15)	0.02
Prob > F	1.00

5 Analysis

In this chapter, deeper interpretations are made from the empirical results obtained. The development of the liquidity premium for the period is investigated as well as the relationship between ETFs and the liquidity premium.

5.1 Quality of the Results

Retrieving the results is just one step when investigating how a study can be applied in practice, the next step is as important, to determine if the results make sense in the real world. One thing that becomes clear in the result is that the magnitude by which return is affected by the systematic liquidity premium is positively impacted by the year 2007, which has a high value compared to the other years. The variable for 2007 was 2.87%, the mean for all the years was 0.15% including 2007 and the mean excluding 2007 was 0.01%. By omitting the 2007 estimate, the systematic risk premium would essentially be zero. However, an aspect that strengthens the idea that systematic risk premium does impact returns positively is that it is significant for the first and last five years, outside the scope of 2007, the first five years on the 5% level and last five years on the 1% level. Likewise, the maximum absolute value for the characteristic liquidity premium was for the year 2007 with -3.57%, however this estimate is not as severe since the coefficient for the spread in general is more volatile. The mean value is -0.49% including 2007 and -0.34% excluding 2007. This means that even though $|3.57|>|2.87|$, -3.57% is still not considered a deviation to the same extent as 2.87. The year 2007 stands out because the financial crisis broke out which affected the risk willingness of market participants, this is discussed in more detail in sub-chapter 5.2 where the development of the premium through the years based on the results is construed.

5.2 Liquidity Premium Development

The purpose of the 1-year regressions was to identify if there is a trend in the liquidity premium. The estimates of the correlation coefficients for both the liquidity proxies have large fluctuations between different years and it is clear that there is no certainty which sign or what magnitude they will have in the short term. There is a positive relationship between the spread and the returns of stocks for 6 years while the remaining 14 years has a negative relationship. In relation, there is a positive systematic liquidity premium for 11 years and a negative relationship for the remaining 9 years. Consider the two graphs below, Figure 1 and Figure 2.

Figure 1: Characteristic Liquidity Premium Trend

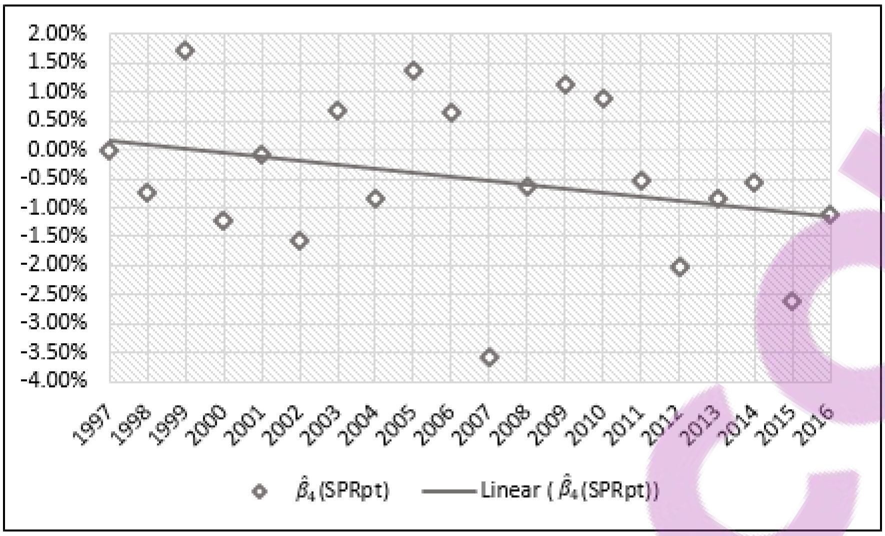


Figure 1 illustrates the trend in the characteristic liquidity premium. The downward trend for $\hat{\beta}_4(SPR_{pt})$ is in line with the findings of a decreasing characteristic liquidity premium by Ben-Rephael et al. (2015). Their results extended to 2011 which means that the findings of significant negative characteristic liquidity for the years 2012, 2015 and 2016 in this study was unaccounted for in their paper. The 5-year results for the characteristic liquidity premium further supports the notion of a decreasing liquidity premium. $\hat{\beta}_4(SPR_{pt})$ for the period 2011-2016 had a negative value of 1.44% and was significant at 1%. The sub-period of 2007-2011 also had a negative premium of 0.55%, however not significant. The results point towards a continuing decrease of the premium.

Figure 2: Systematic Liquidity Premium Trend

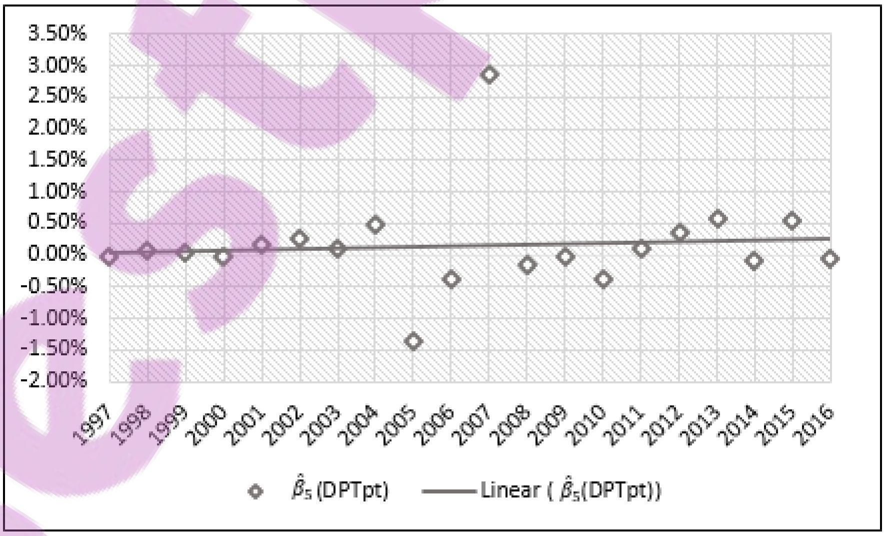


Figure 2 presents a less distinct upward trend for $\hat{\beta}_5(DPT_{pt})$ with small changes. Due to the almost flat looking trend and the presence of a deviant value, it is difficult to draw any further conclusions regarding the forces behind it. The systematic liquidity premium has been positive for the entire 20-year period and for the majority of the sub-periods. These results are also in line with the research done by Ben-Rephael et al. (2015). This could be explained by that investors still requires a liquidity premium for stocks which are more sensitive to systematic shocks to the market liquidity.

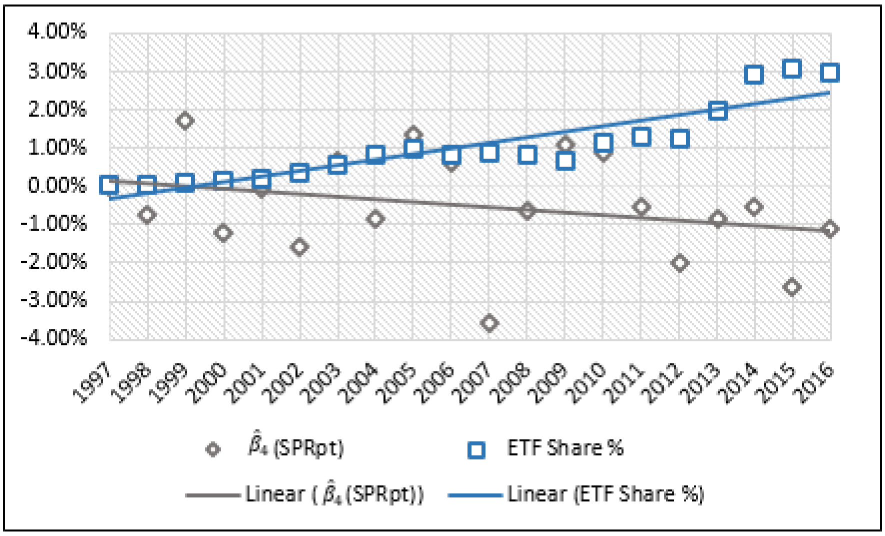
The fluctuations in the 1-year results, which take on both positive and negative significant values on the 5% level, could point to something not discussed in much detail previously in the thesis, the stock market sentiment. The phenomenon called "flight to liquidity" has previously been presented and is something that can be spotted in the results, mainly for the characteristic liquidity premium. The characteristic liquidity premium has a negative value in the years of 2000, 2007 and 2008. These years are characterised by a weak development of the NASDAQ Composite index and are today considered stock-market crashes. The negative characteristic liquidity premium behaves in a similar pattern as the phenomenon flight to liquidity, meaning that there is a sell-off in less liquid stocks (Amihud, 2002; Pástor & Stambaugh, 2003). However, only the result for 2007 was significant, which limits the support for the phenomenon. The systematic liquidity premium shows a similar pattern, with exception for 2007. During 2007 the systematic liquidity premium reached the highest value during the entire 20-year period. This result contradicts the flight to liquidity phenomenon, however, since the market crash began in the second part of 2007, the magnitude of the first part of the year could outweigh the latter part. A second explanation to 2007 could be the influence of possible extreme values in the data sample. Moreover, the systematic liquidity premium for 2000 and 2008 were marginally negative and not significant at any level which further limit the support for a flight to liquidity phenomenon.

Although the majority of the characteristic liquidity premium estimates are negative, some years show a positive premium. During 1999, prior to the burst of the Dot-com bubble, the characteristic liquidity premium was 1.74% and was significant at 5%. Moreover, in 2009, the year after the stock market crash, the characteristic liquidity premium was 1.12% and significant at 5%. We argue that these findings are a result of increased positive sentiment from investors. Specifically, in 1999 when the market was optimistic of the future earnings of companies due to the establishment of the internet. The average first day returns of Initial Public Offerings (IPOs) were 73% and the median was 40% (Ljungqvist & Wilhelm, 2003). This illustrates how investors were valuing newly listed companies and the optimism that was present. In 2009, the stock market began to rise from the financial crisis market trough. The global recovery took several years but the investor sentiment increased as a result of emergency actions taken by the United States Federal Reserve (Foo & Witkowska, 2017). The positive characteristic liquidity premium in 2009 likely reflects a recovery of high spread stocks which underperformed the market in 2007 and 2008. The systematic liquidity premium does not show the same pattern as the characteristic liquidity premium. Even though 1999 had a positive premium, it was not to the same extent as the characteristic liquidity premium and not significant.

5.3 ETFs' Effect on the Liquidity Premium

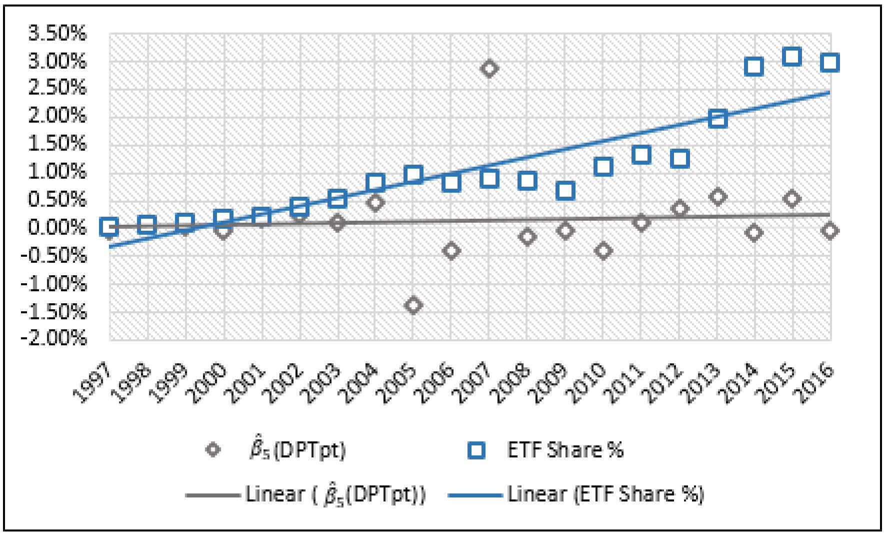
The regression of ETFs' share of the total United States ETF market against the liquidity coefficients did not show any significant relationship as presented in Table 11. In Figure 3 and Figure 4 below, the ETF and liquidity variables are plotted against each other and their relationship is analysed in more detail.

Figure 3: Relationship ETF and SPR



Our hypothesis was that the characteristic liquidity premium would have continued to decrease and that ETFs would have influenced it. This was based on the idea that illiquid stocks are underrepresented in ETFs because it can impair the tracking error. From Figure 3 it becomes clear that the premium has indeed continued to decrease, however, the empirical result presented in Table 11 did not statistically confirm that the impact from ETFs was significant.

Figure 4: Relationship ETF and DPT



We did not make any hypotheses on how the systematic liquidity premium has been affected by ETFs. It is hard to spot any relationship by looking at Figure 4 and the empirical result presented in Table 11 did not indicate any significant correlation as well. The small increasing trend might to some extent be an effect of the extreme value in 2007.

There are several possible reasons for not finding a significant relationship between the ETFs' share of the market and the two liquidity premiums. One possible reason is the drawback of only having a twenty-year period when conducting the statistical tests. Due to difficulties in obtaining monthly data of the growth of ETFs, it was not possible to increase the number of periods. Moreover, in the beginning of the period ETFs' share of the equity market was still negligible as presented in Table 10. For example, the share was 0.03% in 1997, 0.07% in 1998 and 0.12% in 1999 which makes it difficult to spot any correlation when the premiums are quite volatile and can differ by 1% or more between two years. This also implies that including earlier years would not benefit the tests further. When ETFs have existed for more years and possibly have taken a larger share of the market it might become clearer what part they have at determining the level of the premiums. A possible solution to the problem with ETFs' negligible share of the total market could be to use indexed equities as a whole instead of ETFs since indexed products existed before ETFs were created and have a larger share of the equities market than ETFs. However, such a test would not concern ETFs but rather how indexed products affect the liquidity premium which would disregard the distinctive structure ETFs have.

Because this is the first study we know of that have attempted to find a connection between the steep capital inflow to the ETF market and the liquidity premium, it is not possible to compare the results with what other researchers have found. It is likely that as ETFs become a larger and more important part of the financial market more papers will be written that covers the subject.

5.4 Liquidity Based Trading Strategies

In the purpose of this thesis it was stated that one of the reasons why conducting a study of the liquidity premium is interesting is because it can give insight whether investors who are willing to take on the risk of illiquidity can earn excess returns. Creating investment strategies based on backtracking can be flawed, both in the short-term and in the long-term. The outcome is uncertain because backtracking is based on historical data and because it is unclear whether stock market participants will continue to behave in the same way as they have done historically. Disregarding the uncertainty, two different investment strategies are discussed in the two paragraphs below based on the results of this thesis.

There is a slight upward trend in $\hat{\beta}_5(DPT_{pt})$ and it has had a significantly positive relationship with return over the full period at the 1% level. These two factors could open up for a long-term investment strategy. By taking a long position in stocks with a high DPT value, an investor could yield excess returns. However, because of the fluctuations between single years, short-term positions could prove disadvantageous against the broader market. Furthermore, the result could be heavily impacted by 2007 which was a positive deviant estimate, excluding this year could alter the backtracking result and the strategy might then not have been statistically effective. From the results of the study, taking a long position in stocks with a 1% higher DPT value than the market average and taking a short position in the market by the same amount would yield an investor a monthly return of 0.15%.

In contrast, because of the more distinct downward trend in $\hat{\beta}_4(SPR_{pt})$ as well as its significantly negative relationship with return over the entire period at the 5% level, a long-term investment strategy could be to take a short position in stocks with a high spread. Similar to taking a position based on the depth in the short-term, the outcome of taking a position based on the spread in the short-term would be uncertain. Taking a short position in stocks with a high spread contradicts what much previous research have proposed as an investment strategy, but it is an effect of the downward trend in the premium that has been observed. Based on the results, shorting stocks with a 1% higher spread against the market average and taking a long position in the market would yield an investor a monthly return of 0.49%.

6 Conclusions

In this chapter conclusions based to the research questions and with support from the empirical results and the following analysis are presented.

The aim of the thesis was to investigate how the liquidity premium has developed between 1997 and 2016 and if the capital inflow to the ETF market has influenced any alteration in the premium. In order to be able to conduct the statistical tests an extended model of the Fama-French Three-Factor Model was developed which was used to test whether the characteristic or systematic liquidity of a stock influence the return. Moreover, two time-series regressions were performed on the resulting liquidity coefficients in order to test the impact of the relative market size of ETFs on the two types of liquidity risk premiums.

The first research question of the thesis addresses how the liquidity premium has developed between 1997 and 2016. This is interesting because investors who are willing to take on the additional risk of illiquidity have historically been able to receive excess returns. The empirical results from this study suggest that liquidity factors, and in particular, the characteristic liquidity affect stock market returns on NASDAQ Composite. The results reveal that both the characteristic and the systematic liquidity premiums fluctuates year-over-year by sign, magnitude and statistical significance. The characteristic liquidity premium has a distinct negative trend, while smaller fluctuations in the systematic liquidity premium makes it hard to spot any trend. Over the entire test period, the characteristic liquidity premium has been significantly negative while the systematic liquidity premium has been significantly positive. We can conclude that investors have not been adequately compensated for taking on the risk of illiquidity in terms of a high bid-ask spread.

The second research question of the thesis was if changes in the liquidity premium can be linked to capital inflow to the ETF market. This is interesting because if ETFs have impacted the liquidity premium it would have been skewed away from its normal state which in turn might open up for new investing strategies. The results from the statistical tests did not establish any statistical proof of correlation between the ETF variable and the characteristic or systematic liquidity premiums. This means that no conclusions can be made whether the growth of ETFs has contributed to the alteration in the distribution of returns between more and less liquid stocks. Even though the result is insignificant, the plotted relationship between the market share of ETFs and the characteristic liquidity premium is negative which is in line with the expectations of the thesis.

7 Discussion

In this chapter, the conclusions from the thesis are compared to the findings of other papers. There is also a part where possible improvements to the method are highlighted as well as a part where potential future research is discussed.

The results of this thesis support the notion of a negative and decreasing liquidity premium by presenting new evidence. A decreasing premium would indicate that the stock market participants no longer demands a return premium for taking on the risk of illiquidity. The finding of a negative characteristic liquidity premium contradicts previous research made by Amihud and Mendelson (1986a) and Eleswarapu (1997) among others who found a positive premium. On the other hand, our result is in line with the finding of Ben-Rephael et al. (2015) of no positive premium and supports Brennan and Subrahmanyam (1996) finding of a significant negative premium. Using an investing strategy based on a stock's spread appears to have been a less efficient investment strategy for the past two decades than it has been historically. Rather than buying stocks with a high spread, the results of this study would suggest taking a short position in these stocks instead. Given the results, the liquidity risk has not compensated investors with excess returns over a longer time. This should decrease the incentive to invest in less liquid stocks. However, individual stocks with low liquidity could still be beneficial in a portfolio from a diversification perspective.

Although outside the scope of the purpose of this study, the empirical results suggest that size is insignificant for determining return. Investing in stocks with a low market capitalisation would hence not generate excess return over the full period when including the liquidity variables. The finding is in line with the results of Amihud and Mendelson (1986a) but contradicts the results of Eleswarapu and Reinganum (1993) who find a size effect even after including liquidity. Moreover, book-to-market value is not significant either which makes two of the risk factors accounted for in the Fama-French Three-Factor Model insignificant. The empirical results from the thesis therefore indicate that liquidity explains more of a stock's return than the commonly included risk variables size and book-to-market value which is a good reason to include a liquidity proxy when researching stock returns.

We have identified areas with potential for improvement in the method. Firstly, to prevent computational complications stocks with incomplete data were removed. Allowing for incomplete data sets would have led to more observations. This could have made it possible to receive more accurate as well as more significant results. Secondly, as the scope of this study is limited to NASDAQ Composite, the findings can only be applied to that market. Including NYSE as well would have made it possible to generalise the findings for the total United States equities market. Lastly, a possible omitted variable in the ETF regression is the average holding period of investors. Ben-Rephael et al. (2015) proposed that since

ETFs are long-term holders of assets, they are not as susceptible to transaction costs in the form of spreads as short-term holders and would hence not require the same excess return from illiquid assets. Including the average holding period would account for how susceptible investors are to transaction costs and could hence help to explain the variance in the characteristic liquidity premium.

7.1 Further Studies

New ideas of possible research adjacent to the subject were formed based on the problems and results that was found during the process of writing the thesis. A possible further research subject would be to test if total indexed equities rather than solely ETFs has altered the characteristic liquidity premium. In addition to ETFs, index mutual funds would be included in such a study. The reason to focus on the characteristic liquidity premium would be because the systematic liquidity premium lacks a distinct trend.

Furthermore, the limited geographical area of this study allows for similar studies in other countries which could have different results. Such a study would preferably be conducted on a country with an illiquid market to see if there are any differences since this study was made on one of the most liquid markets in the world. Other stock exchanges in the United States could also be of interest. However, since Ben-Rephael et al. (2015) did not find any significant liquidity premium on NYSE, chances are no significant results would be obtained.

This thesis used the bid-ask spread and the ILLIQ variable as proxies for liquidity, but previous research has used other measures. Further studies could be done to test the relationship between the capital inflow to ETFs and the other measures of liquidity. Given the lack of research before this study regarding ETFs and the liquidity premium, additional research could contribute with new insights of the relationship.

Further studies on the same subject on NASDAQ could also be performed in the future when there are more years of data regarding ETFs. Naturally, as the share of ETFs' equity value in relation to the total market increases, there is a higher chance to statistically conclude effects made by them.

References

Amihud, Y., & Mendelson, H. (1986a). Asset Pricing and the Bid-Ask Spread. *Journal of Financial Economics, 17*(2), pp. 223-249.

Amihud, Y., & Mendelson, H. (1986b). Liquidity and Stock Returns. *Financial Analysts Journal*, *42*(3), pp. 43-48.

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets, 5*(1), pp. 31-56.

Baran, L. C., & King, T. H. D. (2014). S&P 500 Index Reconstitutions and Information Asymmetry. *Applied Financial Economics, 24*(11), pp. 777-791.

Ben-Rephael, A., Kadan, O., & Wohl, A. (2015). The Diminishing Liquidity Premium. *Journal of Financial and Quantitative Analysis, 50*(1-2), pp. 197-229.

Beneish, M. D., & Whaley, R. E. (1996). An Anatomy of the "S&P Game": The Effects of Changing the Rules. *Journal of Finance, 51*(5), pp. 1909-1930.

Bernstein, P. L. (1987). Liquidity, Stock Markets, and Market Makers. *Financial Management*, *16*(2), pp. 54-62.

Bodie, Z., & Merton, R. C. (2000). *Finance: International Edition*. Upper Saddle River, NJ: Prentice-Hall, Inc.

Bogle, J. C. (2016). The Index Mutual Fund: 40 Years of Growth, Change, and Challenge. *Financial Analysts Journal, 72*(1), pp. 9-13.

Bradrania, R. M., & Peat, M. (2014). Characteristic Liquidity, Systematic Liquidity and Expected Returns. *Journal of International Financial Markets, Institutions & Money, 33*, pp. 78-98.

Brennan, M. J., & Subrahmanyam, A. (1996). Market Microstructure and Asset Pricing: On the Compensation for Illiquidity in Stock Returns. *Journal of Financial Economics, 41*(3), pp. 441-464.

Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica, 47*(5), pp. 1287-1294.

Buetow, G. W., & Henderson, B. J. (2012). An Empirical Analysis of Exchange-Traded Funds. *Journal of Portfolio Management, 38*(4), pp. 112-127.

Chalmers, J. M. R., & Kadlec, G. B. (1998). An Empirical Examination of the Amortized Spread. *Journal of Financial Economics, 48*(2), pp. 159-188.

Chen, H., Noronha, G., & Singal, V. (2004). The Price Response to S&P 500 Index Additions and Deletions: Evidence of Asymmetry and a New Explanation. *Journal of Finance, 59*(4), pp. 1901-1930.

Chen, J., & Sherif, M. (2016). Illiquidity Premium and Expected Stock Returns in the UK: A New Approach. *Physica A: Statistical Mechanics and its Applications, 458*, pp. 52-66.

Coval, J., & Stafford, E. (2007). Asset Fire Sales (and Purchases) in Equity Markets. *Journal of Financial Economics*, *86*(2), pp. 479-512.

Degutis, A., & Novickyte, L. (2014). The Efficient Market Hypothesis: A Critical Review of Literature and Methodology. *Ekonomika, 93*(2), pp. 7-23.

Eleswarapu, V. R., & Reinganum. M. R. (1993). The Seasonal Behavior of the Liquidity Premium in Asset Pricing. *Journal of Financial Economics*, *34*(3), pp. 373-386.

Eleswarapu, V. R. (1997). Cost of Transacting and Expected Returns in the Nasdaq Market. *Journal of Finance, 52*(5), pp. 2113-2127.

Esterby-Smith, M., Thorpe, R., & Jackson, P. R. (2015). *Management and business research* (5th ed.). London: SAGE Publications Ltd.

Fama, E. F. (1965). The Behavior of Stock-Market Prices. *Journal of Business, 38*(1), pp. 34-105.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance*, *25*(2), pp. 383-417.

Fama, E. F., & French, K. R. (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics, 33*(1), pp. 3-56.

Fama, E. F., & MacBeth, J. D. (1973). Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy, 81*(3), pp. 607-636.

French, K. R. (2018). *Fama/French 3 Factors*. Retrieved March 5, 2018, from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Frino, A., & Gallagher, D. R. (2001). Tracking S&P 500 Index Funds. *Journal of Portfolio Management, 28*(1), pp. 44-55.

Frino, A., Gallagher, D. R., & Oetomo, T. N. (2005). The Index Tracking Strategies of Passive and Enhanced Index Equity Funds. *Australian Journal of Management, 30*(1), pp. 23-55.

FTSE Russell. (2015). *Guide to Calculation FTSE Global Equity Index Series, v3.0*. Retrieved February 19, 2018, from http://www.ftse.com/products/indices/index-support-guides

Foo, J., & Witkowska, D. (2017). A Comparison of Global Financial Market Recovery after the 2008 Global Financial Crisis. *Folia Oeconomica Stetinensia, 17*(1), pp. 109-128.

Grossman, S. J., & Miller, M. H. (1988). Liquidity and Market Structure. *Journal of Finance, 43*(3), pp. 617-633.

Gujarati, D., & Porter, D. (2009). *Basic econometrics* (5.th ed.). Boston, MA: McGraw-Hill.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica, 47*(1), pp. 153-161.

Hegde, S. P., & McDermott, J. B. (2004). The Market Liquidity of DIAMONDS, Q's, and their Underlying Stocks. *Journal of Banking & Finance, 28*(5), pp. 1043-1067.

Hoechle, D. *help xtfmb*. University of Basel, Basel-Stadt, Switzerland. Retrieved April 16, 2018, from http://fmwww.bc.edu/repec/bocode/x/xtfmb.html.

Huberman, G., & Halka, D. (2001). Systematic Liquidity. *Journal of Financial Research, 24*(2), pp. 161-178.

International Organization of Securities Commissions. (2003). *Indexation: Securities Indices and Index Derivatives*. Retrieved February 15, 2018, from http://www.iosco.org/library/pubdocs/pdf/IOSCOPD143.pdf

Investment Company Institute. (2017). *2017 Investment Company Fact Book, A Review of Trends and Activities in the Investment Company Industry*. Retrieved February 13, 2018, from http://www.icifactbook.org/ch7/17_fb_ch7

Kaul, A., Mehrotra, V., & Morck, R. (2000). Demand Curves for Stocks Do Slope Down: New Evidence from an Index Weights Adjustment. *Journal of Finance, 55*(2), pp. 893-912.

Keim, D. B. (1999). An Analysis of Mutual Fund Design: The Case of Investing in Small-Cap Stocks. *Journal of Financial Economics, 51*(2), pp. 173-194.

Kempf, A., & Korn, O. (1999). Market Depth and Order Size. *Journal of Financial Markets, 2*(1), pp. 29-48.

Kluger, B. D., & Stephan, J. (1997). Alternative Liquidity Measures and Stock Returns. *Review of Quantitative Finance and Accounting, 8*(1), pp. 19-36.

Krause, T., Ehsani, S., & Lien, D. (2014). Exchange-Traded Funds, Liquidity and Volatility. *Applied Financial Economics*, *24*(24), pp. 1617-1630.

Linnainmaa, J. T. (2013). Reverse Survivorship Bias. *Journal of Finance, 68*(3), pp. 789-813.

Lintner, J. (1965a). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *Review of Economics and Statistics, 47*(1), pp. 13-37.

Lintner, J. (1965b). Security Prices, Risk, and Maximal Gains from Diversification. *Journal of Finance, 20*(4), pp. 587-615.

Liu, W. (2006). A Liquidity-Augmented Capital Asset Pricing Model. *Journal of Financial Economics, 82*(3), pp. 631-671.

Ljungqvist, A., & Wilhelm, W. (2003). IPO Pricing in the Dot-com Bubble. *Journal of Finance*, *58*(2), 723-752

Madura, J., & Ngo, T. (2008a). Short interest in exchange-traded funds. *Financial Markets and Portfolio Management, 22*(4), pp. 381-402.

Madura, J., & Ngo, T. (2008b). Impact of ETF Inception on the Valuation and Trading of Component Stocks. *Applied Financial Economics, 18*(12), pp. 995-1007.

Markowitz, H. (1952). Portfolio Selection. *Journal of Finance, 7*(1), pp. 77-91.

Maurer, F., & Williams, O. (2015). Physically Versus Synthetically Replicated Trackers: Is There A Difference In Terms Of Risk? *Journal of Applied Business Research, 31*(1), pp. 131-146.

Menard, S. (2002). *Longitudinal research* (2.nd ed.). Thousand Oaks, CA: Sage Publications.

Mossin, J. (1966). Equilibrium in a Capital Asset Market. *Econometrica*, *34*(4), pp. 768-783.

MSCI. (2017). *MSCI US Equity Indexes Methodology*. Retrieved February 17, 2018, from https://www.msci.com/eqb/methodology/meth_docs/MSCI_Nov17_USEI_Methodology.pdf

Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica, 55*(3), pp. 703-708.

Pástor, L., & Stambaugh, R. F. (2003). Liquidity Risk and Expected Stock Returns. *Journal of Political Economy, 111*(3), pp. 642-685.

Ramsey, J. B. (1969). Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis. *Journal of the Royal Statistical Society, 31*(2), pp. 350-371.

Saunders, M., Lewis, P., & Thornhill, A. (2009) *Research Methods for Business Students* (5.th ed.). Essex: Pearson Education.

Schmidt, C., & Fahlenbrach, R. (2017). Do Exogenous Changes in Passive Institutional Ownership Affect Corporate Governance and Firm Value? *Journal of Financial Economics, 124*(2), pp. 285-306.

Sharp, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance, 19*(3), pp. 425-442.

Shleifer, A. (1986). Do Demand Curves for Stocks Slope Down? *Journal of Finance, 41*(3), pp. 579-590.

Standard & Poor's. (2018). *Float Adjustment Methodology*. Retrieved February 17, 2018, from us.spindices.com/documents/index-policies/methodology-sp-float-adjustment.pdf?force_download=true

Stoll, H. R., & Whaley, R. E. (1983). Transaction Costs and the Small Firm Effect. *Journal of Financial Economics, 12*(1), pp. 57-79.

Stoll, H. R. (1989). Inferring the Components of the Bid-Ask Spread: Theory and Empirical Tests. *Journal of Finance, 44*(1), pp. 115-134.

The World Bank. (n.d.). *Stocks traded, total value (current US$)*. Retrieved February 23, 2018, from https://data.worldbank.org/indicator/CM.MKT.TRAD.CD?end=2016&locations=US&start=1997&view=chart

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press.

Appendices

Appendix A. Equation (6) Variables

$RPRF_{pt}$
The downloaded prices were daily closing prices adjusted for capital and stock actions such as dividends and splits, which made it possible to make comparisons of the price over time. Returns were then calculated using the formula: $LN(P_{it+1}/P_{it})$, where P_{it+1} was the price of the last day of the month of interest and P_{it} was the price in the last day of the month before. By subtracting the risk-free rate from the stock returns the excess returns were obtained. Portfolio excess returns were calculated by taking the average excess returns of the underlying stocks.
$RMRF_t$
The excess return of the market was directly obtained from Kenneth French's website.
MV_{pt}
The market values of individual stocks were downloaded and the market values of portfolios was calculated as the average market value of the underlying stocks. Datastream calculates the market values of individual stocks by taking the share price multiplied by the number of shares outstanding.
$BTMV_{pt}$
Datastream calculates market-to-book values as the market value of a company divided by the balance sheet value of equity. To convert the data to book-to-market values, which is more commonly used in asset pricing models, the inverse of the market-to-book values were calculated. The book-to-market value of a portfolio was calculated as the average book-to-market values of the underlying stocks.
SPR_{pt}
After downloading the daily high and daily low prices, the spreads for each day were calculated by using the formula: $(PH_{id}-PL_{id})/PH_{id}$, where PH_{id} was the high price of stock *i* on day *d* and PL_{it} was the low price of stock *i* on day *d*. The average spreads for each month were then calculated. Portfolio spreads were calculated as the average monthly spreads of the underlying stocks.
DPT_{pt}
To calculate the DPT of stocks we had to, in addition to downloading the adjusted prices of stocks, download data on three additional variables depending on the year: turnover by value (VA), turnover by volume (VO) and volume weighted average price (WVAP). In Datastream, the data for the turnover by value of stocks was only available for the years 2015 and 2016. For the rest of the years turnover by volume and a price variable was used to calculate turnover by value. Between 1999 and 2014, value weighted average price was used and for 1997 and 1998 the adjusted closing price was used because the value weighted average price was unavailable. To calculate turnover by value for individual stocks, turnover by volume was multiplied with the price variable, either value weighted average price or the adjusted closing price. The absolute returns for a month was calculated by using the Excel formula =ABS() on the calculated returns. DPT for each day was then calculated as $\|R_{id}\|/VOL_{id}$, with VOL being the turnover by value. The average DPT for a stock in a month was then calculated by taking the average, and for a portfolio the average of the underlying stocks.

Appendix B. Stocks Removed Due to Severe Outliers

Abraxas Petroleum
Electro Scientific
Emmis Communications
Novavax
Sigma Designs

Appendix C. The 210 Included Stocks

Acxiom
Adobe Systems
Adtran
AMD
Advanced Energy
Aegion
Agilysys
AlexionPharms
American Software
American Supercond.
Amgen
Analog Devices
Analogic
Apogee Enterprises
Apple
Applied Materials
ArcBest
ArQule
Ascena Retail Group
ASML Holding
Aspen Technology
Astec Industries
Autodesk
Automatic Data Proc.
Avid Technology
Avis Budget Group
Bed Bath & Beyond
BioScrip
Black Box
Brooks Automation
CONMED
CSX
CA
Cadence Design Sys.
Cascadian Therap
Casey's General Stores
Celgene
Central Garden & Pet
Century Aluminum
Cerner
CP Software Tech.
Cheesecake Factory
Cincinnati Financial
Cintas
Cirrus Logic
Cisco Systems
Citrix Systems
City Holding
Cognex
Coherent
Cohu
Comcast
Commerce Union Banc.
Copart
Costco Wholesale
Cracker Barrel OC
Cree
Cypress Semicond.
DSP Group
Dentsply Sirona
Digi International
Dime Community Banc.
Dish Network
Dollar Tree
E*Trade Financial
Electr. for Imaging
Electronic Arts
Encore Wire
Ericsson
Evine Live
Exp. Int. of Washington
Express Scripts Holding
FD of America
Fastenal
Fifth Third Bancorp
Finish Line
First Financial Bank
First Midwest Bank
Fiserv
Flex
FLIR Systems
Forrester Research
Fulton Financial
Gentex
Geron
Goodyear Tire & Rub
Hain Celestial Group
Harmonic
Hasbro
Heartland Express
Helen of Troy
Henry Schein
Hologic
Hospitality Prop. Trust
Huntington Bancshares
ICU Medical
II-VI
IAC
IDEXX Laboratories
Immunogen
Immunomedics
Incyte
Independent Bank
Ingles Markets
Insight Enterprises
Integra LifeSciences
Integrated Device Tech.
Intel
International Speedway
Intuit
Itron
Hunt JB Transp. Serv.
Jack Henry & Assoc.
Jack in the Box
Jakks Pacific
KLA-Tencor
Kelly Services
Kforce
Kimball International
Kopin
Kulicke and Soffa Indu.
Lam Research
Lamar Advertising
Lancaster Colony
Landstar System
Lattice Semiconductor
Lincoln Electric
Littelfuse
MTS Systems
Mattel
Matthews Int.
Maxim Integr. Products
Maxwell Technologies
MGE Energy
Microchip Technology
Micron Technology
Microsoft
MicroVision
Herman Miller
Mobile Mini
Myriad Genetics
Nanometrics
National Instruments
NBT Bancorp
Neogen
NetApp
Neurocrine Biosciences
Nordson
Northern Trust
Nuance Comm.
OceanFirst Financial
Office Depot
Orbotech
O'Reilly Automotive
Otter Tail
PACCAR
Papa John's Int.
ParkerVision
Patterson-UTI Energy
Paychex
PDL BioPharma
Pegasystems
Penn National Gaming
PepsiCo
Photronic
Popular Inc
Progress Software
Qualcomm
Quidel
Radisys
Sanmina
Scholastic
SeaChange Int.
Selective Ins. Group
Semtech
Shoe Carnival
Sinclair Broadcast
Skyworks Solutions
SLM
Snyder's-Lance
Spectrum Pharma
State Auto Financial
Steel Dynamics
Stein Mart
Stericycle
Strayer Education
SVB Financial Group
Sykes Enterprises
Symantec
Synopsys
Tech Data
Texas Instruments
Trimble
TrustCo Bank Corp NY
Trustmark
Universal Electronics
Universal Forest Prod.
Urban Outfitters
Vaxart
Veeco Instruments
Viavi Solutions
Vical
Vicor
Vodafone Group
Walgreens Boots Alliance
Washington Federal
Wendy's Company
Werner Enterprises
Zions Bancorp
Zix

Appendix D. 20-Year Regression Output Equation (6)

1997 - 2016	*Coefficient*	*Newey-West Std. Err.*	*t-Statistic*	*P-Value*
Intercept	0.02	0.00	7.96	0.00
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)			
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	1.18	0.24
$\hat{\beta}_3(BTMV_{pt})$	0.00	0.00	0.38	0.70
$\hat{\beta}_4(SPR_{pt})$	-0.49	0.20	-2.48	0.01
$\hat{\beta}_5(DPT_{pt})$	0.15	0.05	3.03	0.00

N	*Prob > F*	*avg. R-squared*
2400	0.00	0.61

Appendix E. 5-Year Regressions Output Equation (6)

1997 - 2001	***Coefficient***	***Newey-West Std. Err.***	***t-Statistic***	***P-Value***
Intercept	0.02	0.01	2.40	0.02
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)			
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	1.84	0.07
$\hat{\beta}_3(BTMV_{pt})$	0.00	0.01	-0.56	0.58
$\hat{\beta}_4(SPR_{pt})$	-0.07	0.23	-0.30	0.77
$\hat{\beta}_5(DPT_{pt})$	0.05	0.02	2.11	0.04

N	***Prob > F***	***avg. R-squared***
600	0.08	0.64

2002 - 2006	***Coefficient***	***Newey-West Std. Err.***	***t-Statistic***	***P-Value***
Intercept	0.01	0.00	1.45	0.15
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)			
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	0.28	0.78
$\hat{\beta}_3(BTMV_{pt})$	0.01	0.02	0.39	0.70
$\hat{\beta}_4(SPR_{pt})$	0.03	0.39	0.09	0.93
$\hat{\beta}_5(DPT_{pt})$	-0.19	0.21	-0.91	0.36

N	***Prob > F***	***avg. R-squared***
600	0.93	0.52

2007 - 2011	***Coefficient***	***Newey-West Std. Err.***	***t-Statistic***	***P-Value***
Intercept	0.02	0.01	1.78	0.08
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)			
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	-0.87	0.39
$\hat{\beta}_3(BTMV_{pt})$	-0.01	0.01	-2.11	0.04
$\hat{\beta}_4(SPR_{pt})$	-0.55	0.54	-1.01	0.32
$\hat{\beta}_5(DPT_{pt})$	0.48	0.39	1.23	0.23

N	***Prob > F***	***avg. R-squared***
600	0.10	0.62

2011 - 2016	***Coefficient***	***Newey-West Std. Err.***	***t-Statistic***	***P-Value***
Intercept	0.36	0.00	7.76	0.00
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)			
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	-4.76	0.00
$\hat{\beta}_3(BTMV_{pt})$	0.02	0.01	1.33	0.19
$\hat{\beta}_4(SPR_{pt})$	-1.44	0.16	-9.14	0.00
$\hat{\beta}_5(DPT_{pt})$	0.27	0.07	3.88	0.00

N	***Prob > F***	***avg. R-squared***
600	0.00	0.65

Appendix F. 1-Year Regressions Output Equation (6)

1997	Coefficient	Newey-West Std. Err.	t-Statistic	P-Value	N	Prob > F	avg. R-squared
Intercept	0.01	0.03	0.27	0.79	120	0.61	0.57
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)						
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	1.21	0.25			
$\hat{\beta}_3(BTMV_{pt})$	0.03	0.03	1.10	0.29			
$\hat{\beta}_4(SPR_{pt})$	-0.02	0.76	-0.03	0.98			
$\hat{\beta}_5(DPT_{pt})$	-0.03	0.06	-0.52	0.62			

1998	Coefficient	Newey-West Std. Err.	t-Statistic	P-Value	N	Prob > F	avg. R-squared
Intercept	0.08	0.04	1.98	0.07	120	0.00	0.61
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)						
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	5.65	0.00			
$\hat{\beta}_3(BTMV_{pt})$	-0.06	0.04	-1.45	0.18			
$\hat{\beta}_4(SPR_{pt})$	-0.74	1.15	-0.64	0.54			
$\hat{\beta}_5(DPT_{pt})$	0.08	0.11	0.69	0.50			

1999	Coefficient	Newey-West Std. Err.	t-Statistic	P-Value	N	Prob > F	avg. R-squared
Intercept	-0.06	0.03	-1.79	0.10	120	0.08	0.71
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)						
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	1.00	0.34			
$\hat{\beta}_3(BTMV_{pt})$	-0.01	0.07	-0.20	0.85			
$\hat{\beta}_4(SPR_{pt})$	1.71	0.56	3.05	0.01			
$\hat{\beta}_5(DPT_{pt})$	0.04	0.10	0.44	0.67			

2000	Coefficient	Newey-West Std. Err.	t-Statistic	P-Value	N	Prob > F	avg. R-squared
Intercept	0.08	0.04	2.02	0.07	120	0.03	0.58
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)						
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	-3.40	0.01			
$\hat{\beta}_3(BTMV_{pt})$	0.00	0.03	-0.08	0.93			
$\hat{\beta}_4(SPR_{pt})$	-1.21	0.99	-1.22	0.25			
$\hat{\beta}_5(DPT_{pt})$	-0.03	0.02	-1.47	0.17			

2001	Coefficient	Newey-West Std. Err.	t-Statistic	P-Value	N	Prob > F	avg. R-squared
Intercept	0.00	0.03	0.18	0.86	120	0.00	0.73
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)						
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	-4.91	0.00			
$\hat{\beta}_3(BTMV_{pt})$	0.02	0.02	1.00	0.34			
$\hat{\beta}_4(SPR_{pt})$	-0.07	0.49	-0.15	0.88			
$\hat{\beta}_5(DPT_{pt})$	0.17	0.06	3.11	0.01			

2002	Coefficient	Newey-West Std. Err.	t-Statistic	P-Value	N	Prob > F	avg. R-squared
Intercept	0.00	0.02	0.07	0.95	120	0.01	0.43
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)						
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	1.92	0.08			
$\hat{\beta}_3(BTMV_{pt})$	0.08	0.03	2.49	0.03			
$\hat{\beta}_4(SPR_{pt})$	-1.57	0.61	-2.56	0.03			
$\hat{\beta}_5(DPT_{pt})$	0.25	0.13	2.00	0.07			

2003	Coefficient	Newey-West Std. Err.	t-Statistic	P-Value	N	Prob > F	avg. R-squared
Intercept	0.01	0.01	1.59	0.14	120	0.33	0.63
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)						
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	0.06	0.95			
$\hat{\beta}_3(BTMV_{pt})$	-0.01	0.02	-0.77	0.46			
$\hat{\beta}_4(SPR_{pt})$	0.67	0.38	1.77	0.10			
$\hat{\beta}_5(DPT_{pt})$	0.09	0.07	1.21	0.25			

2004	Coefficient	Newey-West Std. Err.	t-Statistic	P-Value	N	Prob > F	avg. R-squared
Intercept	0.03	0.01	2.35	0.04	120	0.03	0.51
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)						
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	-3.13	0.01			
$\hat{\beta}_3(BTMV_{pt})$	0.01	0.03	0.23	0.83			
$\hat{\beta}_4(SPR_{pt})$	-0.83	0.50	-1.66	0.12			
$\hat{\beta}_5(DPT_{pt})$	0.47	0.30	1.56	0.15			

2005	Coefficient	Newey-West Std. Err.	t-Statistic	P-Value	N	Prob > F	avg. R-squared
Intercept	0.01	0.02	0.63	0.54	120	0.00	0.46
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)						
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	-0.14	0.89			
$\hat{\beta}_3(BTMV_{pt})$	-0.09	0.06	-1.44	0.18			
$\hat{\beta}_4(SPR_{pt})$	1.37	0.63	2.18	0.05			
$\hat{\beta}_5(DPT_{pt})$	-1.36	0.27	-4.97	0.00			

2006	Coefficient	Newey-West Std. Err.	t-Statistic	P-Value	N	Prob > F	avg. R-squared
Intercept	-0.02	0.02	-1.31	0.22	120	0.35	0.57
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)						
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	0.34	0.74			
$\hat{\beta}_3(BTMV_{pt})$	0.03	0.02	1.31	0.22			
$\hat{\beta}_4(SPR_{pt})$	0.64	0.41	1.55	0.15			
$\hat{\beta}_5(DPT_{pt})$	-0.39	0.42	-0.91	0.38			

2007	Coefficient	Newey-West Std. Err.	t-Statistic	P-Value	N	Prob > F	avg. R-squared
Intercept	0.09	0.01	11.30	0.00	120	0.00	0.56
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)						
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	-3.68	0.00			
$\hat{\beta}_3(BTMV_{pt})$	-0.02	0.02	-1.42	0.18			
$\hat{\beta}_4(SPR_{pt})$	-3.57	0.39	-9.22	0.00			
$\hat{\beta}_5(DPT_{pt})$	2.87	0.37	7.86	0.00			

2008	Coefficient	Newey-West Std. Err.	t-Statistic	P-Value	N	Prob > F	avg. R-squared
Intercept	-0.01	0.03	-0.34	0.74	120	0.82	0.51
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)						
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	-0.12	0.90			
$\hat{\beta}_3(BTMV_{pt})$	0.00	0.02	0.24	0.81			
$\hat{\beta}_4(SPR_{pt})$	-0.64	0.55	-1.17	0.27			
$\hat{\beta}_5(DPT_{pt})$	-0.16	0.30	-0.53	0.61			

2009	Coefficient	Newey-West Std. Err.	t-Statistic	P-Value	N	Prob > F	avg. R-squared
Intercept	0.00	0.01	0.00	1.00	120	0.05	0.83
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)						
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	1.19	0.26			
$\hat{\beta}_3(BTMV_{pt})$	-0.03	0.02	-1.62	0.13			
$\hat{\beta}_4(SPR_{pt})$	1.12	0.38	2.96	0.01			
$\hat{\beta}_5(DPT_{pt})$	-0.04	0.21	-0.18	0.86			

2010	Coefficient	Newey-West Std. Err.	t-Statistic	P-Value	N	Prob > F	avg. R-squared
Intercept	0.00	0.01	0.35	0.74	120	0.05	0.59
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)						
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	-0.84	0.42			
$\hat{\beta}_3(BTMV_{pt})$	-0.01	0.01	-1.20	0.26			
$\hat{\beta}_4(SPR_{pt})$	0.89	0.60	1.49	0.17			
$\hat{\beta}_5(DPT_{pt})$	-0.38	0.14	-2.83	0.02			

2011	Coefficient	Newey-West Std. Err.	t-Statistic	P-Value	N	Prob > F	avg. R-squared
Intercept	0.01	0.01	0.86	0.41	120	0.74	0.58
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)						
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	0.99	0.34			
$\hat{\beta}_3(BTMV_{pt})$	0.00	0.04	0.04	0.97			
$\hat{\beta}_4(SPR_{pt})$	-0.54	0.66	-0.81	0.44			
$\hat{\beta}_5(DPT_{pt})$	0.11	0.15	0.73	0.48			

2012	Coefficient	Newey-West Std. Err.	t-Statistic	P-Value	N	Prob > F	avg. R-squared
Intercept	0.05	0.01	4.26	0.00	120	0.00	0.59
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)						
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	-4.26	0.00			
$\hat{\beta}_3(BTMV_{pt})$	0.01	0.05	0.13	0.90			
$\hat{\beta}_4(SPR_{pt})$	-2.02	0.74	-2.73	0.02			
$\hat{\beta}_5(DPT_{pt})$	0.37	0.20	1.87	0.09			

2013	Coefficient	Newey-West Std. Err.	t-Statistic	P-Value	N	Prob > F	avg. R-squared
Intercept	0.04	0.01	4.35	0.00	120	0.13	0.67
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)						
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	-1.02	0.33			
$\hat{\beta}_3(BTMV_{pt})$	0.00	0.02	0.03	0.97			
$\hat{\beta}_4(SPR_{pt})$	-0.86	0.62	-1.38	0.20			
$\hat{\beta}_5(DPT_{pt})$	0.56	0.23	2.44	0.03			

2014	Coefficient	Newey-West Std. Err.	t-Statistic	P-Value	N	Prob > F	avg. R-squared
Intercept	0.03	0.01	4.37	0.00	120	0.59	0.65
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)						
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	1.19	0.26			
$\hat{\beta}_3(BTMV_{pt})$	-0.02	0.03	-0.83	0.43			
$\hat{\beta}_4(SPR_{pt})$	-0.55	0.61	-0.90	0.39			
$\hat{\beta}_5(DPT_{pt})$	-0.08	0.18	-0.43	0.68			

2015	Coefficient	Newey-West Std. Err.	t-Statistic	P-Value	N	Prob > F	avg. R-squared
Intercept	0.04	0.01	2.80	0.02	120	0.00	0.67
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)						
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	-2.34	0.04			
$\hat{\beta}_3(BTMV_{pt})$	0.06	0.02	3.31	0.01			
$\hat{\beta}_4(SPR_{pt})$	-2.63	0.74	-3.53	0.01			
$\hat{\beta}_5(DPT_{pt})$	0.54	0.26	2.10	0.06			

2016	Coefficient	Newey-West Std. Err.	t-Statistic	P-Value	N	Prob > F	avg. R-squared
Intercept	0.02	0.01	2.28	0.04	120	0.00	0.68
$\hat{\beta}_1(\mathrm{RMRF}_t)$	(Omitted)						
$\hat{\beta}_2(\mathrm{MV}_{pt})$	0.00	0.00	-2.68	0.02			
$\hat{\beta}_3(BTMV_{pt})$	0.05	0.03	1.63	0.13			
$\hat{\beta}_4(SPR_{pt})$	-1.12	0.26	-4.30	0.00			
$\hat{\beta}_5(DPT_{pt})$	-0.04	0.16	-0.28	0.79			

Appendix G. ETF Regression Output Equation (8)

1997 - 2016	Coefficient	Std.Err.	t-Statistic	P-Value
Intercept	0.00	0.00	0.01	0.99
ETF_y	-0.39	0.32	-1.21	0.24

N	Prob > F	R-squared
20	0.24	0.07

Appendix H. ETF Regression Output Equation (9)

1997 - 2016	*Coefficient*	*Std.Err.*	*t-Statistic*	*P-Value*
Intercept	0.00	0.00	0.40	0.69
ETF_y	0.04	0.19	0.23	0.82

N	*Prob > F*	*R-squared*
20	0.82	0.00